

# **THE 10<sup>TH</sup> INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION**

*SEPTEMBER 12-14 2011  
ROME, ITALY*



**EDITED BY**  
***AGOSTINO BRUZZONE***  
***CLAUDIA FRYDMAN***  
***MARINA MASSEI***  
***MIKE MCGINNIS***  
***MIQUEL ANGEL PIERA***  
***GREGORY ZACHAREWICZ***

**PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2011**

ISBN 978-88-903724-5-2

## © 2011 DIPTTEM UNIVERSITÀ DI GENOVA

RESPONSIBILITY FOR THE ACCURACY OF ALL STATEMENTS IN EACH PAPER RESTS SOLELY WITH THE AUTHOR(S). STATEMENTS ARE NOT NECESSARILY REPRESENTATIVE OF NOR ENDORSED BY THE DIPTTEM, UNIVERSITY OF GENOA. PERMISSION IS GRANTED TO PHOTOCOPY PORTIONS OF THE PUBLICATION FOR PERSONAL USE AND FOR THE USE OF STUDENTS PROVIDING CREDIT IS GIVEN TO THE CONFERENCES AND PUBLICATION. PERMISSION DOES NOT EXTEND TO OTHER TYPES OF REPRODUCTION NOR TO COPYING FOR INCORPORATION INTO COMMERCIAL ADVERTISING NOR FOR ANY OTHER PROFIT – MAKING PURPOSE. OTHER PUBLICATIONS ARE ENCOURAGED TO INCLUDE 300 TO 500 WORD ABSTRACTS OR EXCERPTS FROM ANY PAPER CONTAINED IN THIS BOOK, PROVIDED CREDITS ARE GIVEN TO THE AUTHOR(S) AND THE WORKSHOP.

FOR PERMISSION TO PUBLISH A COMPLETE PAPER WRITE TO: DIPTTEM UNIVERSITY OF GENOA, DIRECTOR, VIA OPERA PIA 15, 16145 GENOVA, ITALY. ADDITIONAL COPIES OF THE PROCEEDINGS OF THE *MAS* ARE AVAILABLE FROM DIPTTEM UNIVERSITY OF GENOA, DIRECTOR, VIA OPERA PIA 15, 16145 GENOVA, ITALY.



# THE 10<sup>TH</sup> INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION

SEPTEMBER 12-14 2011  
ROME, ITALY

## ORGANIZED BY



DIPTeM - UNIVERSITY OF GENOA



LIOPHANT SIMULATION



SIMULATION TEAM



IMCS - INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF  
SIMULATION



MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA



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



MODELING AND SIMULATION CENTER OF EXCELLENCE (MSCOE)



MISS LATVIAN CENTER - RIGA TECHNICAL UNIVERSITY



LOGISIM



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



MISS - UNIVERSITY OF PERUGIA



MISS - BRASILIAN CENTER, LAMCE-COPPE-UFRJ



MISS - MCLEOD INSTITUTE OF SIMULATION SCIENCES



M&SNET - MCLEOD MODELING AND SIMULATION NETWORK



LATVIAN SIMULATION SOCIETY



ECOLE SUPERIEURE D'INGENIERIE EN SCIENCES APPLIQUEES



FACULTAD DE CIENCIAS EXACTAS. INGENIERIA Y AGRIMENSURA



UNIVERSITY OF LA LAGUNA



CIFASIS: CONICET-UNR-UPCAM



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

#### I3M 2011 INDUSTRIAL SPONSORS



PRESAGIS



CAE



CAL-TEK



MAST



AEgis TECHNOLOGIES

#### I3M 2011 MEDIA PARTNERS



MILITARY SIMULATION & TRAINING MAGAZINE



EURO MERCI

## EDITORS

### **AGOSTINO BRUZZONE**

*MISS-DIPTM, UNIVERSITY OF GENOA, ITALY*  
[agostino@itim.unige.it](mailto:agostino@itim.unige.it)

### **CLAUDIA FRYDMAN**

*LSIS, DOMAINE UNIVERSITAIRE DE SAINT-JERME, FRANCE*  
[claudia.frydman@lsis.org](mailto:claudia.frydman@lsis.org)

### **MARINA MASSEI**

*LIOPHANT SIMULATION, ITALY*  
[massei@itim.unige.it](mailto:massei@itim.unige.it)

### **MIKE MCGINNIS**

*UNIVERSITY OF NEBRASKA, USA*  
[mmcginnis@nebraska.edu](mailto:mmcginnis@nebraska.edu)

### **MIQUEL ANGEL PIERA**

*AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*  
[miquelangel.piera@uab.cat](mailto:miquelangel.piera@uab.cat)

### **GREGORY ZACHAREWICZ**

*IMS UNIVERSITY OF BORDEAUX 1, FRANCE*  
[gregory.zacharewicz@u-bordeaux1.fr](mailto:gregory.zacharewicz@u-bordeaux1.fr)

# **THE INTERNATIONAL MEDITERRANEAN AND LATIN AMERICAN MODELING MULTICONFERENCE, I3M 2011**

## **GENERAL CO-CHAIRS**

**AGOSTINO BRUZZONE**, *MISS DIPTÉM, UNIVERSITY OF GENOA, ITALY*

**MIQUEL ANGEL PIERA**, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*

## **PROGRAM CHAIR**

**FRANCESCO LONGO**, *MSC-LES, UNIVERSITY OF CALABRIA, ITALY*

# **THE 10<sup>TH</sup> INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION, MAS 2011**

## **GENERAL CO-CHAIRS**

**CLAUDIA FRYDMAN**, *LSIS, DOMAINE UNIVERSITAIRE DE SAINT-JERÔME, FRANCE*

**MIKE MCGINNIS**, *UNIVERSITY OF NEBRASKA, USA*

## **PROGRAM CO-CHAIRS**

**MARINA MASSEI**, *LIOPHANT SIMULATION, ITALY*

**GREGORY ZACHAREWICZ**, *IMS UNIVERSITÉ BORDEAUX 1, FRANCE*

## MAS 2011 INTERNATIONAL PROGRAM COMMITTEE

THÈCLE ALIX *IMS UNIVERSITÉ BORDEAUX 1, FRANCE*  
ENRICO BOCCA, *SIMULATION TEAM, ITALY*  
AGOSTINO BRUZZONE, *MISS DIPTM UNIVERSITY OF  
GENOA, ITALY*  
DUILIO CURCIO, *MSC-LES, ITALY*  
FABIO DE FELICE, *UNIVERSITY OF CASSINO, ITALY*  
DAVID DEL RIO VILAS, *UNIVERSITY OF LA CORUNA, SPAIN*  
CLAUDIA FRYDMAN, *LSIS, FRANCE*  
ANDREA GRASSI, *UNIVERSITY OF MODENA AND REGGIO  
EMILIA, ITALY*  
YILIN HUANG, *DELFT UNIVERSITY OF TECHNOLOGY,  
NETHERLANDS*  
JANOS SEBESTYEN JANOSY, *MTA KFKI ATOMIC ENERGY  
RESEARCH INSTITUTE, HUNGARY*  
CLAUDIA KRULL, *MAGDEBURG UNIVERSITY, GERMANY*  
FRANCESCO LONGO, *MSC-LES, UNIVERSITY OF CALABRIA,  
ITALY*  
FRANCESCA MADEO, *SIMULATION TEAM, ITALY*  
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*  
MIKE MCGINNIS, *UNIVERSITY OF NEBRASKA, USA*  
AZIZ NAAMANE, *LSIS, FRANCE*  
FEDERICA PASCUCCI, *UNIVERSITY OF ROMA 3, ITALY*  
ANTONELLA PETRILLO, *UNIVERSITY OF CASSINO, ITALY*  
ROBERTO SETOLA, *UNIVERSITY OF ROMA 3, ITALY*  
ADRIANO SOLIS, *YORK UNIVERSITY, CANADA*  
FEDERICO TARONE, *SIMULATION TEAM, ITALY*  
ALBERTO TREMORI, *UNIVERSITY OF GENOA, ITALY*  
ALEXANDER VERBRAECK, *DELFT UNIVERSITY OF  
TECHNOLOGY, NETHERLAND*  
GIUSEPPE VIGNALI, *UNIVERSITY OF PARMA, ITALY*  
GREGORY ZACHAREWICZ, *IMS UNIVERSITÉ BORDEAUX 1,  
FRANCE*

## TRACKS AND WORKSHOP CHAIRS

### AUTOMATION

**CHAIR:** NAAMANE AZIZ, LABORATOIRE DES SCIENCES DE  
L'INFORMATION ET DES SYSTEMES (FRANCE)

### DECISION SUPPORT SYSTEMS APPLICATIONS

**CHAIR:** FABIO DE FELICE, UNIVERSITY OF CASSINO (ITALY)

### INVENTORY MANAGEMENT SIMULATION

**CHAIRS:** ADRIANO SOLIS, YORK UNIVERSITY (CANADA);  
DUILIO CURCIO, MODELING & SIMULATION CENTER  
LABORATORY OF ENTERPRISE SOLUTIONS M&SNET CENTER  
(ITALY)

### MODELING AND SIMULATION OF CRITICAL INFRASTRUCTURES

**CHAIRS:** ROBERTO SETOLA, UNIVERSITY OF ROMA 3 (ITALY);  
FEDERICA PASCUCCI, UNIVERSITY OF ROMA 3 (ITALY).

### MODELING AND SIMULATION IN FOOD, BEVERAGE AND PERISHABLE GOODS INDUSTRY

**CHAIRS:** ANDREA GRASSI, UNIVERSITY OF MODENA AND  
REGGIO EMILIA (ITALY); GIUSEPPE VIGNALI, UNIVERSITY OF  
PARMA (ITALY)

### PRODUCTION SYSTEMS DESIGN

**CHAIRS:** DAVID DEL RIO VILAS, UNIVERSITY OF LA CORUNA,  
(SPAIN); ANTONIO CIMINO, UNIVERSITY OF CALABRIA (ITALY)

### PROMOTING ENTERPRISE INTEROPERABILITY BY SERVICE MODELING & SIMULATION

**CHAIRS:** THECLE ALIX, IMS UNIVERSITE BORDEAUX 1,  
(FRANCE); GREGORY ZACHAREWICZ, IMS UNIVERSITE  
BORDEAUX 1 (FRANCE)

### SIMULATION BASED DESIGN

**CHAIRS:** YILIN HUANG , DELFT UNIVERSITY OF TECHNOLOGY,  
(NETHERLANDS); ALEXANDER VERBRAECK, DELFT  
UNIVERSITY OF TECHNOLOGY (NETHERLANDS)

### SIMULATION IN ENERGY GENERATION AND DISTRIBUTION

**CHAIR:** JANOS SEBESTYEN JANOSY, MTA KFKI ATOMIC  
ENERGY RESEARCH INSTITUTE (HUNGARY)

## CONFERENCE CHAIRS' MESSAGE

### *WELCOME TO MAS 2011*

The Annual Modeling and Applied Simulation Conference is the ideal International Framework for presenting the very best research results, problem solutions and insight on new challenges facing the field of M&S Applications; MAS is in fact a strong pillar of I3M and, even if its among the most young I3M conferences, the 2011 edition represents the 10<sup>th</sup> time MAS event was organized in the Simulation International Community.

Nowadays, Modeling & Simulation plays a significant role in many critical areas and new Simulation Applications are strongly required by many different sectors that expect to use M&S as enabling technology for introduction break-through innovation.

Obviously to introduce simulation into new and complex problems provides big challenges; from market evolution, sustainability issues, product innovation, interactive communications, globalization, mass customization, intensification of competition, and many others. So today it becomes more and more important to develop multidisciplinary teams and networks for succeeding in creating effective M&S application; it is critical to create different specific skills and strong expertise by creating networks involving both academicians, technicians, subject matter experts and practitioners; MAS 2011 is focusing exactly on this goal and represents a very effective moment for the International Community to discuss up-to-date initiatives, opportunities provided by enabling technologies and scenario evolutions.

In fact MAS 2011 provides us with a timely opportunity to reflect and discuss problems and solutions, to identify new issues and to shape future directions for research and industry users. MAS 2011 is also an opportunity for researchers and practitioners from industry, academia and government to come together to share and advance their knowledge, discussing and working together for setting up new R&D proposals and projects based on state of art of M&S techniques and technologies.

The submitted papers was subjected to hard review not only to select best manuscripts, but even to feedback authors with comments and suggestions to improve the overall quality of the contributions; as results the selection shows that all the most innovative themes and issues in this area are under consideration at the MAS 2011 and all mainstream features of Applied Simulation are addressed.

Last but not least, Rome is probably among the most imaginative Metropolis world-wide; in fact this City was Capital of the whole known world for Many Centuries and it is still holding major cultural and tourist attractions: so we hope that this year edition will leave you with a unique cultural, scientific and networking experience in MAS 2011, Rome.



*Claudia Fryman*  
*LSIS*  
*University of Aix Marseille III*



*Mike McGinnis*  
*The Peter Kiewit Institute*  
*University of Nebraska*



*Marina Massei*  
*DIPTM*  
*University of Genoa*



*Gregory Zacharewicz*  
*IMS*  
*University of Bordeaux*

## **ACKNOWLEDGEMENTS**

The MAS 2011 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The MAS 2011 IPC would like to thank all the authors as well as the reviewers for their invaluable work.

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

## **LOCAL ORGANIZATION COMMITTEE**

AGOSTINO G. BRUZZONE, *MISS-DIPTEM, UNIVERSITY OF GENOA, ITALY*

ENRICO BOCCA, *SIMULATION TEAM, ITALY*

FRANCESCO LONGO, *MSC-LES, UNIVERSITY OF CALABRIA, ITALY*

FRANCESCA MADEO, *UNIVERSITY OF GENOA, ITALY*

MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*

LETIZIA NICOLETTI, *CAL-TEK, ITALY*

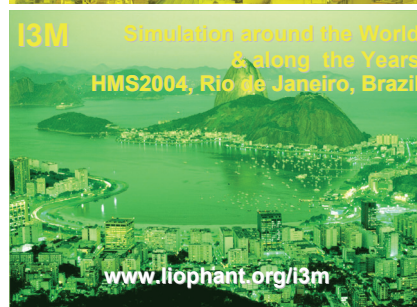
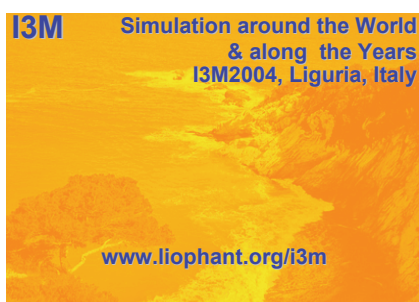
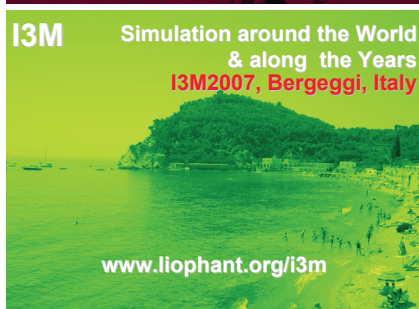
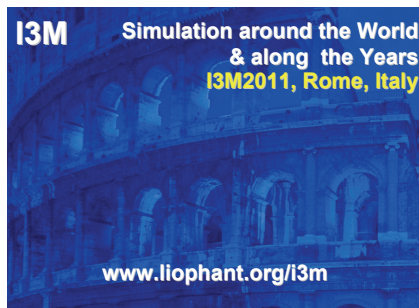
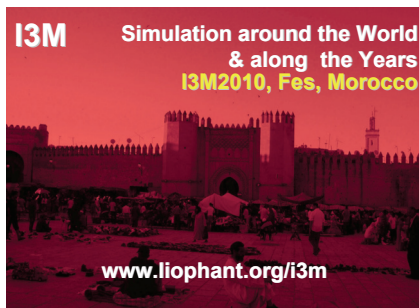
FEDERICO TARONE, *SIMULATION TEAM, ITALY*

ALBERTO TREMORI, *SIMULATION TEAM, ITALY*





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





## **Index**

Part I: Mothballing (Asset Preservation) of a Crude Oil Gathering Center <i>Musleh Al-Otaibi, Abdul-Hassan Mohammed</i>	1
An efficient hybrid approach based on SVM and Binary ACO for feature Selection <i>O. Kadri, L. H. Mouss, F. Merah, A. Abdelhadi, M. D. Mouss</i>	8
Study for a Mechanism Aided by Asynchronous Actuator Powered by Asynchronous Diesel Generator <i>Meglouli H., Ikhlef B.</i>	16
Mathematical Modeling and Simulation for Detection of Suicide Bombers <i>William Fox, John Vesecky, Kenneth Laws</i>	22
Comparing Decision Support Tools for Cargo Screening Processes <i>Peer-Olaf Siebers, Galina Sherman, Uwe Aickelin, David Menachof</i>	31
Improvement of the Forecast of Economic Processes Parameters <i>Yuri Menshikov</i>	40
Solving Fully Fuzzy Linear Programming Problem with Equality and Inequality Constraints <i>Ahmad Jafarnejad, Mahnaz Hosseinzadeh, Hamed Mohammadi Kangarani</i>	45
Restoring Aquatic Ecosystems on the Basis of the GMAA DSS <i>Antonio Jiménez, Alfonso Mateos</i>	50
Planning Highways Resurfacing Using Computer Simulation <i>Mohamed Marzouk, Marwa Fouad, Moheeb El-Said</i>	56
3 Years of Funding Initiative for Computational Mathematics in Austria <i>Peter Kersch, Alexander Pogány</i>	61
Dynamic Phenomena and Quality Defects in Laser Cutting <i>Dieter Schuöcker, Joachim Aichinger, Georg Schuöcker</i>	65
Analysis of a Warehouse Management System by Means of Simulation Experiments <i>Giuseppe Aiello, Mario Enea, Cinzia Muriana</i>	70
CoMMoDO - Complex Material Modeling Operations, A Comprehensive Approach to the Modeling of Complex Materials with Machine Learning Models within Finite Element Simulations <i>Andreas Kuhn, Toni Palau, Gerolf Schlager, Helmut Böhm, Sergio Nogales, Victor Oancea, Ritwick Roy, Andrea Rauh, Jürgen Lescheticky</i>	80
Multi-Objective Optimization in Urban Design <i>Michele Bruno, Kerri Henderson, Hong Min Kim</i>	90
Additive Fault Tolerant Control for Delayed System <i>Abdelkrim Nouceyba, Tellili Adel, Abdelkrim Mohamed Naceur,</i>	96
Centralized and Decentralized Adaptive Fault-Tolerant Control applied to Interconnected and Networked Control System <i>Amina Challouf, Adel Tellili, Christophe Aubrun, Mohammednaceur Abdelkrim</i>	101
Modelling NONI (Morinda Citrifolia L.) Pulp Spray Drying Using a Crossed Mixture-Process Design and Feed Optimization <i>Eliosbel Márquez González, Débora Castro Espín, Margarita Núñez de Villavicencio, José Luis Rodríguez</i>	107
Factory layout Planner: how to speed up the factory design process in a natural and comfortable way <i>Antonio Avai, Giovanni Dal Maso, Paolo Pedrazzoli, Diego Rovere</i>	113

Release Kinetics of Cardamom Oil from Microcapsules Prepared by Spray and Freeze Drying <i>Masoud Najaf Najafi, Rassoul Kadkhodae</i>	120
Genetic Algorithm approach to modelling fractal manufacturing layout <i>Julian Aririguzo, Sameh Saad,</i>	126
Mechanical Damage to Navy Beans as Affected by Moisture Content, Impact Velocity and Seed Orientation <i>Feizollah Shahbazi</i>	136
How important is price leadership in the UK fresh fruit and vegetable market? <i>Cesar Revoredo-Giha, Alan Renwick</i>	139
Simulations of liquid film flows with free surface on rotating silicon wafers (RoWaFlowSim) <i>Markus Junk, Frank Holsteyns, Felix Staudegger, Christiane Lechner, Hendrik Kuhlmann, Doris Prieling, Helfried Steiner, Bernhard Gschaider, Petr Vita</i>	149
PLC Code Processing for Automatic Simulation Model Generation <i>Gergely Popovics, András Pfeiffer, Botond Kádár, Zoltán Vén, László Monostori</i>	155
Modelling and Analysing the Impact of Manual Handling Processes in Seaports <i>Bernd Scholz-Reiter, Michael Görges, Ralf Matthies</i>	160
Numerical Analysis of Wind Induced Pressure Loads on an Integrated Roof-Based Photovoltaic System <i>Marco Raciti Castelli, Sergio Toniato, Ernesto Benini</i>	166
Modeling of Drying Process of Candies Obtained with Starch Molding Technique <i>Noemi Baldino, Lucia Seta, Francesca Romana Lupi, Domenico Gabriele, Bruno de Cindio</i>	173
Benchmarking Real-World Job-Shop Scheduling using a Genetic Algorithm with a Simulation Approach <i>Peter Steininger</i>	179
Improving Patient Safety: A Model of Drift in Health Care Teams <i>R. Severinghaus, C. Donald Combs</i>	185
Modelling of Gestures with Differing Execution Speeds: Are Hidden non-Markovian Models Applicable for Gesture Recognition? <i>Sascha Bosse, Claudia Krull, Graham Horton</i>	189
Modelling and interpolation of spatial temperature during food transportation and storage by the Variogram <i>Reiner Jedermann, Javier Palafox-Albarrán, Pilar Barreiro, Luis Ruiz-García, Jose Ignacio Robla, Walter Lang</i>	195
Modeling and Simulation of an assembly line: a new approach for assignment and optimization of activities of operators <i>Domenico Falcone, Alessandro Silvestri, Antonio Forcina, Antonio Pacitto</i>	202
Dynamic Spectrum Management with Minimizing Power Allocation <i>Tawiwat Veeraklaew, Settamong Malisuwan</i>	207
Large Target-Weapon Assignment Problems with Evolutionary Programming as applied to the Royal Thai Armed Forces ground-to-ground rocket simulation system <i>Settamong Malisuwan, Teeranan Nandhakwang, Tawiwat Veeraklaew</i>	211
Data Presentation and Transformation in Train Schedule Information Systems <i>Eugene Kopytov, Vasilij Demidovs, Natalia Petukhova</i>	216
Dynamic Simulation Based Design of Powered Support in Integrated	226

<b>Mechanized Coal Mining</b> <i>Zhifeng Dong, Guozhu Liu, Jian Wang, Jiangang Qian, Shouxiang Zhang, Dongsheng Bao</i>	
<b>Impact of Occupants Behavior on Building Energy Use: an Agent-Based Modeling Approach</b> <i>Elie Azar , Carol Menassa</i>	232
<b>Sheeting Process Modelling and Rheological Analysis of an Olive-Oil-Emulsion-Based Puff Pastry</b> <i>Francesca Romana Lupi, Noemi Baldino, Lucia Seta, Domenico Gabriele, Bruno de Cindio</i>	242
<b>New Evolutionary Algorithm Based on Particle Swarm Optimization and Adaptive Plan System With Genetic Algorithm</b> <i>Pham Ngoc Hieu, Hiroshi Hasegawa</i>	249
<b>The Art of Data-Driven Modelling in Logistics Simulation</b> <i>Rainer Frick</i>	255
<b>Evaluation of Climate-Efficient Cross-Company Logistics Models based on Discrete Event Simulation</b> <i>Christian Hillbrand, Susanne Schmid</i>	259
<b>Building a System for Automated Modeling and Simulation of Plants</b> <i>Robert Schöech, Susanne Schmid, Christian Hillbrand</i>	266
<b>Dynamics of direct transports in autonomously controlled production networks</b> <i>Bernd Scholz-Reiter, Michael Görges</i>	273
<b>Advanced Design of a Static Dryer for Pasta with Simulation Tools</b> <i>Mattia Armenzoni, Federico Solari, Davide Marchini, Roberto Montanari, Gino Ferretti</i>	281
<b>Bakery Production Scheduling Optimization -a PSO Approach</b> <i>Florian Hecker, Walid Hussein, Mohamed Hussein, Thomas Becker</i>	289
<b>Simulation Based Design Optimization Framework for a Gait Pattern Generation of a small Biped Robot with Tiptoe Mechanism</b> <i>Krissana Nerakae, Hasegawa Hiroshi</i>	295
<b>Rfid-Based Real-Time Decision Support in Supply Chains</b> <i>Tobias Hegmanns, Michael Toth</i>	303
<b>Feature Selection for Datasets of Wine Fermentations</b> <i>Antonio Mucherino, Alejandra Urtubia</i>	309
<b>Demand-Supply Interaction and Inventory Buildup Strategies for Short Life Cycle Products</b> <i>Rong Pan, Adriano Solis, Bixler Paul</i>	314
<b>Evaluation of the Performance of Solar Air Collector by Using Bond Graph Approach</b> <i>Hatem Oueslati, Salah Ben Mabrouk, Abdelkader Mami</i>	322
<b>The Use of Simulation as a Pedagogical Tool in Construction Education</b> <i>Ronald Ekyalimpa, Jangmi Hong, Simaan AbouRizk</i>	330
<b>Logical Operator Usage in Structural Modelling</b> <i>Ieva Zeltmate</i>	338
<b>A Survey on Performance Modelling and Simulation of Sap Enterprise Resource Planning Systems</b> <i>Manuel Mayer, Stephan Gradl, Veronika Schreiber, Harald Kienegger, Holger Wittges, Helmut Krcmar</i>	347
<b>Understanding the Performance Behavior of a Sap Erp System for the Use of</b>	353

<b>Queuing Models</b>	
<i>Stephan Gradl, Manuel Mayer, Alexandru Danciu, Holger Wittges, Helmut Krcmar</i>	
<b>Object-Oriented Model of Fixed-Bed Drying of Coffee Berries</b>	361
<i>Emilio de Souza Santos, Paulo Cesar Corrêa, Brian Lynn Steward, Daniel Marçal de Queiroz</i>	
<b>Scheduling Patients Based on Provider's Availability</b>	370
<i>Jose Sepulveda, Waldemar Karwowski, Francisco Ramis, Pablo Concha</i>	
<b>An Overall DHM-Based Ergonomic and Operational Assessment of a Manufacturing Task: a Case Study</b>	375
<i>Nadia Rego Monteil, David del Rio Vilas, Diego Crespo Pereira, Rosa Rios Prado</i>	
<b>A Computationally Efficient and Scalable Shelf Life Estimation Model for Wireless Temperature Sensors in the Supply Chain</b>	383
<i>Ismail Uysal, Jean-Pierre Emond, Gisele Bennett</i>	
<b>Simulation as a Decision Support Tool in Maintenance Float Systems - System Availability Versus Total Maintenance Cost</b>	389
<i>Francisco Peito, Guilherme Pereira, Armando Leitão, Luís Dias</i>	
<b>ANP Approach For Improving Public Participation In Strategic Environmental Management Planning</b>	397
<i>Fabio De Felice, Antonella Petrillo</i>	
<b>Distributed M&amp;S for Product Service System</b>	406
<i>Thecle Alix, Greg Zacharewicz</i>	
<b>Design and Development of Realistic Food Models with Well Characterised Micro and Macro- Structure and Composition</b>	412
<i>Monique Axelos, Jean-Dominique Daudin, Guy DellaValle, Nathalie Perrot, Catherine MGC Renard, Caroline Sautot, Jean-Louis Sebedio</i>	
<b>From Expert Knowledge to Qualitative Functions: Application to The Mixing Process</b>	419
<i>Kamal Kansou, Guy Della Valle, Amadou Ndiaye</i>	
<b>Simulation Based Modeling of Warehousing Operations in Engineering Education Based on an Axiomatic Design</b>	427
<i>Rafael S. Gutierrez, Sergio Flores, Fernando Tovia, Olga Valerio, Mariano Olmos</i>	
<b>A Multi-Item Multi-Rack Approach for Designing LIFO Storage Systems: a Case Study from the Food Industry</b>	437
<i>Elisa Gebennini, Andrea Grassi, Bianca Rimini, Rita Gamberini</i>	
<b>A Process Mining Approach Based on Trace Alignment Information for Very Large Sequential Processes with Duplicated Activities</b>	442
<i>Pamela Viale, Claudia Frydman, Jacques Pinaton</i>	
<b>Modeling Psychological Messages and Their Propagation</b>	452
<i>Colette Faucher</i>	
<b>NAV - The Advanced Visualization Station: a Mobile Computing Center for Engineering Project Support</b>	464
<i>Gabriel A. Fernandes, Gerson G. Cunha, Tiago Mota, Celia Lopes</i>	
<b>Competition and Information: Cumana a Web Serious Game for Education in the Industrial World</b>	471
<i>Marina Massei, Alberto Tremori, Alberto Pessina, Federico Tarone</i>	
<b>Modeling Logistics &amp; Operations in a Port Terminal for Environmental Impact Evaluation and Analysis</b>	480
<i>Marina Massei, Francesca Madeo, Francesco Longo</i>	
<b>Modelling a Traceability System for a Food Supply Chain: Standards,</b>	488

Technologies and Software Tools	
<i>Bruno De Cindio, Francesco Longo, Giovanni Mirabelli, Teresa Pizzuti</i>	
Command Feedback and Response (CFR) - the Evolution of Command and Control in an Immersive and Interactive Environment (C2I2)	495
<i>Marco Biagini, Michele Turi</i>	
Modelling and Simulation of the Sterilization Process Of Pouch Packaging In an Aseptic Line	503
<i>Paolo Casoli, Gabriele Copelli, Michele Manfredi, Giuseppe Vignali</i>	
Authors' Index	510



## Part-1: Mothballing (Asset Preservation) of a Crude Oil Gathering Center

Dr. M.A. Al-Otaibi<sup>a</sup>, A.H.M. Abdul-Hassan<sup>b</sup>

<sup>(a)</sup>Affiliation: Visiting lecturer at the Applied Technology Institution, Kuwait.

<sup>(b)</sup>Affiliation: MIChemE (UK), EuroIng.

<sup>(a)</sup> [MAOtaibi@kockw.com](mailto:MAOtaibi@kockw.com)

<sup>(b)</sup> [EmailAAbdulHassan@kockw.com](mailto:EmailAAbdulHassan@kockw.com) & [alhassan@aol.com](mailto:alhassan@aol.com)

### ABSTRACT

Kuwait Oil Company (KOC) produces Crude Oil from various fields and adjusts the production rates in order to meet OPEC's quota.

Due to expected weak Crude Oil demand worldwide following the global economic crisis in 2008, KOC is considering equipment Preservation by Mothballing one or more GGs in S&EK assets in such a way that the GC would be readily available to be put back in operation when required.

The success or failure of mothballing the facilities depends on a number of factors such as:

- Shutdown duration, preparation and removing all process materials.
- Responsibility allocation for proper implementation of the Mothballing Procedure.
- Materials and equipment used during the preservation.
- Maintaining & recording the preservation conditions during the idle period.
- Personnel expertise in mothballing materials proper implementation.
- Safety programs and active routine audits..

Keywords: KOC, Asset Preservation, Mothballing Procedure, Oil & Gas Facilities.

### 1. INTRODUCTION

KOC produces crude oil from various fields in Kuwait which are divided into three different Assets as follows:

- North Kuwait (NK)
- West Kuwait (WK)
- South & East Kuwait (S&EK).

In order to meet the market demand internally and also adhere to OPEC's crude production quota, crude oil production from NK and WK is normally kept at maximum rates allowable from each Oil Producing Well, while production from Wells feeding the fourteen GCs in S&EK varies since S&EK is considered as the

Swing Producer adjusting production upwards or downwards, as needed.

S&EK is dominated by the Greater Burgan Fields. Wells effluents are directed via dedicated flow lines, and then commingled into a number of dedicated headers prior to feeding the fourteen Gathering Centers (GCs) for phases' separation and crude oil treatment to the required specification.

The treated crude oil to specification from each GC is pumped by a Booster pump then by a Main Oil Line pump via a dedicated transit pipeline to the Export Oil Tank Farms.

Due to decreasing oil demand worldwide following the global economic crisis in 2008, and in order to avoid operating one or more GCs at reduced capacity (inefficient operation), or shutting down GC(s) with no preservation for an extended period of time, KOC is considering mothballing of one or more GGs in S&EK assets in such a way that the mothballed GC(s) would be readily available to be put back in operation when required. Mothballing helps in preserving facilities without reduction in useful operating life and keeps them in "ready for start-up" conditions.

A Taskforce Team (TFT) comprising specialists from various groups within KOC was formed in September 2010 for the purpose of defining the following:

- The steps necessary for identifying the GC(s) to be mothballed.
- The Capacity to be mothballed.
- The duration of mothballing.
- The Mothballing Procedures to be implemented.
- Lessons learned from other organizations related to mothballing.

There are different mothballing procedures for various facilities preservation periods since the preservation procedures differ for periods up to 18 months and for periods exceeding 18 months up to five years.

Equipment and piping in a process industry are subjected to faster deterioration during periods of idleness. Carbon steel equipment can suffer general and/or pitting type corrosion due to the presence of moisture, oxygen, stagnant pockets of corrosive process fluids, deposits of scale, etc.

The objective of mothballing is to ensure that the equipment including piping and fittings are preserved in good conditions such that it will not be deteriorate due to internal and external corrosion and that the GC(s) would be readily available for commissioning and start-up within a short period (weeks) of de-mothballing.

This write up covers Part-1 of Mothballing related to the steps we have taken and our approach for the practical and successful implementation of mothballing procedure incorporating lessons learned from previous application of similar procedures, and what do we really need to safeguard our assets from deterioration during the idle time.

Part-2 shall be presented sometime in year 2012/2013 after the GC is put back in operations and shall cover our statistical analysis of our findings during the mothballing idle period, lessons learned including implications of our findings for any problem identified, and our conclusion.

## **2. BRIEF DESCRIPTION**

### **2.1. Definition**

Mothballing is the preservation of production facilities for a specified period of time and kept in good working order in idle conditions without production so that when needed, the facilities can be put back in production within a short period of de-mothballing.

### **2.2. Mothballing Objectives**

- To preserve equipment without any considerable loss of useful operating life and without excessive costs.
- To maintain the preservation status with a minimum of maintenance activities.
- To be able to activate and re-commission the equipment / plant at a minimum cost and delay.

### **2.3. Mothballing Strategy**

The Mothballing Taskforce Team (TFT) has studied the three alternative Mothballing Execution Routes as shown below.

- In house
- Consultant
- Turnkey

The execution of Mothballing and subsequent monitoring and inspection depends on the strategy Route selected. In all cases, HSE shall be the paramount factor.

### **2.4. Mothballing Duration**

The preservation process (different corrosion control method and procedures), and the associated costs are influenced by the idle period duration in addition to risk insurance company requirements.

In general, the idle period duration may be classified into three periods as shown below:.

- 1 month to 2 years.
- Between 2-5 years
- More than 5 years.

For storage tanks the short term is 6 to 12 months.

### **2.5. Mothballing Procedures**

In order to preserve the facilities for an extended period of time successfully and without being deteriorated mainly due to corrosion, it is imperative that the necessary approved preservation procedures are adhered to without any deviation. These procedures are divided into three categories:

- Internally (KOC) developed procedures.
- Equipment's Vendors' Procedures.
- Third Party (Consultants/Contractors) Procedures.
- Internationally Accredited Procedures.

### **2.6. Identification Process for selection a GC.**

The TFT members have discussed a number of criteria that deemed to be relevant and lead to identifying the GC(s) that could be mothballed and excluding other GC(s) from Mothballing. These criteria included the topics listed below. Section 3.6 below shows the flowcharts for identifying a GC for mothballing.

## **3. DETAILED DESCRIPTION**

A global decline in crude oil demand is anticipated and that would make it difficult to sustain the current level of production. Cut in production quotas are expected. Anticipating future outlook for oil demand, it would be prudent to develop and embark on a strategy that is best result oriented with regard to the production capacity utilization of KOC facilities.

When a large production cut-back is required for a prolonged period of time, the current operating practice is to shut down an entire GC(s) without mothballing. or operating a number of GC(s) at reduced capacities. This practice is detrimental to plan and equipment integrity mainly due to corrosion attack such as:

- Internal corrosion
- Pitting corrosion
- Galvanic Corrosion
- Microbiological Corrosion

Internationally acceptable practice for a prolonged production shutdown is to mothball and preserve the facilities by employing suitable preservation method depending on the shutdown duration.



KOC has no the experience regarding mothballing practice and therefore, the management has decided to set up a specific Taskforce Team responsible for:

- Establishing the necessary mothballing procedures.
- Evaluating the available mothballing technologies.
- Short listing the execution strategies.
- Identify the GC(s) to be mothballed.

### 3.1. TFT Members:

The TFT comprised Core Members and Supporting Members.

Core Members:

- Production Operation (Process)
- Inspection and Corrosion (I&C)
- Maintenance, Support and Reliability (M&SR)
- Operational Technical Services (OTS)

The two Authors mentioned above were also part of the Core members. The TFT team was led by Team Leader Production Operations (EK-I).

Supporting members:

- Field Development (FD)
- Health, Safety, & Environment (HSE)
- Standards
- Planning

### 3.2. Typical Equipment in a GC

The fourteen GCs are not all identical in terms of equipment's size, number and type but typically, each GC has the following systems:

- Main Plant comprising 2-phase & 3-phase separators and knock-out drums.
- Desalting and Dehydration Trains
- Condensate Recovery Unit including Tank Vapour Compression.
- Effluent Water Treatment.
- Crude Oil & Effluent Water Tanks
- Crude Oil Booster and Main Oil Line Pumps.
- Utilities and Offsite facilities including Flares.
- Piping, fittings, and Controls.
- Main Control Room.

### 3.3. Available Technology

The preservation materials listed below are identified for external & internal protection:

- Inert gases such as Nitrogen and dry air.
- Proprietary Vapour Corrosion Inhibitors (VCIs)
- Desiccants
- Solvents and Emulsions
- Hydrocarbons such as diesel, kerosene and treated light crude oil.
- Greases and Oils

- Coatings (paints, varnishes
- Wrapping and insulation.

The selection of the appropriate preservative is influenced by the following considerations:

- Climatic Conditions
- Material of Construction and product handled
- Duration of preservation
- Disadvantages, e.g. flammable, sensitive to air, etc.
- Cost of mothballing vs. cost of equipment
- Available/Approved budget.
- Re-commissioning duration after period of idleness.

### 3.4. Mothballing Procedures

KOC developed in-house preservation procedures covering stationary mechanical equipment, pipelines, and rotating equipment. The preparation of these procedures was based on international standards and published literature and also on vendors' guidelines. Two alternative periods are considered in the internal procedures developed by KOC relevant to 1-18 months and long term 19-60 months' preservation periods. These procedures also cover re-commissioning (pre-commissioning) procedure.

### 3.5. Mothballing Strategy (KOC)

The alternative strategies (options) addressed by the TFT for mothballing execution are shown below. The major area of concern irrespective of the selected option is the duration taken for the preparation and final approval of the Contract documents within KOC.

#### A. Option: In-house

This option leads to KOC personnel full involvement, responsibility, and accountability to the high management.

Although this option could be implemented in shorter time and cheaper than the other two options B & C shown below, there is an element of risk, as this is the first time such activities are undertaken by KOC. On the other hand, KOC will gain direct experience in mothballing application.

The execution of work shall be through an existing maintenance contract under KOC supervision.

- In-house development and approval of the necessary mothballing procedures.
- Quantifying all equipment and components such as instruments, electrical & mechanical equipment, tanks, vessels, piping, junction boxes, enclosures, cabinets, panels, flanges, hinges, etc.
- Developing detailed scope of work including de-mothballing and commissioning.
- Identifying and selecting the applicable preservation materials for the various equipment

and components in the GCs based on the specified mothballing duration.

- Estimating the quantities of the preservation material necessary.
- Prequalify, select, and Supervise suitable local contractors for job execution.

**B. Option: Consultant**

KOC prepare Consultant's Scope of Work, identifies an experienced Consultancy Company responsible for the activities listed below:

The appointed Consultancy would be experienced in the mothballing field, risk is averted and KOC involvement shall be minimum and limited to coordination with the Consultancy Company.

- Preparation of Mothballing Scope of Work.
- Mothballing Procedures shall either be KOC in-house developed, Contractors' procedures or internationally accredited procedures.
- Screening and identifying suitable Contractors for the supply and application of suitable mothballing materials.
- Supervising the mothballing execution and monitoring including equipment inspections during the mothballing period.

**C. Option: Turnkey**

This option is identical to Option "A" above but with less KOC involvement in the tasks' execution and the accountability is shifted to the Main Contractor..

KOC prepares the Scope of Work, identify and select an experienced Single-Accountability Contractor who would be responsible for:

- Selecting, supply and application of all the necessary preservative materials and tools in accordance with approved procedures (KOC or external).
- Supply the necessary supervisory personnel.
- Monitoring and inspection of the mothballed equipment during the mothballing period.

KOC shall provide the necessary labours working under the Contractor's direct supervision.

The flowchart in section 3.10 below shows the Options related to the Mothballing studies considered by the TFT members.

**3.6. Identification Process for selecting a GC:**

The flowcharts in section 3.10 below show the five steps briefly described below that KOC TFT considered necessary for identifying the GC(s) than can be mothballed.

**Step 1: Facility HSE Concerns covering:**

- Gas Flaring issues

- Effluent Disposal issues

**Step 2: External Factors/Directives covering:**

- GC's location in Strategic Area
- Bordering Inhibited Area
- Other Strategic Factors

• **Step 3: Reservoir Management covering:**

- GCs with ongoing Projects
- Wells with ESP feeding GCs
- Dry/Wet/Critical Wells production

• **Step 4: Operational Factors covering:**

- Equipment Constraints
- Underutilization Capacity
- OPEX/BBL
- Reallocation of resources
- Ease of rerouting production to another GC.

• **Step 5: Cost Estimate shall be based on:**

- Duration of Mothballing
- Mothballing Strategy (Option)
- Scope of Work

In addition to the mothballing cost and its associated monitoring/maintenance cost, the cost estimate shall also take into account the restart cost to achieve full production.

**3.7. Visit to ADMA-OPCO**

In December 2010, selected TFT members visited ADMA-OPCO offices in Abu Dhabi for the purpose of finding out the mothballing executing process and lessons learned.

The purpose of the visit was to gain experience from ADMA on how did they manage to maintain the equipment idle for such a long period (about 25 years) without deterioration and negative impact on equipment's integrity.

Apparently, a number of equipment has been replaced due to production enhancement reasons and not due to deterioration.

Summary of ADMA's mothballing activities:

- Utilized in-house generated procedures.
- Dedicated TFT followed-up Contractor's activities.
- Subcontracted the Mothballing activities to a third party.
- Nitrogen at 15 psig was used for equipment blanketing.
- Rotating equipment mothballing was as per vendors' procedures, e.g. Pumps have been removed mechanical seals and impellers flashed-off and kept under nitrogen pressure.
- Motors: Wrapped in plastic.
- All vessels' internals have been removed.
- All valves (except PSV's) and instrumentations have been removed, tagged, and stored in a warehouse.
- Small bore piping: filled with inhibited water.
- Pipelines: filled with treated crude.

- ADMA has carried out Lab tests and found that an alternative media, Vaporized Chemical Inhibitors (VCI), is very effective in preserving equipment and has recently applied VCI on a leg in a Platform

Based on the visit to ADMA-OPCO, VCI preservation media seems to be an attractive choice to be considered.

### 3.8. Visit to ARAMCO

A team visited ARAMCO in May 2011 and toured Gas/Oil Separation Plant (GOSP), which was mothballed since November 2010 but it can be ready for re-commissioning within a two weeks.

ARAMCO employed Consultants to define the Criteria for selecting a GOSP for mothballing and to produce the mothballing procedure.

Preservation:

- Nitrogen was used for mothballing all stationary mechanical equipment and piping.
- Diesel oil for the rotating equipment.
- The shafts were stored vertically in a warehouse without any preservation in and without air conditioning.
- The crude transit pipelines were kept full of dry crude oil ( $^{\circ}$ API about 36).
- Power was kept on to provide electricity for the site and for air conditioning.

### 3.9. Recommended Mothballing Steps

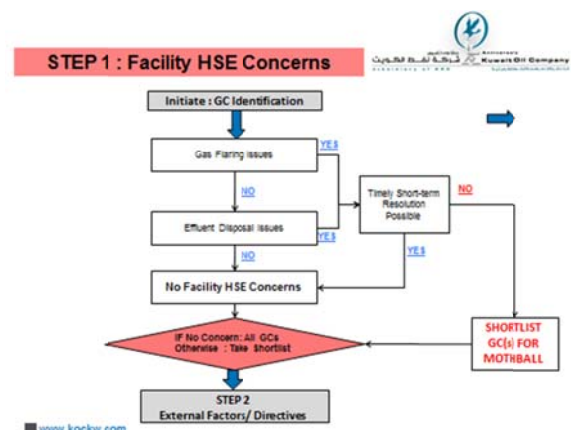
- The plan shall clearly state the length of the intended preservation time span and the required re-commissioning period.
- Plan for adequate funding and manpower are provided throughout the life of the mothball process to maintain preservation effectiveness and equipment readiness.
- For an effective and successful mothballing, the following steps are recommended:
  - a. Establish a Task Force Team comprising personnel from various relevant disciplines.
  - b. Identify facilities to be mothballed and re-commissioning concerns.
  - c. Identify Options and Alternatives.
  - d. Decide on the mothballing procedures (internally developed or otherwise).

- e. Identify the execution methodology (internally managed or contracted out).
- f. Establish Mothballing cost including routine Inspection and Maintenance cost during the idle period.
- g. Establish cost benefits.
- h. Seek Management approval to proceed.

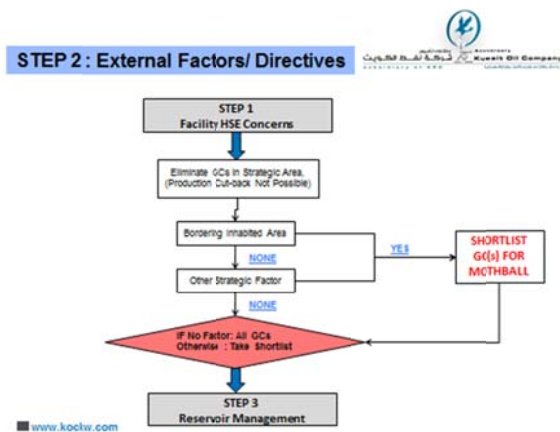
### 3.10. Figures



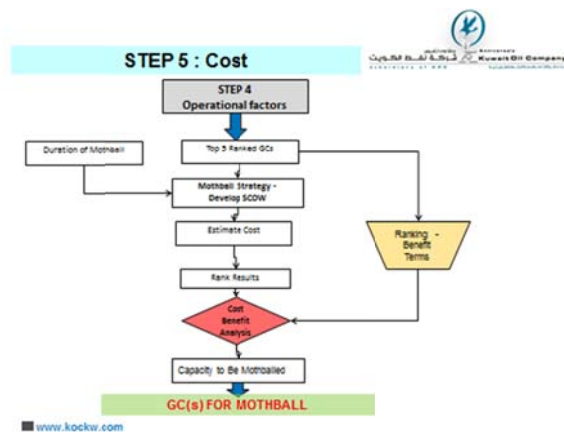
Mothballing Strategy (Options)



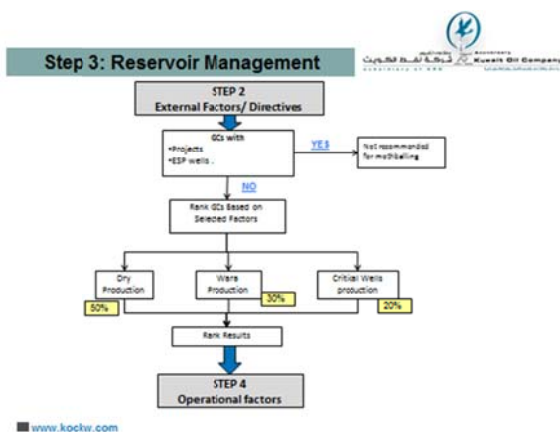
Step 1: Facility HSE Concerns



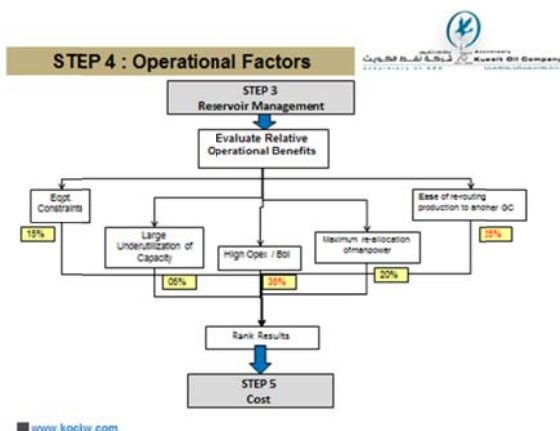
Step 2: External Factors/Directives.



Step 5: Cost



Step 3: Reservoir Management



Step 4: Operational Factors

### 3.11. Lessons Learned from KOC visits

- Irrespective of the idle period, severe corrosion can occur under some circumstances, e.g. ammonium or amine chloride deposits on equipment can be very corrosive if equipment is opened to atmosphere.
- For successful equipment's preservation during idle time, a well-developed mothballing Plan should be executed in a timely fashion and closely monitored.

### REFERENCES

- API Guide for Inspection of Refinery Equipment, Chapter XVIII: Protection of Idle Equipment, 1982.
- MTI publication No. 34: Guidelines for the Mothballing of Process Plants, Third Edition.
- NACE RP0170: Protection of Austenitic Stainless Steels and Other Austenitic Alloys from Polythionic Acid Stress Corrosion Cracking During Shutdown of Refinery Equipment.
- NACE RP 0487: Consideration in the selection and evaluation of Rust Preventive s and Vapour Corrosion Inhibitors for Interim (Temporary) Corrosion Protection.

### AUTHORS BIOGRAPHY

**Dr. Musleh Baddah Al-Otaibi:**

PhD. Chemical Engineering, Loughborough University, UK, 2004.

Master Degree Chemical Engineering, Kuwait University, 1999.

BSc. Petroleum and Natural Gas Engineering, Penn State University, PA, the USA, 1993.

Memberships:

- Member of the Society of Petroleum Engineers (SPE).
- Member of Kuwait Society of Engineers.
- Member of the Institution of Chemical Engineers (IChemE), the USA.

- Chartered Arbitrator and Expert in Commercial Companies Laws and suits; the Arab World of Arbitrators, Cairo, Egypt.

Dr. Al-Otaibi has over 17 years' experience in the process plant design and engineering for onshore facilities Gas, Oil, Water Treatment and Injection. He has a vast exposure over commissioning & start-up of crude oil and Gas gathering centers. His knowledge to the Mothballing practices in the Gulf region has put him on heading Kuwait Oil Company's Mothballing Standards Team. He has been in the circle of middle management for over 10 years exposed to executive decisions and participating effectively in many strategic visions that helped shaping the Oil Sector future plans. He is still acting Manager Operations in East Kuwait fields.

#### **Abdul-Hassan Mohammed Abdul-Hassan:**

B.Sc (Hons.) Chemical Engineering, UK, 1974.

Memberships:

- Member of the Institution of Chemical Engineers (IChemE), UK,.
- Chartered Engineer (C.Eng.), UK.
- European Engineer (Euro. Ing).

Mr. Abdul-Hassan has over thirty years' experience in the process plant design and engineering for both onshore and offshore facilities. He has experience in Gas, Oil, Water Treatment and Injection, Refineries and Petrochemicals fields including commissioning & start-up of refineries and LNG trains. His worldwide experience included working in thirteen countries in Asia, Africa, the Middle East, Australia, and Europe for both EPCC (Engineering, Procurement, Construction, & Commissioning) companies and also for National and International Operating Companies. He was a member of the Project Management Team on a number of onshore and offshore projects.

# An efficient hybrid approach based on SVM and Binary ACO for feature Selection

O. KADRI\*, L.H. MOUSS\*, F. MERAH\*\*, A. ABDELHADI\* & M. D. MOUSS\*

*\*Laboratory Automation & Production Engineering*

*University of Batna*

*1, Rue Chahid Boukhrouf 05000 Batna, ALGÉRIE*

*\*\*Department of Mathematics*

*University of Khenchela*

*Route de Batna BP:1252, El Houria, 40004 Khenchela ALGÉRIE*

*ouahabk@yahoo.fr, merahfateh@yahoo.fr,*

*hayet\_mouss@yahoo.fr, abdelhadi.adel@yahoo.fr & djmouss@yahoo.fr*

## Abstract

*One of the most important techniques in data pre-processing for classification is feature selection. In this paper, we propose a novel hybrid algorithm for feature selection based on a binary ant colony and SVM. The final subset selection is attained through the elimination of the features that produce noise or, are strictly correlated with other already selected features. Our algorithm can improve classification accuracy with a small and appropriate feature subset. Proposed algorithm is easily implemented and because of use of a simple filter in that, its computational complexity is very low. The performance of the proposed algorithm is evaluated through a real Rotary Cement kiln dataset. The results show that our algorithm outperforms existing algorithms.*

## Keywords

*Binary Ant Colony algorithm, Support Vector Machine, feature selection, classification.*

## 1. Introduction

Our work falls under the Condition monitoring and diagnosis of industrial system which is an important field of engineering study (in our case is a Rotary Cement kiln, see fig. 1). In substance, condition monitoring is a classification problem [12]. The principal function of the condition monitoring is to check the operating condition of the system. It is made up of two parts which are detection and the diagnosis.

The phase of detection makes it possible to determine the state of the system as being normal or abnormal. The phase of diagnosis consists in identifying the failing components and to find the causes starting from a whole of symptoms observed [7, 10, 12].

An industrial system is described by a vector of numeric or nominal features. Some of these features may be irrelevant or redundant. Avoiding irrelevant or redundant features is important because they may have a negative effect on the accuracy of the classifier [7,10]. In addition, by using fewer features we may reduce the cost of acquiring the data and improve the comprehensibility of the classification model (fig. 2).



Fig. 1 Rotary Cement kiln

Feature extraction and subset selection are some frequently used techniques in data pre-processing. Feature extraction is a process that extracts a set of new features from the original features through some functional mapping [15]. Subset selection is different from feature extraction in that no new features will be

generated, but only a subset of original features is selected and feature space is reduced [14].

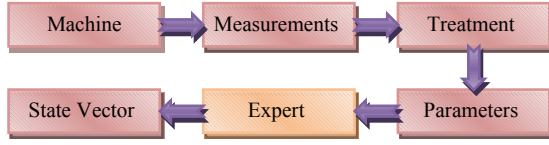


Fig. 2 Construction of the set of features

The idea behind the selection approach is very simple and is shown in Fig. 3. Any method of selection of features consists of four essential points:

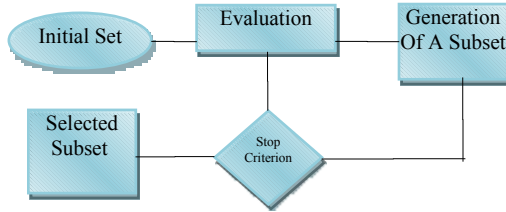


Fig. 3 Subset Selection Method

A starting subset, which represents the subset of features, initially is used by a procedure of research. This set can be empty, or contains all the features or a random subset.

The procedure of research is the essential element of any method of selection. It turns over as result the subset of features which answer the quality standard better. This criterion is turned over by a function of evaluation. This function determines the quality of classification obtained by using a subset of feature. A criterion of stop is used to finish the procedure of research. This criterion depends to the evaluation function or with the parameters of configuration which are defined by the user [1].

We present in this paper a hybrid approach based on ant colony optimization (ACO) and support vector machine (SVM) for feature selection problems using datasets from the field of industrial diagnosis. This paper presents a novel approach for heuristic value calculation, which will reduce the set of available features.

The rest of this paper is organized as follows. In section 2, different methods for feature selection problems are presented. An introduction on ACO applications in feature selection problems is discussed in Section 3. A brief introduction of SVM is presented in Section 4. In Sections 5 and 6, the proposed algorithm is discussed, followed by a discussion on the experimental setup, datasets used and the results.

## 2. Feature subset selection

Feature selection is included in discrete optimization problems. The whole search space for optimization contains all possible subsets of features, meaning that its size is  $2^n$  where  $n$  is the dimensionality (the number of features). Usually FS algorithms involve heuristic or random search strategies in an attempt to avoid this prohibitive complexity. However, the degree of optimality of the final feature subset is often reduced [2, 4, 8].

Two broad categories of optimal feature subset selection have been proposed based on whether feature selection is performed independently of the learning algorithm that constructs the classifier. They are the filter approach and the wrapper approach [3, 22]. The filter approach initially selects important features and then the classifier is used for classification while the wrapper uses the intended learning algorithm itself to evaluate the usefulness of features [13]. The two famous algorithms of this category are Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) [1, 3]. In Sequential forward selection, the features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion but in Sequential backward selection the features are sequentially removed from a full candidate set until the removal of further features increase the criterion. In our work, we use a hybrid wrapper/filter approach aiming to explore the qualities of both strategies and try to overcome some of their deficiencies [22].

The criterion of stop represents the dimension of the vector obtained by the algorithm where the quality standard does not evolve/move if we add another feature [9].

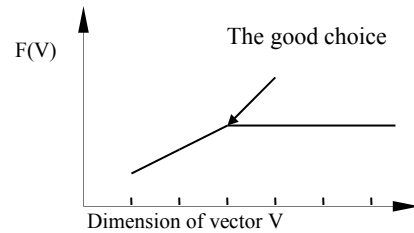


Fig. 4 The choice of the dimension of vector V

The first good use of ACO for feature selection seems to be reported in [9]. A. Al-Ani [9] proposes to use a hybrid evaluation measure that is able to estimate the overall performance of subsets as well as the local importance of features. A classification algorithm is used to estimate the performance of subsets. On the other hand, the local importance of a given feature is measured using the Mutual Information Evaluation



Function. Susana [19] proposes an algorithm for feature selection based on two cooperative ant colonies, which minimizes two objectives: the number of features and the classification error. The first colony determines the number (cardinality) of features and the second selects the features based on the cardinality given by the first colony. C.L. Huang [20] presents a hybrid ACO-based classifier model that combines ant colony optimization (ACO) and support vector machines (SVM). In his work, an ant's solution represents a combination of the feature subset and the classifier parameters,  $C$  and  $g$ , based on the radial basis function (RBF) kernel of the SVM classifier. The classification accuracy and the feature weights of the constructed SVM classifier are used to design the pheromone update strategy. Based on the pheromone table and measured relative feature importance, the transition probability is calculated to select a solution path for an ant. The major inconvenience with this method is the parameters of classifier are fixed during the execution of program and they may have different value in each solution.

### 3. Ant colony optimization (ACO)

Ant colony optimization (ACO) is based on the cooperative behavior of real ant colonies, which are able to find the shortest path from their nest to a food source. ACO algorithms can be applied to any optimization problems that can be characterized as follows [5, 16]:

1. A finite set of components  $C = \{c_1, c_2, \dots, c_N\}$  is given.
2. A finite set of  $L$  of possible connections/transitions among the elements of  $C$  is defined over a subset  $C'$  of the Cartesian product  $C \times C$ ,  $L = \{C_i C_j | (c_i, c_j) \in C'\}$ ,  $|L| \leq N^2 c'$ .
3. For each  $l C_i C_j \in L$  a connection cost function  $J_{C_i C_j} \equiv J(l C_i C_j, t)$ , possibly parametrized by some time measure  $t$ , is defined.
4. A finite set of constraints  $\Omega \equiv \Omega(C, L, t)$  is assigned over the elements of  $C$  and  $L$ .
5. The states of the problem are defined in terms of sequences  $s = (c_{i_1}, c_{i_2}, \dots, c_{i_k}, \dots)$  over the elements of  $C$  or of  $L$ .  $S'$  is a subset of  $S$ . The elements in  $S'$  define the problem's feasible states.
6. A neighborhood structure is assigned as follows: the state  $s_2$  is said to be a neighbor of  $s_1$  if  $s_1$  and  $s_2$  are in  $S$  and the state  $s_2$  can be reached from  $s_1$  in one logical step, that is, if  $c_i$  is the last component in the sequence determining the state  $s_1$ , it must exist  $c_2 \in C$  such that  $l C_i C_2 \in L$  and  $s_2 \equiv \langle s_1, c_2 \rangle$ .

7. A solution  $\Psi$  is an element of  $S'$  satisfying all the problem's requirements. A solution is said multi-dimensional if it is defined in terms of multiple distinct sequences over the elements of  $C$ .

8. A cost  $J_\Psi(L, t)$  is associated to each solution  $\Psi$ .  $J_\Psi(L, t)$  is a function of all the costs  $J_{C_i C_j}$  of all the connections belonging to the solution.

It is worth mentioning that ACO makes probabilistic decision in terms of the artificial pheromone trails and the local heuristic information. This allows ACO to explore larger number of solutions than greedy heuristics. Another characteristic of the ACO algorithm is the pheromone trail evaporation, which is a process that leads to decreasing the pheromone trail intensity over time. Pheromone evaporation helps in avoiding rapid convergence of the algorithm towards a sub-optimal region [5, 9, 16].

### 4. Support vector machines

In our wrapper approach, we have used SVM as classifier. SVM is an attractive learning algorithm first introduced by Vapnik [23]. It has a competitive advantage Compared to neural networks, and decision trees.

Given a set of data  $S = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_m, y_m)\}$ . Where  $x_i \in R^N$  is features vector and  $y_i \in \{-1, +1\}$  is a class label. The goal of the SVM is to find an of the form

$$w\phi(x) + b = 0 \text{ with } y_i(w\phi(x_i) + b) \geq 1 - \xi_i \quad (1)$$

that separating the  $S$  set of training data into two classes (positive and negative) (fig. 5). In general,  $S$  cannot be partitioned by a linear hyperplane. However  $S$  can be transformed into higher dimensional feature space for making it linearly separable.

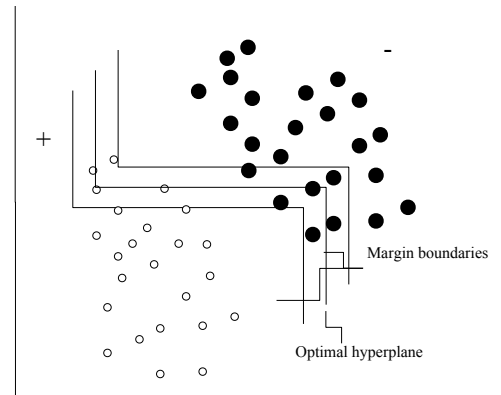


Fig. 5 Two-Class SVM used in linear classification



The mapping  $\phi(x_i)$  need not be computed explicitly; instead, an inner product Kernel of the form  $K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle$  (2)

To solve the optimal hyperplane problem, we can construct a Lagrangian and transforming to the dual. Then, we can equivalently maximize

$$\sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3)$$

subject to

$$\sum_{i=1}^m \alpha_i y_i = 0 ; 0 \leq \alpha_i \leq C, \quad (4)$$

For a test example  $z$ , we define the decision function as follow

$$\text{sign} \left( \sum_{i=1}^m \alpha_i y_i K(z_i, z_j) + b \right) \quad (5)$$

where

- $w$  is the weight vector.
- $b$  is the bias term.
- $C$  is the punishment parameter.
- $\alpha$  is the Lagrange multiplier.

In the next section, we present our proposed SVM/Binary ACO algorithm [6], and explain how it is used for selecting an appropriate subset of features.

## 5. Proposed approach

### 5.1. Description of the proposed approach

This research proposes a new implementation of Binary ACO algorithm [6] applied to feature selection, where the best number of features is determined automatically. In this approach, each ant searches the same routine, and pheromone is left on each edge. As an intelligent body, each ant just chooses one edge of the two as shown in Fig 6. The intelligent behavior of ant is very simple, and the incidence matrix traversed by each ant needs only  $2 \times n$ 's space, which to some extent solves the descriptive difficulty generated from long coding and the reduction of solution quality.

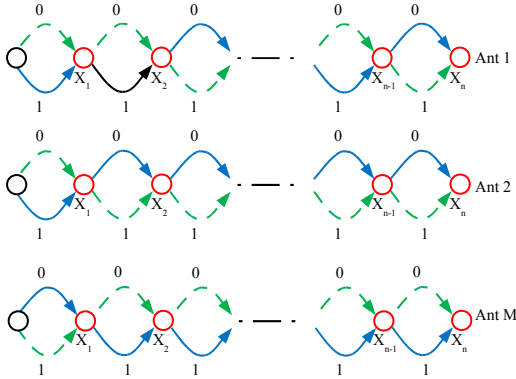


Fig.6 The final net obtained by the binary ACO algorithm

### 5.2. Probabilistic rule

Initially, the quantity of information in each routine is randomly generated. During the movement, ant  $k$  shifts its direction according to the values of pheromone concentration  $FP$  and the heuristic value  $FH$ . The heuristic value  $FH$  is computed using the Fisher discriminate criterion for feature selection [11] [17], which determines the importance of each feature, and it is described in more detail in Section 5.4. The probability that an ant  $k$  chooses the feature  $X_i$  is given by:

$$PS_{il} = \frac{FP_{il} + \frac{FP_{i0}}{\text{Max}(FH)} FH_i}{FP_{il} + FP_{i0}} \quad (6)$$

### 5.3. Updating rule

After all ants have completed their solutions, pheromone evaporation on all nodes is triggered, and then according to (2), pheromone concentration in the trails is updated.

$$FP \leftarrow (1 - \rho) FP + \Delta FP \quad (7)$$

Where  $\rho \in ]0,1[$  is the pheromone evaporation and  $\Delta FP$  is the pheromone deposited on the trails by the ant  $k$  that found the best solution for this tour:

$$\Delta FP = \frac{1}{1 + F(V) - F(V')} \quad (8)$$

Where  $F(V)$  represents the best solution built since the beginning of the execution and  $F(V')$  represents the best solution built during the last tour.

$F$  is the objective function of our optimization algorithm and  $V$  is the solution funded by the ant  $k$ .

The optimal subset is selected according to classifier performance and their length.

The results of this wrapper approach will be compared to a filter approach. The filter approach uses  $F'(V)$  an evaluation function.  $F'$  is calculated using two concepts: the variance in each class and the variance between classes.

$$F'(V) = \text{trace} \left( \sum_w^{-1} \sum_B \right) \quad (9)$$

Where the matrix of variance intra-class is calculated as follows:

$$\sum_w = \frac{1}{N} \sum_{C=1}^M \sum_{V=1}^{N_C} (X_{CV} - m_C)(X_{CV} - m_C)^t \quad (10)$$

Whereas the matrix of the variance inter-classes is calculated as follows:

$$\sum_B = \frac{1}{N} \sum_{C=1}^M (m_C - m)(m_C - m)^t \quad (11)$$

With:

- $m$ : General centre of gravity
- $M$ : A number of classes
- $m_c$ : Centre of gravity of the class number  $C$
- $X_{cv}$ :  $V^{th}$  vector of the class number  $C$
- $NR_c$ : A number of vectors of the class number  $C$
- $NR$ : Numbers total vectoriels

## 5.4. Heuristics

The heuristic value is computed using the Fisher discriminant criterion for feature selection [17]. Considering a classification problem with  $M$  possible classes, the Fisher discriminant criterion is described as follow:

$$FH(\alpha) = \sum_{c=1}^M \sum_{r=1}^{M-1} \frac{m_c(\alpha) - m_r(\alpha)}{N_c \sigma_c^2(\alpha) - N_r \sigma_r^2(\alpha)} \quad (12)$$

Where:

$M$  represents the number of classes;

$m_c(\alpha)$  Represent the centre of gravity of the class number  $C$  by considering only the parameter  $\alpha$  it is calculated as follows:

$$m_c(\alpha) = \frac{1}{N_c} \sum_{v=1}^{N_c} X_{cv}(\alpha) \quad (13)$$

With  $X_{cv}$  is the number  $v$  of the class number  $C$ . the value of  $NR$  equal to the number of vectors of the class in question is the vector.

$\sigma_r^2(\alpha)$  is the variance of the component  $\alpha$  of the vectors of the class number  $C$ .

$$\sigma_r^2(\alpha) = \frac{1}{N_c} \sum_{v=1}^{N_c} [X_{cv}(\alpha) - m_c(\alpha)]^2 \quad (14)$$

Algorithm 1 presents the description of the Binary ACO-SVM feature selection algorithm.

---

ALGORITHM I  
BINARY ACO/SVM FEATURE SELECTION ALGORITHM

---

- 1) Initiate the pheromone of the net;
  - 2) Compute the  $FH(\alpha)$  using (7);
  - 3) Ants search using (1);
  - 4) Evaluate the solutions founded by ant colony algorithm using SVM classifier, and reserve the optimal;
  - 5) Upgrade pheromone of the net by optimal solution using (2);
  - 6) Judge whether the stopping condition of the qualification is met, if qualified, ends; otherwise, Goto 3;
- 

The time complexity of proposed algorithm is  $O(I m)$ , where  $I$  is the number of iterations, and  $m$  the number of ants. This can be seen from Fig. 6. In the worst case, each ant selects all the features. As the objective function is evaluated after all ants have completed their solutions, this will result in  $m$  evaluations. After  $I$  iterations, the objective function will be evaluated  $I m$  times.

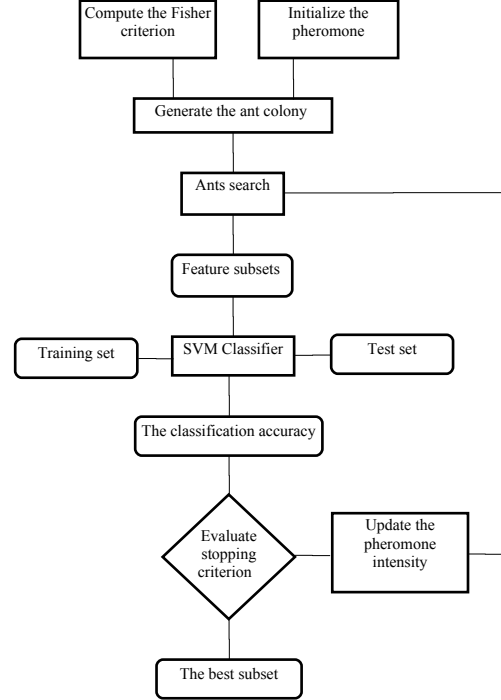


Fig. 7 SVM-Binary ACO feature selection algorithm

## 6. Experimental results

### 6.1. Data of test

The experimental results comparing the binary ACO algorithm with genetic algorithm are provided for real-life dataset (Vehicle) [18] and industrial dataset (Rotary Cement kiln) [7]. The Vehicle base is collected in 1987 by Siebert. Vehicle consists of 846 recordings which represent 4 classes. RCK consists of 200 recordings which represent 4 classes.

### 6.2. Parameters of the algorithm of selection

Like any other algorithm, before passing to the phase of selection. Some parameters should be fixed. This problem represents one of the disadvantages of the biomimetic methods. Since the values of the parameters are related to the number of individuals and the distribution of the data on the beach of

representation. The following table presents the values of the parameters of our algorithm. These parameters are fixed after the execution of several simulations by using as entered a restricted whole of data.

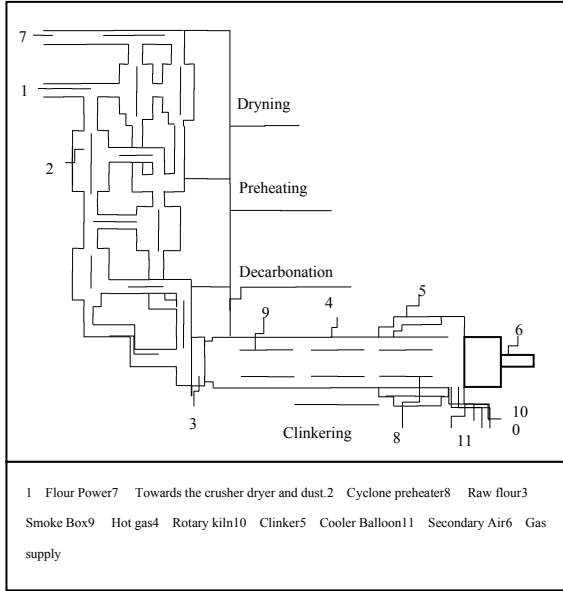


Fig. 8 Flow Diagram of a Rotary Cement Kiln [7]

TABLE 1  
PARAMETERS OF BINARY ANT COLONY

Parameter	Value	Description
N_A	20	A number of agents
F_a	0.2	Random rate of behaviour
$\rho$	0.3	Rate of evaporation
S_Min_FP	Min FH	Minimal threshold of pheromone
S_Max_FP	Max FH	Maximum threshold of pheromone

### 6.3. Heuristic factor FH

Heuristic factor FH is taken into account that by the ants which have a behavior related to the probability PS. The ants which have a random behavior are used to discover new spaces of research. The following figures represent the values of heuristic factor FH by using the dataset of the Rotary cement kiln and Vehicle dataset.

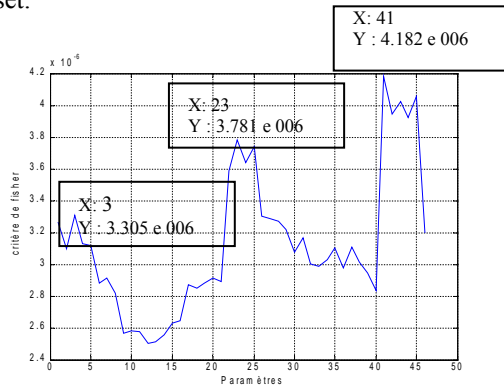


Fig. 9 The Fisher discriminant criterion for RCK dataset

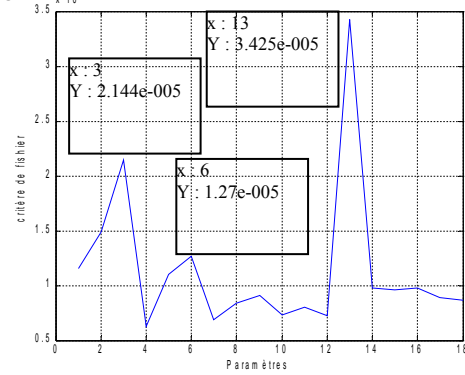


Fig. 10 The Fisher discriminant criterion for Vehicle dataset

### 6.4. Results

We tested the performances of our algorithm by using the evolutionary method of classification ECMC [21] and the following table shows the quality of classification while using:

- The best discriminating feature;
- The best subset of features generated;
- All features.

The implementation platform was implemented in *Matlab 7.9*, which is a general mathematical development tool. The Bioinformatics Toolbox functions *svmclassify* and *svmtrain* were used as the SVM classifier. The empirical evaluation was performed using an Intel Pentium Dual Core T4400 2.2GHz with 3GB RAM.

Using the parameters presented in the Table.1, the following results were obtained by taking the best solution after 20 BACOs trials. The Table.1 gives the best solutions obtained for each dataset (Rotary Cement kiln & Vehicle). For the two datasets, the FV of the best solution is indicated with the corresponding Rate of error. We conducted a performance comparison between the proposed wrapper-based ACO (ACO-SVM), the filter-based ACO and the filter-based GA.

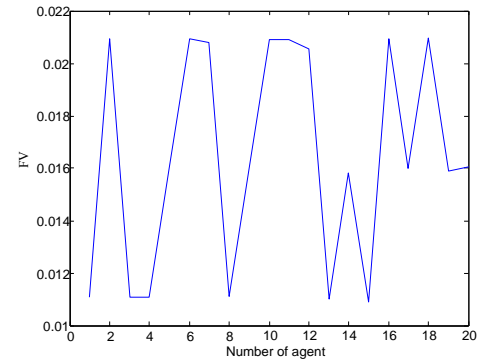


TABLE 2  
PERFORMANCES OF CLASSIFICATION BY USING THE VARIOUS ENTRIES

Algorithm	Features	Vehicle		Rotary Cement kiln	
		Error rate	F (V)	Error rate	F (V)
Wrapper-based ACO/SVM	One feature	75 %	0.0325	46 %	0.0194
	Generated subset	11 %	0.4717	10 %	0.5078
	All features	07 %	0.7875	10 %	0.5082
Filter-based ACO	One feature	75 %	0.0325	46 %	0.0061
	Generated subset	13 %	0.6537	15 %	0.0210
	All features	07 %	0.7875	10 %	0.0218
Filter-based GA	One feature	75 %	0.0325	46 %	0.0061
	Generated subset	11 %	0.7717	15 %	0.0210
	All features	07 %	0.7875	10 %	0.0218

Fig. 11 FV obtained by each agent during the last iteration

Table 2 shows that we obtain an acceptable rate of error with the subset generated by our algorithm. It is also noticed that the value of FV reflects well the quality of classification. The Fig. 9 shows the value of FV obtained by each agent during the last iteration using Rotary Cement kiln dataset.

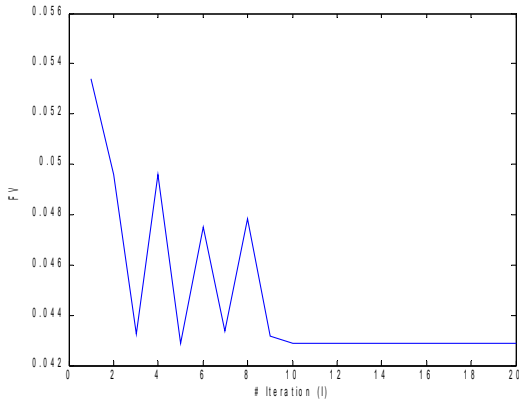


Fig. 12 The best solution obtained during each iteration

We notice that after the last iteration, more than 33% of agents find the optimal solution. This is due to the pheromone density which is updated at the end of each iteration.

Fig. 12 shows that we obtain the optimal solution after the 5th iteration which shows the effectiveness and the speed of our algorithm. The time of convergence of the presented algorithm can be reduced using a lower number of ants. This number is related to the number of features in the dataset.

Fig. 14 shows that our algorithm discards a bigger percentage of features for the case of Vehicle dataset. However, the selected features are not always the

same, once there are features that are weakly relevant and have a similar influence in the classifier.

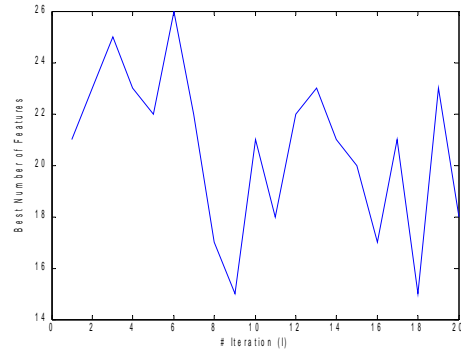


Fig. 13 Best number of features for each iteration in Rotary Cement kiln dataset

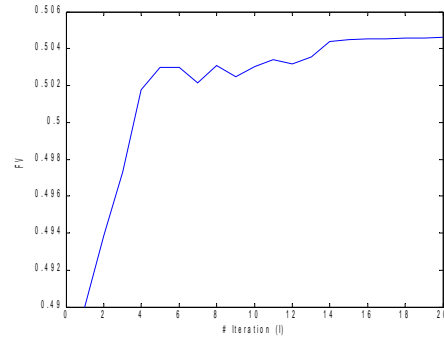


Fig. 14 The best solution obtained during each iteration (Vehicle)

The results given in figures 11-14 and Tables 2 show that our approach (Wrapper-based ACO-SVM) is very precise. In other word, it gives the optimal solution in compare to those of obtained by the other algorithms. In fact, the results obtained on the Rotary Cement kiln dataset show that our approach converges to the global optimum in all of runs.

## 7. Conclusion

In this work, a new approach for selecting best discriminates features subset using Binary ACO algorithm is presented. The ACO is chosen for this study because it is the newest metaheuristic. The goal is to select the best subset that is sufficient to perform a good classification and obtain acceptable rate of error. We have tested the proposed method on two datasets. The experimental results indicate that the proposed Binary ACO algorithm can be applied for larger number of features.

The classifier induced in the experiments was a SVM. This classifier was chosen because it does not

suffer from the local minima problem, it has fewer learning parameters to select, and it produces stable and reproducible results, but our wrapper method can be used with any other supervised classifiers.

In the near future, the performance of the proposed algorithm will be compared with other features selection methods to improve that our algorithm achieving equal or better performance. And we will combine our algorithm with other intelligent classifiers, such as neural networks classifiers.

## 8. References

- [1] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning artificial," *Artificial Intelligence*, vol. 97, pp. 245–271, 1997.
- [2] P.M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Transactions on Computers*, vol. 26, pp. 917–922, 1977.
- [3] J. Kittler, "Feature set search algorithms," In C. H. Chen, editor, *Pattern Recognition and Signal Processing*. Sijhoff and Noordhoff, the Netherlands, 1978.
- [4] J. H. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems*, Vol. 13, no. 2, pp. 44–49, 1998.
- [5] V. Maniezzo, M. Dorigo and A. Colomi, "The ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, vol. 26, no.1, pp. 29–41, 1996.
- [6] X. Weiqing, Y. Chenyang and W. Liuyi, "Binary Ant Colony Evolutionary Algorithm," *International Journal of Information Technology*, Vol. 12, no. 3, 2006.
- [7] R. Mahdaoui, L.H. Mouss, "NEFDIAG. a New Approach for Industrial diagnosis by Neuro-Fuzzy systems: Application to manufacturing System," *ERIMA*, vol.2, no. 2, pp. 144–151, 2008.
- [8] M.L. Raymer, W.F. Punch, E.D. Goodman, L.A. Kuhn and A.K. Jain, "Dimensionality Reduction Using Genetic Algorithms," *IEEE Transactions on Evolutionary Computation*, vol 4, no. 2, pp 164–171, 2000.
- [9] A. Al-Ani, "Feature Subset Selection Using Ant Colony Optimization," *International Journal of Computational Intelligence*, vol 2, no. 1, pp 53–58, 2005.
- [10] Z. Zhang, W. Cheng and X. Zhou, "Research on Intelligent Diagnosis of Mechanical Fault Based on Ant Colony Algorithm," [\*The Sixth International Symposium on Neural Networks\*](#), Springer Berlin, Heidelberg, Vol 56, pp 631–640, 2009.
- [11] E. Youn, "Support vector-based feature selection using Fisher's linear discriminant and Support Vector Machine," *Expert Systems with Applications*, vol, 37, no 9, pp. 6148–6156, 2010.
- [12] Q. Wu, Fault diagnosis model based on Gaussian support vector classifier machine. *Expert Systems with Applications*, vol, 37, no 9, pp.6251–6256, 2010.
- [13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol 3, pp 1157–1182, 2003.
- [14] M. Dash and H. Liu, "Feature selection methods for classifications," *Intelligent Data Analysis*, vol 3, no.1, pp. 131–156, 1997.
- [15] N. Wyse, R. Dubes, and A.K. Jain, "A critical evaluation of intrinsic dimensionality algorithms," In E.S. Gelsema and Kanal L.N, editors, *Pattern Recognition in Practice*, Morgan Kaufmann Publishers, Inc, pp 415–425. 1980.
- [16] M. Dorigo and G. Di Caro, "The ant colony optimization meta-heuristic," In *New Ideas in Optimization*, McGraw-Hill, London, UK, pp 11–32, 1999.
- [17] R. O. Duda, P. E. Hart and D. G. Stork, in *Pattern classification*, 2nd ed. Wiley Interscience Publication, 2001.
- [18] A. Asuncion, and D. J. Newman, *UCI machine learning repository*, 2007.
- [19] S. M. Vieira, J.C. Sousa and T.A. Runkler, "Two cooperative ant colonies for feature selection using fuzzy models," *Expert Systems with Applications*, vol. 37, no.4, pp. 2714–2723, 2010.
- [20] C. L. Huang, "ACO-based hybrid classification system with feature subset selection and model parameters optimization," *Neurocomputing*, vol 73, pp 438–448, 2009.
- [21] M. F. Pasha, R. Budarto, "Evolvable-NEURAL- Based Fuzzy Inference System and Its Application for Adaptive Network Anomaly Detection," *Advances in machine learning and cybernetics*, vol. 3930, pp. 662–671, 2006.
- [22] O. Boz, "Feature subset selection by using sorted feature relevance," *International Conference on Machine Learning and Application*, Las Vegas City, , P147–153, 2002.
- [23] W.N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions of Neural Networks*, vol 10, pp. 988–999, 1999.

# STUDY FOR A MECHANISM AIDED BY ASYNCHRONOUS ACTUATOR POWERED BY ASYNCHRONOUS DIESEL GENERATOR

H.Meglouli <sup>(a)</sup>, B. Ikhlef <sup>(b)</sup>

<sup>(a)</sup> Research Laboratory of industrial electrification

<sup>(b)</sup> Research Laboratory of applied automatic  
Faculty of hydrocarbons and chemicals.  
University of Boumerdes

<sup>(a)</sup> [hmechlouli@yahoo.fr](mailto:hmechlouli@yahoo.fr), <sup>(b)</sup> [blm\\_ikhlef@yahoo.fr](mailto:blm_ikhlef@yahoo.fr)

## ABSTRACT

The modern electric facilities are equipped by a great number of different mechanisms and devices activator by Asynchronous electric Motor (ASM), the power of these motors is equal to the power of the generating devices, where their most complicated working regime is the starting when their power is equal to the power of the generating devices.

In this regime we can have an overcharge of the generating devices by the active and reactive power.

For this reason, this article is dedicated to the study of the starting methods of asynchronous motors that action the mechanisms and that are powered by Asynchronous Generating Diesel (AGD) with a limited capacity of DRY value and a given couple of resistance.

Keywords: Reliability, Autonomous asynchronous generator, starting of the asynchronous motors, Tension converter.

## 1. INTRODUCTION

There are several factors that considerably influence the characteristics of the asynchronous motors starting process from a AGD, amongst these factor:

- The initial conditions of the process,
- The oscillation of frequency and amplitude of the AGD tension, - The non linear character of the electric machines parameters used as an activator motor for the generator,
- The mechanisms resistant couple.

If we take into account these factors, the study of the transient regime in the AGD-ASM system using analytical methods will be complex and will induce high calculations uncertainties. In this article, we will study the analysis of the regimes dynamics of the starting from AGD of mechanisms with asynchronous electric activator using a numerical method to achieve a given precision of the calculations.

## 2. MATHEMATICAL MODEL OF THE STARTING REGIME OF MECHANISMS WITH ASYNCHRONOUS ACTIVATOR POWERED BY AGD

### 2.1. General characteristic of the model

Studies performed previously have shown that for a complete analysis of the common operating regime of AGD and ASM, it is necessary that the mathematical model takes into account the review of the transient regimes for the direct starting of the motor, the starting through an auxiliary resistance in the statoric circuit as well as the starting through the Tension Converters with Thyristors (TCT). In addition to considering the electromagnetic systems properties, the supplying of the consumers by a three-phased tension under neutral line, the necessity in a large interval of the regulation of the key elements starting angle value (because the activating mechanisms has the same power that the generator) In the mathematical model, the functioning of the TCT of 3TT type that is composed of two thyristors connected head to foot in every phase of the supplying line is described.

### 2.2. Mathematical model of the asynchronous motor with short-circuited rotor.

In order to study the operating regime of ASM with the auxiliary elements connected to its statoric coil, we have to write the composed differential equations in relation to the statoric current and to the rotoric hooking flux in a simple shape, in the  $(\alpha, \beta, \sigma)$  coordinate system.

The ASM equations are the following:

$$\frac{di_{\alpha M}}{d\tau} = [U_{\alpha} - R_{SM}i_{\alpha M} - L_1(r_{rM}L_1L_{\alpha M} - T_1\psi_{\alpha M} - w_{rM}\psi_{\beta M})] \frac{1}{L_S\Sigma}$$

$$\frac{di_{\beta M}}{d\tau} = [U_{\beta} - R_{SM}i_{\beta M} - L_1(r_{rM}L_1L_{\beta M} - T_1\psi_{\beta M} - w_{rM}\psi_{\alpha M})] \frac{1}{L_S\Sigma}$$

$$\frac{d\psi_{\alpha M}}{d\tau} = r_{rM}L_1i_{\alpha M} - T_1\psi_{\alpha M} - w_{pM}\psi_{\beta M}$$

$$\frac{d\psi_{\beta M}}{d\tau} = r_{rM}L_1i_{\beta M} - T_1\psi_{\beta M} - w_{rM}\psi_{\alpha M}$$

where :

$$R_{SM} = r_{rM} + R_n$$

$$L_{S\Sigma} = L_{SM} + L_n$$

$$L_1 = \frac{L}{L_{mM} + L_{rM}}$$

$$T_1 = \frac{r_{rM}}{L_{mM} + L_{rM}}$$

$L_{rM}$  and  $L_{sM}$ : ASM's rotor and stator flux scattering inductances.

$L_n$ : Inductance of the auxiliary resistance.

$L_{mM}$ : Mutual inductance of the ASM's stator and rotor coils.

$w_{rM}$ : The pulsation of the ASM's rotor rotation speed.

$i_M$ : ASM stator current.

$\psi_M$ : Hooking flux of the ASM's rotor

The calculation of the transient regime in the ASM is performed in a relative units system, the nominal current amplitudes and the generator tension are taken as basis value, the basis time is the same for the processes calculation. The instantaneous values of the motor current in the generator relative units system are transformed using of the transfer coefficients

$$K_\sigma = \frac{i_{\sigma M}}{i_{\sigma G}}$$

$i_{\sigma M}$  and  $i_{\sigma G}$ : Basis values of the currents in the motor and the generator relative units system.

The electromagnetic couple developed by ASM is determined by the formula (2):

$$C_M = \psi_{\alpha M} \cdot i_{\beta M} - \psi_{\beta M} \cdot i_{\alpha M}$$

The equation of the movement of the mechanism axis that action the ASM is:

$$\frac{dw_{rM}}{d\tau} = \frac{1}{J_\Sigma} \left( C_M - \frac{C_{cM}}{C_{\sigma M}} \right)$$

$J_\Sigma$ : Sum of the inertias couples of the ASM mobile mass and of the mechanism translated to the motor rotor

$C_{cM}$ : Couple of the mechanism resistance (N.m).

$C_{\sigma M}$ : Basis value of the ASM couple (N.m).

To take into account the influences of the rotor current and the saturation of the machine iron on the variation of the rotation frequency at the starting time we include in the mathematical model the relations linking the rotor scattering inductance and the motor rotor active resistance with the motor sliding.

$$r_{rM} = (r_{nr} - r_{rM}) g_M$$

$$L_{rM} = (L_{nr} - L_{rM}) g_M$$

Where:

$r_{nr}$ : Active resistance of the rotor.

$L_{nr}$ : Rotor scattering inductance at the starting.

$g_M$ : Motor sliding ( $g_M = 1$ )

C - Mathematical model of the Tension Converter with Thyristors TCT For the development of the TCT mathematical model we take into account the following: The arm of each branch is composed of two thyristors connected head to foot when the command signal arrives to the thyristor trigger and become in the closed state regardless of the system tension where it is present at that given moment.

The thyristor remains closed so far the value of the current that crosses it is higher than the upholding current.

In its closed state, the thyristor is replaced by an active resistance; the drop of tension in this latter one corresponds to the drop of tension value in the thyristor in the closed state.

In its open state, the thyristor is replaced by an active resistance in which the current becomes equal to the inverse current of the chosen thyristor.

Taking into account these simplifications, the control of the thyristors state is achieved by the analysis of every step of the command tension value calculation and the value of the current that crosses it at that given moment. To achieve this objective, the mathematical model takes into consideration the equations of the ASM phase current derived from the first two equations of the system (1)

$$\frac{di_a}{d\tau} = [(U_a - U_N) - R_{SM}i_a - L_1(r_{rM}L_1i_a - w_{rM}\psi_{\beta M})] \frac{1}{L_{SM}}$$

$$\begin{aligned} \frac{di_b}{d\tau} = & [(U_b - U_N) - R_{SM}i_b - L_1[\frac{1}{2}(r_{rM}L_1i_a - T_1\psi_{\alpha M} - w_{rM}\psi_{\beta M}) \\ & + \frac{\sqrt{3}}{2}(r_{rM}L_1\frac{i_b i_c}{\sqrt{3}} - T_1\psi_{\beta M} + w_{rM}\psi_{\alpha M})]] \frac{1}{L_{s\Sigma}} \end{aligned}$$

$$\begin{aligned} \frac{di_c}{d\tau} = & [(U_c - U_N) - R_{SM}i_c - L_1[-\frac{1}{2}(r_{rM}L_1i_a - T_1\psi_{\alpha M} - w_{rM}\psi_{\beta M}) \\ & - \frac{\sqrt{3}}{2}(r_{rM}L_1\frac{i_b i_c}{\sqrt{3}} - T_1\psi_{\beta M} + w_{rM}\psi_{\alpha M})]] \frac{1}{L_{s\Sigma}} \end{aligned}$$

Where:

$$U_a = U_\alpha$$

$$U_b = \frac{-U_\alpha}{2} + \frac{\sqrt{3}}{2}U_\beta$$

$$U_c = \frac{-U_\alpha}{2} - \frac{\sqrt{3}}{2}U_\beta$$

$U_a, U_b$  and  $U_c$ : tensions of phases of AG

$U_N$ : Tension between the neutral points of the AG and ASM stator coils.

The functioning algorithm of TCT in the lack of a neutral link line in the naval network is limited by three possible conduction regimes that are:

- Three-phased conduction: closed arm for all phases.
- Two-phased conduction: closed arm for any phase.



- Neutral conduction: open state of the arms for the three phases.

For the TCT chosen types, taking into account the previous simplifications on the asynchronous machines symmetry,

$U_N$  value can be determined for any time moment. At the time of the functioning of the symmetrical electric machines,  $U_N$  is different from zero only in the case of the two-phased conduction. For the thyristors that are used in the model and at the time of passing from TCT to any

conduction state that precedes the two-phased one, the initial values of the currents in the phases of open thyristor are equal in value but opposed in phase, on the other hand,  $U_N$  because in the AGD-ASM system, the symmetrical regime still exists at the following time moment. The same current will cross the two phases with open thyristor.

In these conditions, the instantaneous value of  $U_N$  can be calculated using the system of equations (7) that leads to the following values:

The phase (a) closed:

$$U_N = \left( U_b + U_c + L_1 \cdot \frac{d\psi_{\alpha M}}{d\tau} \right) \cdot \frac{1}{2.L_{SM}}$$

The phase (b) closed:

$$U_N = \left( U_b + U_c - L_1 \cdot \left( \frac{1}{2} \frac{d\psi_{\alpha M}}{d\tau} - \frac{\sqrt{3}}{2} \frac{d\psi_{\beta M}}{d\tau} \right) \right) \cdot \frac{1}{2.L_{SM}} T$$

he phase (c) closed:

$$U_N = \left( U_a + U_b - L_1 \cdot \left( \frac{1}{2} \frac{d\psi_{\alpha M}}{d\tau} - \frac{\sqrt{3}}{2} \frac{d\psi_{\beta M}}{d\tau} \right) \right) \cdot \frac{1}{2.L_{SM}}$$

The opening of the thyristors command signal is formed using the command law taken for the TCT.

### 3. ALGORITHM OF CALCULATION OF THE STARTING REGIME

The simulation by MATLAB software of the system ((1) - (5)), using the Runge-Kutta method, has allowed us to get the following results of the ASM starting powered by AGD.

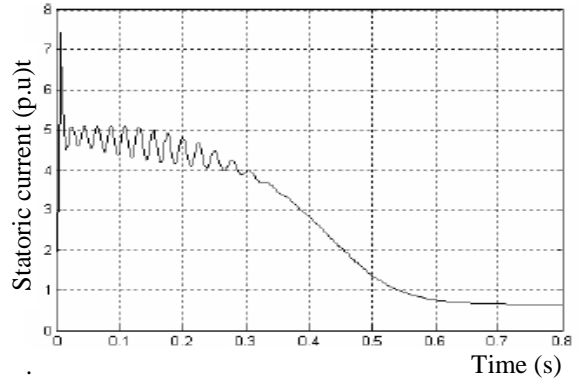


Figure 1: Variations of the  $I_s$  current in terms of the time for a direct ASM starting powered by AGD

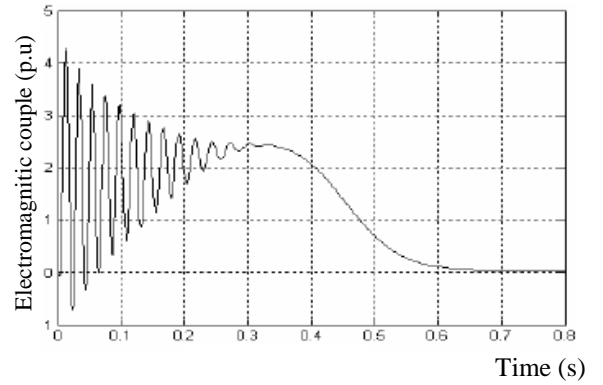


Fig. 2. Variations of the electromagnetic couple according to the time for a direct ASM starting powered by AGD

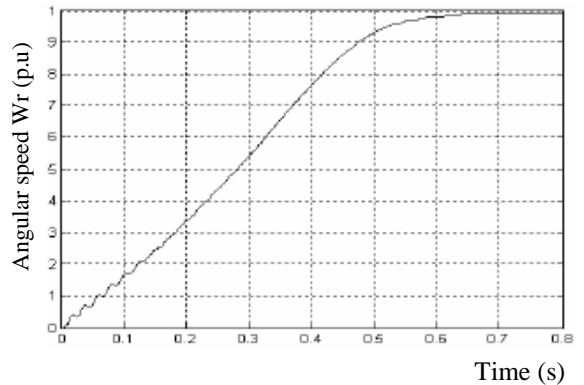


Fig. 3. Variations of the angular speed  $W_r$  according to the time for a direct ASM starting powered by AGD.

We notice on the three previous figures that at the time of the direct ASM starting powered by AGD, the existence of peaks of statoric currents and peaks of couple that can be the origin of the machine destruction by overheating, especially in the case of excessive repetitions



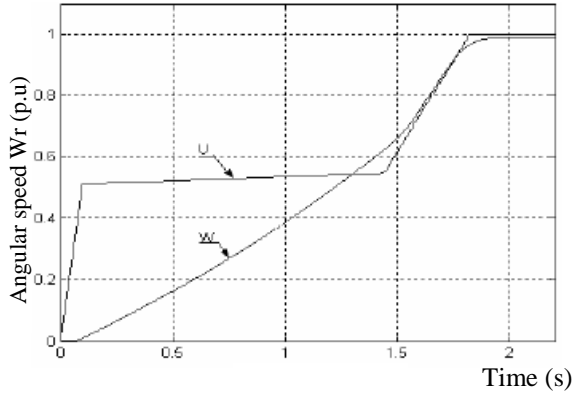


Fig. 4. Variations of the angular speed  $W_r$  and the  $U$  tension according to the time for an ASM starting with a TCT powered by AGD

We notice on this figure that for this type of starting the variation of tension that powers the ASM is progressive.

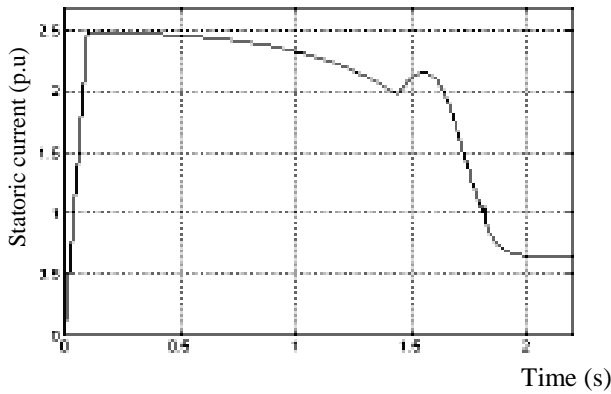


Fig. 5. Variations of the  $I_s$  current in terms of time for an ASM starting with a TCT powered by AGD

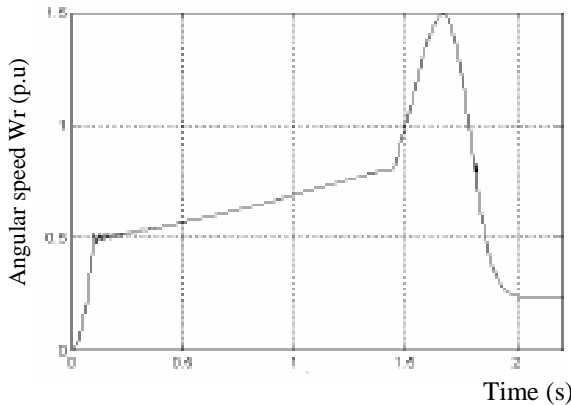


Fig. 6. Variations of the electromagnetic couple  $C_{em}$  in terms of time for an ASM starting with a TCT powered by AGD

We notice on the two previous figures, the absence of any current or couple abrupt peaks. The resulting drop of tension that and mechanical shocks due to the brutal apparition of the couple. The starting time in this case can exceed the direct starting time by several times.

#### 4. COMMAND OF THE STARTING REGIME BY A TENSION CONVERTER WITH THYRISTORS

For an ASM starting, the limitation of the current peaks can be obtained not only by the reduction of the tension amplitude via the introduction of the auxiliary resistances in its statoric circuit, but also using other regulation methods of this tension value in the devices that allow the command of this starting regime, by means of varying the commutation of the allowed or blocked state of the semiconductor components (thyristor, power transistor, triac).

The most efficient activators of asynchronous mechanisms are the starting devices constructed on the basis of Tension Converters with Thyristor (TCT) commanded by a phase angle. For an automatic command of the starting regimes of the asynchronous activators mechanisms, several solutions exist currently. Amongst them we can mention the solution that uses gradators, where the power circuit includes in every phase two thyristors assembled head to foot; the variation of tension that powers the ASM is progressive and is obtained via varying the conduction time by phase angle of these thyristors during every half period (fig.7).

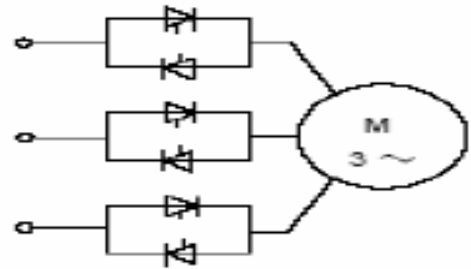


Fig. 7. Block diagram of the ASM starting through a gradator

This type of starting limits the call of current, the ensuing drop of tension and the mechanical shocks resulting from the brutal apparition of the couple. For the ADG energizing systems linked to an excitation device (DRY) whose action rapidity can be compared to the action rapidity of the command system by phase angle of the TCT key elements and which can have a positive effect if we introduce in the TCT automatic system a negative return loop between the starting angle of the thyristors' triggers and the drop of tension between the AGD limits. The possibility to use a system with TCT for the ASM starting command from an AGD is represented on the figure (8). For the elements of commutation we use some photothyristors that can be commanded by a luminous impulse. To assure a galvanic insulation between the power circuit and the command circuit at the time of the of the installation functioning, the primaries of the impulsion transformers T1-T3 are joined with the AGD statoric coils, the secondary of these transformers are plugged with the Zener diodes DZ7-DZ12 stabilizing the phases tensions.

If a drop of tension appears the phase changes and the length of the command luminous impulse that determines the value of the starting angle of the photothyristor also changes. The choice of the command tensions phase angle and the parameters of the photothyristors allow the creation of a negative return loop by a drop of tension. We can have a large regulation interval of the photothyristors starting angle by means of the formation of the phase command signal that has a tension which exceed the anode-cathode tension of these photothyristors in a varying interval between  $30^\circ$  and  $120^\circ$ . The relation of the photothyristors starting angle with the tension phase amplitude is determined by the equation (8)

$$\alpha = \frac{\pi}{3} + \arcsin \cdot \frac{U_{sT} \cdot k_T}{U_m}$$

Where:

$\alpha$  : Pothothyristors starting angle

$U_{sT}$  : Stabilizer tension

$k_T$  : Transformers transformation coefficient

$U_m$  : Amplitude of the tension phase instantaneous value

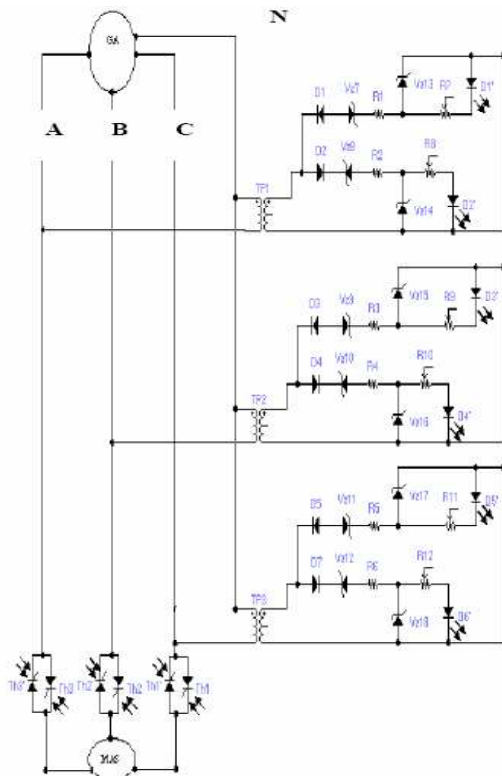


Fig. 8. Electric diagram of the ASM starting powered by AGD through a TCT

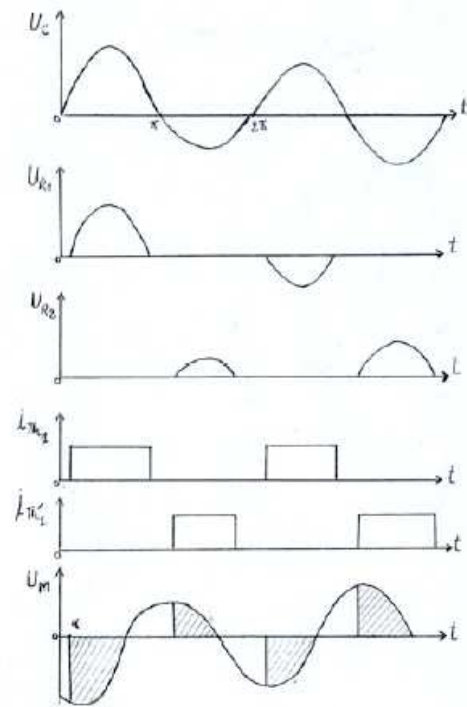


Fig . 9 . Functioning chronogram of the photothyristors

We take into consideration the variation character of the ASM power coefficient at the time of its starting. The chosen regulation interval of the starting angle is sufficient to command this starting because it ensures the conditions of TCT functioning in the case of three or two conductor arms.

On the figure (10) we represented the oscillogramms of the regime in the case of an ASM starting through a TCT. The curves represent the effective tension value, the frequency of the diesel generator activator and the starting duration of ASM for different powers. The comparison between these features and the parameters of the ASM's direct starting regimes (fig(1)-(2)) permits to point out that the use of the TCT lead to the decrease the drop of tension in the load, the starting time in this case can exceed the time of direct starting by several multiples.

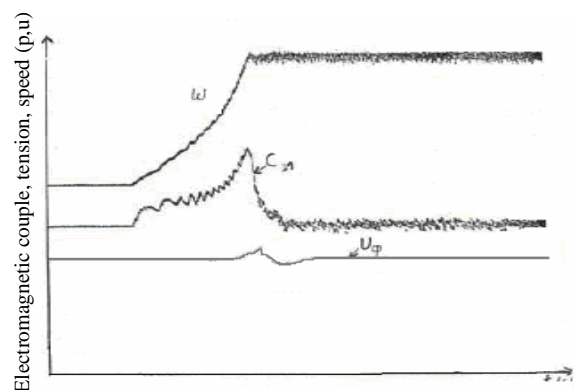


Fig . 10. Oscillogramme of the ASM starting powered by AGD through a TCT

## 5. CONCLUSION

We have developed a mathematical model of the common functioning regime dynamics of the AGD and of the mechanisms with asynchronous actionneur. This model takes into consideration the elements of the following system:

**1-** Diesel, asynchronous generating, asynchronous motor with auxiliary resistances in the statoric circuit with a resistance couple on the corresponding axis of the different mechanisms, tension regulator with thyristors, the calculation parameters of the transient regimes are closer of the experimental data with a precision around 13%.

**2-** The calculations of the parameters of the asynchronous machines starting regimes of 4A, AM and AO series, for a capacity of DRY equal to 1.45 pu and for a limitation of tension drop on the borders of AGD of 20% of  $U_n$ , give a steady direct starting of the asynchronous machines of a power of 20 to 25% of the nominal power of AG, without limitation of the tension drop value and for the same value of DRY capacity the AGD can assure a direct ASM starting with a power of 30 to 40% of the nominal power of AG.

**3-** The analysis of the possibilities of the starting regime command of the mechanisms with asynchronous actionneur by a AGD with the use of a TCT with negative return loop between the angle of the thyristors triggers opened state and the drop of tension on the borders of AGD, shows that by this method we can limit the drop of tension on the borders of the load and can increase the unit of motors power started for a limited value of the DRY capacity. We noticed that the couple developed by ASM in the time of starting through a TCT has a smaller value compared to the starting through auxiliary resistances for the same value of tension drop that appeared on the AGD borders.

## REFERENCES

- Glasinko T.A., Khrisanov V.I., 1981 Semiconductor systems electric actuators, low-power asynchronous impulsive, Leningrad, *Inergoatomisdat*, 144 p.
- Abudura S., Konacof G.A. 2000 The determination of the power of Diesel in naval operations *Installation énergétique navale*, N°2, Odessa, p 43-49.
- Caron. J.P. J-P.Hautier, 1995 Modelling and control of the asynchronous machine, éditions TECHNIP, 324p
- Kobilov, I.B. 1986 Electrical machinery, Moscou, *Inergoatomisdat*, 360p
- Kobilov, I.B. 1987, Mathematical modeling of electrical machines, Moscou, The upper school, 247p.
- Abudura, S. 2002 The regimes of operation of naval diesel engines and diesel generators, Odessa, Konsalting, 236 p.

## AUTHORS PROFILE

Dr. H. Meglouli is currently working as Assistant Professor at the Department of Automatisation and electrification of process, in university of Boumerdes-Algeria. He obtained his Ph. D degree in technical sciences at petroleum institute of Moscow.

B. IKHLEF is currently a doctoral student at the University of Boumerdes, he received his engineering degree in automatic control at the University of Tizi Ouzou, Algeria, and master degree in applied automation and signal processing at the University of Boumerdes.

# MATHEMATICAL MODELING AND SIMULATION FOR DETECTION OF SUICIDE BOMBERS

William P. Fox<sup>(a)</sup>, John Vesecky<sup>(b)</sup>, Kenneth Laws<sup>(b)</sup>

(a) Department of Defense Analysis: Naval Postgraduate School, Monterey, CA 993943

(b) Department of Electrical Engineering, University of California at Santa Cruz

(a) wpfox@nps.edu, (b) vesecky@cse.ucsc.edu, (c) kip@soe.ucsc.edu

## ABSTRACT

We examined the use of radar to detect humans wearing a suicide bomb vest with detonation wires. In our research we used the GunnPlexer Doppler radar at 12.5 GHz to collect experiment data of humans both with wires and a vest and without wires. We collected data and broke the reflected radar signal into both horizontal and vertical polarization (HH and VV). We developed several metrics from this data that could be used in building models or algorithms to more accurately detect subjects wearing wires. We discovered additional information about the metrics and used combinations of the metrics so we could increase the detection probability. We built a Monte Carlo simulation to test our theories. To date, we have a success rate over 98% and a false positive rate of under approximately 2%. This research and the results encourage us to think that suicide bombers can be found prior to their detonation of their bombs at a safe range.

Keywords: military application, radar cross section, detection, suicide bomber

## 1. INTRODUCTION

Improvised Explosive Devices (IEDs) are a major problem in the world we face today (Meigs, 2007). A major IED concern is the suicide bomber. The suicide bomber generally does not present their action prior to the event and can more easily accomplish their goal. We examine the dynamics involved in the suicide bomber and possible detection strategies using a stand-off radar.

The general observational situation we consider is illustrated in figure 2, below. We see one or more radars observing a crowd of people of whom one or more have wires on their bodies. Those with wires might be terrorists who plan to explode their suicide bomb. We anticipate that the range from the radar to the people (or animals) under observation would be typically 50 to 100 meters. Our plan is to make observations with one or more radars (and likely other sensors as well, such as video surveillance cameras or thermal imaging). The results of these observations become the essential input data to our mathematical model that assesses the system's ability to detect

*suspects* (persons suspected of harmful intent) from among a crowd of *subjects* who are largely harmless.

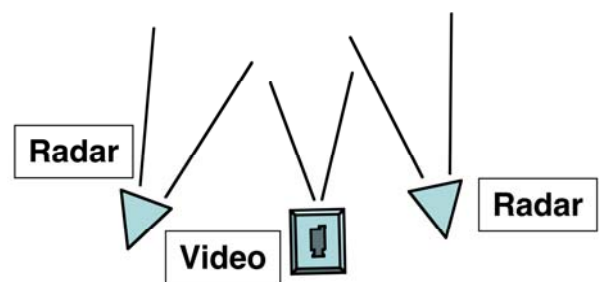
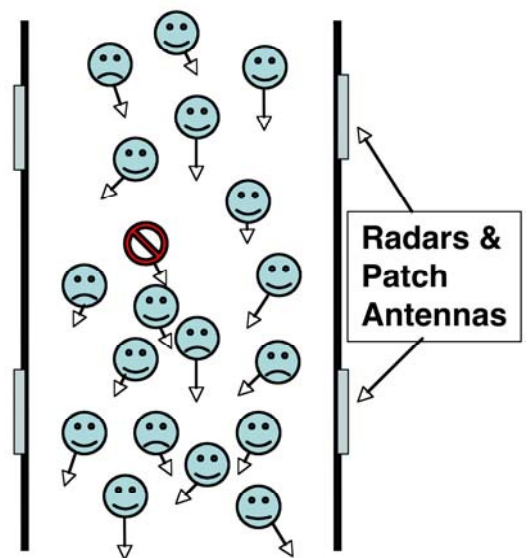


Figure 2. Radar observational geometry. One or more radars observe a group of people with one or two having wires on their bodies and hence becoming suspects.

We discuss the radar observational systems, the radar cross sections of humans both with and without wires on their bodies (from both experimental measurements and computational electromagnetic estimates), our mathematical models with metrics and our findings and conclusions with recommendations.

## 2. DETECTING SUICIDE BOMBERS

Data was collected using the GunnPlexer radar on persons both with and without wires & vests. This data has been analyzed. We begin by displaying the scatterplots, see Figure 1-3. Each plot indicates a visual exponential distribution. Using goodness of fit chi-squared analysis we found each does follow an exponential distribution.

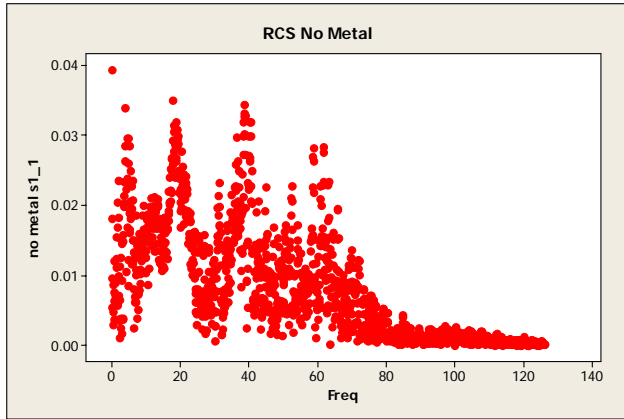


Figure 1. No wire on persons

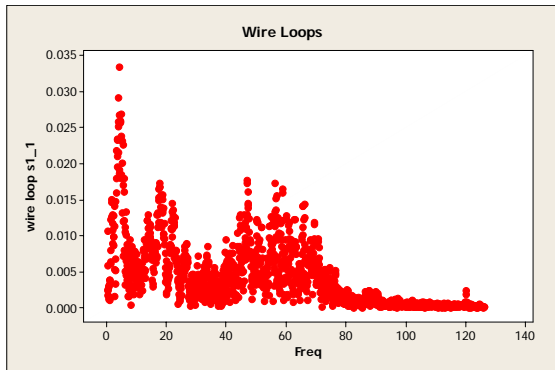


Figure 2. Persons wearing wire loops

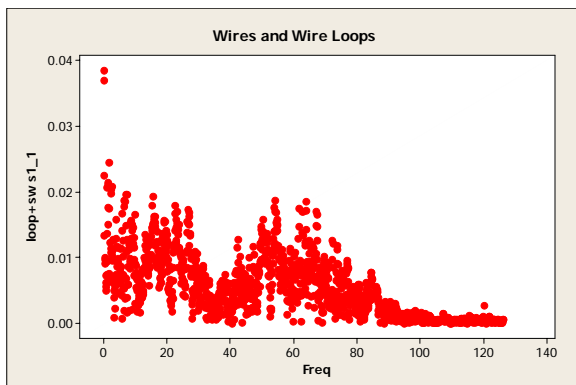


Figure 3. Person wearing wires and wire loops

Analysis of the data used to create these graphs show that each follows an exponential distribution. We used a Chi-squared goodness of fit test at  $\alpha = 0.05$  for each test.

First, we took the scaled or normalized the data and then display a histogram of the data in Vest 1, see the figure 4 below.

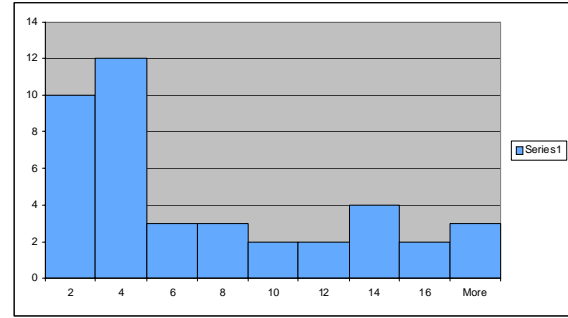


Figure 4. Histogram of Data set 1, Vest Configuration 1

We used the  $\chi^2$  goodness of fit test to a truncated exponential distribution:

$$H_0 : f(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda x_0}}, 0 \leq x \leq x_0$$

Since our test statistic value is less than my critical value then, we conclude that the truncated exponential with empirical mean 0.15209355 is a good fit at an  $\alpha$  level of 0.05

$$\chi^2 = 5.11619$$

$$\chi^2_{.05,4} = 9.48$$

We perform the same analysis for the data from Vest 2 seen in figure 5.

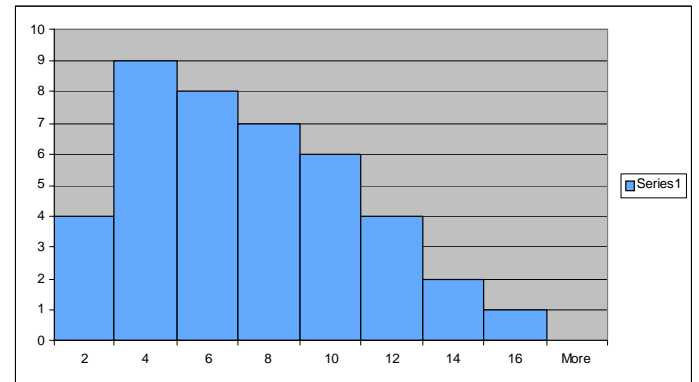


Figure 5. Histogram of data set 2, Vest Configuration 2

We tested using a goodness of fit test to a truncated exponential distribution:

$$H_0 : f(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda x_0}}, 0 \leq x \leq x_0$$

We conclude that the truncated exponential with empirical mean 0.156108622 is a good fit at an  $\alpha$  level of 0.05.

$$\chi^2 = 4.6898$$

$$\chi^2_{.05,4} = 9.48$$

Both empirical distributions are essentially exponential distributions and that is supported by both the literature and other's research (Dogaru et al., 2007, Fox, et al, 2010, Angell et al. 2007a, 2007b ).

We examined the vertical and horizontal polarization of the data that according to the literature might be able to distinguish certain objects. Linear polarization has been found to detect metal. We found that comparing the VV to HH polarization of our subjects was useful to identify metal. Our plots of the polarization data very closely resemble those of Dogaru et al. (2007), shown below in figure 6-7.

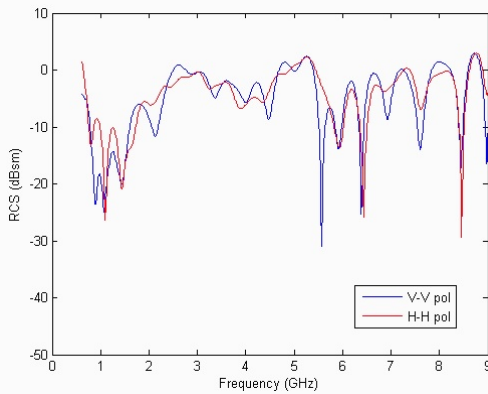


Figure 6. (Radar cross section of a simulated human body in both VV and HH polarization over the frequency range from 0.5 to 9 GHz. After Dogaru et al. (2007))

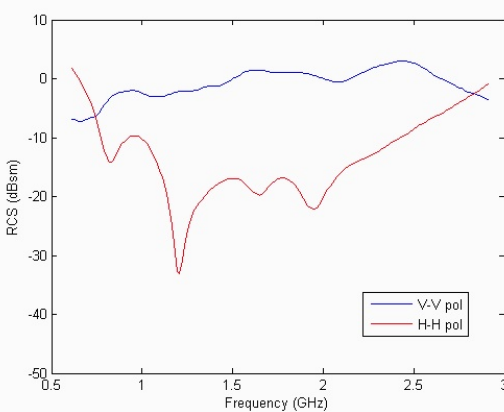


Figure 7. (Radar cross section of a human body carrying a thin, 1 m metal rod in front of the body. After Dogaru et al. (2007))

It is easy to see that the two figures are different. Figure 6 shows the two graphs (blue and red functions) and the plots are close to being the same. Figure 7 clearly shows visually that the two plots (blue and red)

appear to be different. In fact, statistical analysis shows this is true. We analyzed two sets of data for person with wires in different arrays and tested the means in pairs to show they are different.

$\mu_1$  = mean for person with wires

$\mu_2$  = mean for person with wires (Vest 2)

$\mu_3$  = mean for persons with wires and loops (Vest 3)

Case 1:

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

Case 2:

$$H_0 : \mu_1 = \mu_3$$

$$H_a : \mu_1 \neq \mu_3$$

Case 3:

$$H_0 : \mu_2 = \mu_3$$

$$H_a : \mu_2 \neq \mu_3$$

Rejection region with  $\alpha = 0.05$  in each case is reject if  $|Z| > 1.96$ . The test statistics were found and are

$$\text{Case 1: } |Z| = |1.03 - 1.520 / (0.1425)| = 3.439$$

$$\text{Case 2: } |Z| = |1.03 - 1.430 / (0.1628)| = 2.457$$

$$\text{Case 3: } |Z| = |1.52 - 1.43 / (0.186)| = 0.483$$

Our decisions from the hypothesis tests are:

Reject the null hypothesis in Case 1 and Case 2 concluding the ratios are different. we fail to reject the null hypothesis in Case 3, so we conclude the ratios for the wires on humans are statistically the same. This confirms they are different.

Previous results were weak in two areas (Fox, et al. 2010):

- (1) our probability of detection was at most approximately 85% and
- (2) our probability of false detections was high between 22-56%

We have created the wave forms of the polarization data using sinusoidal regression on the data in hopes of finding some new indicators. We obtained the following plots from the sinusoidal regression (Fox, 2011). Figure 8-12 show these results.

## 2.1 The Data

$vvdata := [2.9981, 2.568, 3.334, 2.3511, 1.0501, 2.0599, 2.3896, 2.4332, 2.0329, 3.7039, 2.5939, 2.3927, 1.846, 1.6338, 1.259, 4.8333, 3.031, 2.4694, 1.0318, 0.88241, 3.8399, 6.7319, 4.7824, 1.3003, 1.896, 2.2515, 3.1435, 1.9894, 5.0677, 1.8283];$

```
HHdata := [1.3997, 1.8211, 2.0562, 3.3373, 2.091, 1.4585,
3.2779, 0.98183, 1.2876, 0.75924, 0.79999, 2.2327, 3.3146,
3.7066, 1.6449, 1.7559, 1.3638, 2.665, 2.7001, 1.1329,
3.8667, 0.95611, 3.0781, 0.79202, 1.9711, 1.4571, 2.5922,
2.3116, 1.6376, 1.6722];
```

```
xdata := [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30];
```

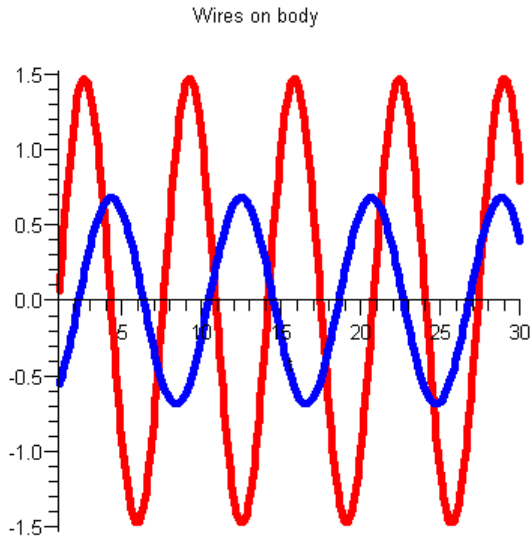


Figure 8. Waves with wires on person

```
nwvv := [-0.02289, 0.045738, 0.055329, 2.0421, 3.0388,
1.9826, 2.6181, 1.0328, 2.1586, 2.2184, 1.1283, 0.92269,
2.347, 2.4937, 5.0086, 3.3956, 3.9786, 3.6112, 0.9557,
2.1656, 2.4783, 1.7162, 1.3098, 3.4382, 0.67549, 0.53344,
3.8273, 4.5605, 1.8271, 2.5243, 2.4916, 1.4358, 1.4161,
2.4914, 1.8025, 2.5435, 1.7122, 2.5345, 2.5091, 1.3373,
4.5793, 2.4049, 6.4391, 2.7667, 2.4826, 4.3883, 3.8871,
4.3853, 3.073, 3.3666];
```

```
nwhh := [2.9902, 2.3461, 1.5768, 2.6801, 2.1605, 1.7474,
1.2864,
2.6163, 3.234, 3.546, 2.338, 3.8477, 2.2325, 3.5133,
3.5618, 2.5593, 2.3327, 2.1272, 1.6574, 2.1195, 3.5193, 2.8781,
1.5764, 1.4311, 3.2563, 3.2845, 2.564, 1.4189, 1.0782, 1.3227,
3.5949, 2.0295, 3.1405, 1.4265, 1.4768, 2.3642, 2.327, 2.0575,
1.0367, 2.6106, 3.4525, 2.6518, 2.4876, 2.4157, 4.0245,
0.87177, 2.7099, 2.25, 3.2173, 2.7642];
```

```
nwxdata := [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25,
26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41,
42, 43, 44, 45, 46, 47, 48, 49, 50];
```

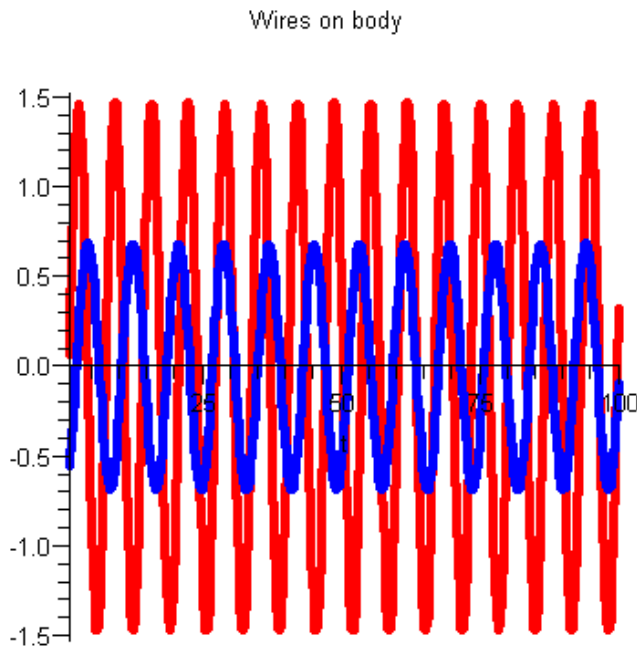


Figure 9. Waves with wires on person

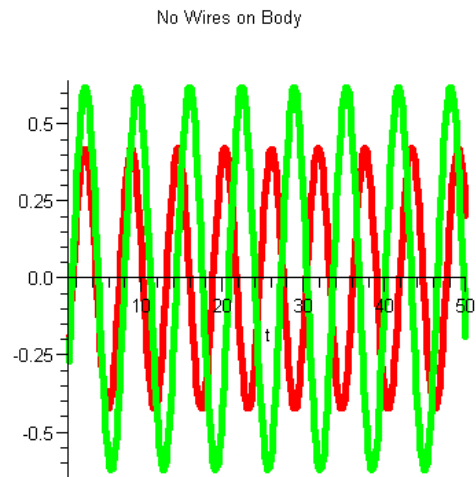


Figure 10. Waves with no wires on person



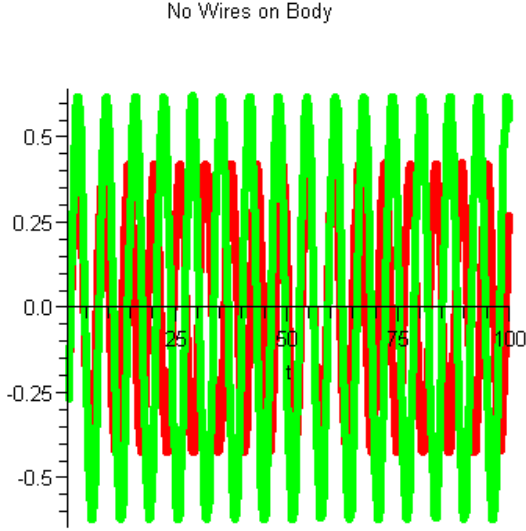


Figure 11. Waves with no wires on person

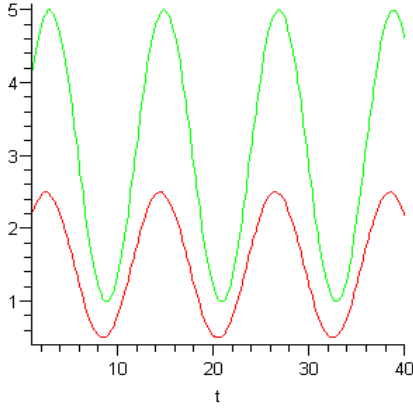


Figure 12. No wires on person.

The bottom line from analysis, confirmed with hypothesis testing at  $\alpha=0.05$ , of these plots of the sinusoidal regression are:

For persons wearing wires the periodicity is different while for persons without wires the periodicity is the approximately the same.

### 3. INDICATORS FOR METRICS FOR DETECTION (INFORM DETECTION)

#### 3.1 DETECTION METHODS and METRICS

Previously, we examined the absolute differences and the ratio of polarization. Now we move on to new constructs. Signal-to-noise ratios has been shown to be useful in detection and lowering false positives as shown by Kingsley et al (1992).

Our definition of SNR is shown in equation 1:

$$SNR = \frac{\mu}{\sigma}. \quad (1)$$

Rather than look at this metric alone, we devise a ratio of SNR ratios for polarization wave forms. Our new metrics are shown in equations 2 and 3 below:

$$dm_1 = \frac{\frac{\mu_{vv}}{\sigma_{vv}}}{\frac{\mu_{HH}}{\sigma_{HH}}} \quad (2)$$

or

$$dm_1 = \frac{\frac{\mu_{vv}}{\sigma_{vv}}}{\frac{\mu_{HH}}{\sigma_{HH}}} = \frac{\mu_{vv} \sigma_{HH}}{\mu_{HH} \sigma_{vv}} \quad (3)$$

For example, for a person without wires on their person this calculation leads to

$$dm_1 = \frac{\frac{\mu_w}{\sigma_w}}{\frac{\mu_{HH}}{\sigma_{HH}}} = \frac{\mu_w \sigma_{HH}}{\mu_{HH} \sigma_w} = \frac{2.44 \cdot 0.83779}{2.36857 \cdot 1.35613} = 0.637069.$$

For persons wearing wires for detonation purposes, we have

$$2.7757 \cdot 0.8206 / (1.824 \cdot 1.1267) = 1.10773.$$

The values are both different and are significant using a level of significance of  $\alpha = 0.05$ .

#### 3.2 Levels of Detection

##### Level 1 Detection

We introduce a concept of Level 1 detection using radar only. Level 1 detection stems from a combination of output from radar capabilities. We use the following metrics for my support matrix in Table 1:

Metric 1:  $M1 = |VV_{mean} - HH_{mean}|$

Metric 2:  $M2 = \frac{VV_{mean}}{HH_{mean}}$

Metric 3:  $dm_1 = \frac{\frac{\mu_{vv}}{\sigma_{vv}}}{\frac{\mu_{HH}}{\sigma_{HH}}} = \frac{\mu_{vv} \sigma_{HH}}{\mu_{HH} \sigma_{vv}}$

Metric 4:  $M4 = \text{Periodicity of the polarizations scaling weighting (same}=0, \text{weak}=0.5, \text{different}=1)$



Table 1. Detection Levels

Radar Scan				
Metric 1	M1			
Metric 2	0	M2		
Metric 3	0	0	dm1	
Metric 4	0	0	0	M4

The product of these values along the main diagonal yield a strength measure of the level 1 detection:

$$\text{Detection level} = M1 \cdot M2 \cdot M3 \cdot M4$$

The interpretation is as follows:

Detection Level = 0 (not a person of interest)

Detection level > 0 (a person of interest)

The larger the value the higher the interest.

M1 usually has to be greater than 0.6, M2 greater than 1.35, dm1 greater than 1 or a Detection Level > 0.8 units.

Simulation models for Level 1 detection finds the following statistics supporting our claim in Table 2. We ran the model 824,000 trials.

Table 2. Simulation Results

	$P(\text{Success})$	$P(\text{False Success})$
Mean	0.96189	0.092336
Stand error	0.00202	0.004426
Median	0.970508	0.0808
St. Deviation	0.02864	0.0626
Minimum	.875	0
Maximum	1	0.303
Count	824,000	824,000
95% CI length	0.00399	0.0087

$P(\text{Successful detection})$  is about 0.96189 and  $P(\text{False Positives})$  is about 0.092336.

Level 1 detection shows us that we have a good success rate but still need a little improvement as well as improvement is needed in the false detections. Level 1 detection is based on single radar.

Multiple radars operating in independently and orthogonal in direction (if feasible) provides the best improvement. This will improve our  $p(\text{detection})$  to greater than 99.84%. The  $P(\text{False Detections})$  is reduced to less than 1%  $\approx 0.85\%$ .

Further improvement can be made with Level 2 and Level 3 Detections described next.

### Level 2 Detection

We move on to using radar and video together. We take the level detection probabilities and combine with video of the crowd. Our video portion is examining for several quick indicators that will enhance the probabilities for detection and for false positives. Video indicators coupled after We get a Level 1 detection would include:

- (1) folds or shape in clothing over suicide vests which have been clearly modeled by previous people.
- (2) More clothing than needed based on weather and climate
- (3) Sweat or massive perspiration not caused by climate
- (4) Motion not consistent with others (saying when walking)

Each of these will be Bernoulli variable (1 if positive or 0 if negative). Any one positive indicator yields a 1 for the total. We total the values of each giving us a scale from 0 to 4.

The more one's the stronger the multiplier is for the detection.

We create a Level 2 detection metric:

$$L2DM = p(\text{success from Level 1}) \cdot \sum \text{Indicators}$$

Video will most likely be used when then we have a probability flag for a suicide bomber. The video feed then is observed and processed by a computer program for abnormal traits of movement. We use the information as we showed previously in figure 2.

### Level 3 Detection

Level 3 detection concerns using the radar to measure speed of the suspect. This can be done simultaneously with the other 2 detection strategies. We will use the work done by the Bornstein's on "Pace of Life" to measure discrepancies from the norm based on the findings of the FBI that the two suicide bombers caught were under the influence of drugs that modified their speed and perceptive abilities. First, in a region of concern we want to measure the typical speed of movement under normal circumstances. The radar gun is keyed to pick up a plus or minus one standard deviation changes in the normal speed as a flag indicator. This flag is coupled with either a strong Level 1 and/or a Level 2 indicators to better detect the person wearing wires as a suicide bomber. We use three sensors (radar cross section (RCS), video, and radar speed) for my calculations for the probability of detection.

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A) \cdot P(B) -$$

$$P(A) \cdot P(C) - P(B) \cdot P(C) + P(A) \cdot P(B) \cdot P(C)$$

for three events.

In our simulations introducing speed we find that speed needs to be coupled with other indicators to raise

its ability to predict. In a simulation model with the use of single radar with metrics discussed earlier only raised the probability of detection to approximately 100% and lowered the probability of false positive to approximately 0.5%.

#### Level 4 Detection

We consider more sensors in this level of detection to include thermal imagery and terahertz radar. My advanced algorithm uses Bayesian statistical updating analysis in order to calculate the probabilities of detection based on scans and visual sightings of the persons on interests.

For more than three events we use the inclusion - exclusion principle and the general form for four sensors of:

$$P(A_1 \cup A_2 \cup \dots \cup A_4) = P(A_1) + P(A_2) + \dots + P(A_4) - \sum_{i \neq j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_4) + P(A_2 \cap A_3 \cap A_4) + P(A_1 \cap A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_3 \cap A_4)$$

For the general case of the principle, let  $P(A_1)$ , ...,  $P(A_n)$  be finite sets of Probabilities.

$$|\bigcup_{i=1}^n P(A_i)| = \sum_{i=1}^n P(A_i) - \sum_{i,j:1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{i,j,k:1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n)$$

Each sensor adds to the probability of detection and decreases the probability of a false detection.

#### More metrics considered

We experimented with phased metrics using metrics (a) through (d) that proved to better than a single metric alone. By phased metric we mean using more than one metric in the algorithm, i.e., using two or more radar RCS metrics in the detection scheme. In a simulation, we achieved a probability of detection of approximately 99%.

The sensitivity of the device (radar) and the collection apparatus is critical. The threshold values chosen are vital to the detection algorithm. For example, the higher the probability the further away from the mean the statistic is (see our figure below). Therefore, the SE now becomes an essential element.

Only data was used from identical subjects. We determine the following baseline data shown in Table 3 and Table 4:

Table 3. Baseline data for polarization differences

Status	Polarization	Mean	SD	1 SD
--------	--------------	------	----	------

				Range
No wires	VV	2.44	.19	2.23,2.63
	HH	2.37	.11	2.26,2.48
	VV-HH	<b>0.09</b>	<b>.3</b>	<b>-.21,.39</b>
Wires (no loop)	VV	2.78	.19	2.57,2.97
	HH	1.83	.11	1.72,1.94
	VV-HH	<b>0.95</b>	<b>.30</b>	<b>0.65,1.25</b>
Wire (loop)	VV	2.87	.16	2.71,3.03
	HH	2.00	.17	1.83,2.17
	VV-HH	<b>0.87</b>	<b>0.33</b>	<b>0.54,1.20</b>

Table 4. Baseline data for polarization ratios

Status	Ratio VV/HH	Mean	SD	1 SD Range	3 SD Range
No wires	VV/HH	1.03	.12	.91,1.15	.67,1.37
Wires (no loop)	VV/HH	1.52	.15	1.37,1.67	1.07,1.97
Wires (loop)	VV/HH	1.43	.11	1.32,1.54	1.12,1.76

From a probabilistic standpoint we see that at 3 SD there is some slight overlap of values between no wires and wires in this case which are our false positives come from using only one sensor.

Enhancing our simulation to take advantage of this we find much improved results. Using both metrics together in a series fashion, the |VV-HH| and VV/HH, We found 100% of the bombers over a wider range of threshold values. We also am able to reduce the false positives to approximately 10-15%.

Video is an integral component to improve on detection. Video obtain simultaneous input that is couple with the radar infusion as shown in figure 2:

The radar becomes Flag 1 when it identifies through the combination of metrics above as potential subject. The video then analyzes the subject for deviations from the norm, approximately 1 SD. This becomes Flag 2. Two flags increase the probability of detection substantially. Adding a speed component to the radar is easy. Speed becomes the Flag 3 using the work done by the Bornstein's (1976) in walking speed of a crowd in world cities. Again, speeds that differ by approximately 1 SD are deemed critical. If all three flags are persistent then our probability of detection is over 99% and the false positive detections are less than 1% as evidence by simulation models.

Further, the addition of Thermal imagery can provide significant advantages. If the video camera or other surveillance device is added with thermal capability then we can measure the temperature change in a person. Significant temperature changes indicate that cold, hard substances are present that are different than 98.6°. Again we look for 1 SD from the mean to

create a flag. This flag help increase the probability of detection as well as decrease the probability of a false detection.

Thus, adding the other sensors, speed and video, help reduce the percentages of false positives as well as the use of thermal imagery and increases our probability of a valid detection.

The detection algorithm for a real device will be realistic modification of the simulation algorithm below:

INPUTS: N, number of runs, assumed distribution for the number of suicide bombers in a crowd, distributions for probability metric for radar detections, **threshold value**  
 OUTPUTS: the number of positive detections, the number of false detections

Step 1. Initialize all counters: detections = 0, false alarms=0, suicide bombers =0

Step 2. For i = 1,2,..., N trials do

Step 3. Generate a random number from an integer interval [a,b].

Step 4. Obtain an event of a suicide bomber based upon our hypothesized distribution of the number of suicide bombers in a crowd of size X. Basically if random number  $\leq a$  specified small value then I have a suicide bomber, otherwise I do not.

For example, I might generate random numbers between [1,300] and if the random number is  $\leq 2$  then the random number represents a suicide bomber.

Step 5. Generate characteristics for each person in the crowd by either being a bomber with random bomber characteristics or a non-bomber with random non-bomber characteristics based upon updated data collection feedback loop. I want to create a smart system.

Step 6. Allow the sensors to randomly detect the measures from Step 5 and use Step 7 to identify the characteristics based upon the metric used.

These distributions are described previously.

Step 7. Compare results from step 5-step 6 to threshold value using the following:

Target present:  $y(t) > Y \rightarrow$  correct detection

Target present:  $y(t) < Y \rightarrow$  missed detection

Target not present:  $y(t) > Y \rightarrow$  false alarm

Target not present:  $y(t) < Y \rightarrow$  no action

Step 8. For each correct detection, obtain a video and a speed input. Generate a random speed for each of the N trials above based upon Speed normal about 1 m/sec for a non-suicide bomber and Speed is  $1-.5(\text{rand}())$  or  $1+.5\text{rand}()$  for a bomber on drugs.

Step 9. Compare for detection with speed and video.

Target present:  $z(t) > Z \rightarrow$  correct detection

Target present:  $z(t) < Z \rightarrow$  missed detection

Target not present:  $z(t) > Z \rightarrow$  false alarm

Target not present:  $z(t) < Z \rightarrow$  no action

Step 10. If any are positive then use thermal imagery. Generate a random number for thermal imaging for temperature difference based upon

$$\frac{100\% \cdot (\text{temperature}_h - \text{temperature}_l)}{\text{temperature}_h}$$

Thermal difference for a normal person  
 temperature percent differential of

$$\frac{100\% \cdot (\text{temperature}_h - \text{temperature}_l)}{\text{temperature}_h}$$

using

temperature<sub>h</sub>= 98.6 and temperature<sub>l</sub> = 95

Thermal difference for a normal person  
 temperature percent differential of

$$\frac{100\% \cdot (\text{temperature}_h - \text{temperature}_l)}{\text{temperature}_h}$$

using

temperature<sub>h</sub>= 98.6 and temperature<sub>l</sub> = a random number between 70-95 degrees)

Step 11. Compare for detection by thermal imaging

Target present:  $w(t) > W \rightarrow$  correct detection

Target present:  $w(t) < W \rightarrow$  missed detection

Target not present:  $w(t) > W \rightarrow$  false alarm

Target not present:  $w(t) < W \rightarrow$  no action

Step 12. Increase all Counters as necessary

Step 13. Output statistics under the assumption of independence and use Inclusion-Exclusion as explained previously.

$$\begin{aligned} |\cup_{i=1}^n P(A_i)| = & \sum_{i=1}^n P(A_i) - \sum_{i,j:1 \leq i < j \leq n} P(A_i \cap A_j) \\ & + \sum_{i,j,k:1 \leq i < j < k \leq n} [P(A_i \cap A_j \cap A_k) - \\ & \dots + (-1)^{n-1} P(A_i \cap \dots \cap A_n)] \end{aligned}$$

END of Algorithm

Figure 13. Simulation Algorithm for Methodology Model for RCS, Radar, Video, and Thermal Imagery

## REFERENCES

- Meigs, Montgomery C. Gen(Ret), 2007 JIEDDO PowerPoint Update Report to Congress, Director, Joint IED Defeat Organization, 19 November 2007.
- Kingsley, S.,Quegan, S., 1992. *Understanding Radar Systems*, London, UK: McGraw Hill .
- Bornstein, M H., Bornstein, H.G., 1976. The Pace of Life. *Nature*, 259, 557-559.

Angell A., Rappaport, C., 2007. Computational Modeling Analysis of Radar Scattering by Clothing Covered arrays of Metallic Body-Worn Explosive Devices”, *Progress In Electromagnetics Research*, PIER 76, 285–298.

Fox, W. P., Vesecky, J., Laws, K. 2010. Detecting Suicide Bombers, *Journal of Defense Modeling and Simulation (JDMS)*, pp 1-20

Dogaru, T., Nguyen, L. Le, C., 2007. Computer models of the human body signature for sensing through the wall radar applications, Tech. Rpt. ARL-TR-4290, Army Research Laboratory .

Fox, W.P., Giordano, F., Horton, S., Weir, M. 2009. *A First Course in Mathematical Modeling*, 4<sup>th</sup> Ed. Belmont: Cengage Publishing.

Fox, W.P., 2012. *Mathematical Modeling with Maple*, Belmont: Cengage Publishing.

Gore’s Environmental Task Force as Chair of the Sensors Panel. His areas of research expertise include radar and radar systems, Earth remote sensing, ocean sensors and autonomous vehicles. He has a lifelong interest in amateur radio and holds an amateur extra class FCC license, AE6TL. He is an IEEE Fellow.

Adjunct Professor Kenneth Laws studied Physics at the University of California, finishing with a Doctorate in 2001. Since then he has done research on topics in radar applications to ocean remote sensing, emphasizing the use of HF (decimeter wavelength) radars along the coast line to measure ocean currents. The main thrust of his research has been the analysis of errors in surface current measurements and tracking ships using HF radar. He has also worked with autonomous ocean surface vehicles and renewable energy projects in the coastal environment. His teaching experience includes engineering design classes and bringing the engineering and science aspects of sustainable energy to a wide spectrum of undergraduate students.

## I

### AUTHORS BIOGRAPHY

Dr. William P. Fox is a professor in the Department of Defense Analysis at the Naval Postgraduate School. He received his BS degree from the United States Military Academy at West Point, New York, his MS at the Naval Postgraduate School, and his Ph.D. at Clemson University. Previous he has taught at the United States Military Academy and Francis Marion University where he was the chair of mathematics for eight years. He has many publications including books, chapters, journal articles, conference presentations, and workshops. He directs several math modeling contests through COMAP. His interests include applied mathematics, optimization (linear and nonlinear), mathematical modeling, statistical models for medical research, and computer simulations. He is Vice-President of the Military Application Society in INFORMS.

Professor John Vesecky studied Electrical Engineering at Rice University before attending graduate school at Stanford. After academic posts at the University of Leicester UK, Stanford and Michigan he was selected in 1999 as Founding Chairman of the Electrical Engineering Department in the new Jack Baskin Engineering School at the University of California, Santa Cruz. He has served as Chair or Associate Chair ever since. Recent teaching experience has focused on Capstone Design Courses for undergraduates and the impact of technological innovation on environmental challenges. In the 1990s he served on Vice-President

# COMPARING DECISION SUPPORT TOOLS FOR CARGO SCREENING PROCESSES

Peer-Olaf Siebers<sup>(a)</sup>, Galina Sherman<sup>(b)</sup>, Uwe Aickelin<sup>(c)</sup>, David Menachof<sup>(d)</sup>

<sup>(a, c)</sup> School of Computer Science, Nottingham University, Nottingham NG8 1BB, UK.

<sup>(b, d)</sup> Business School, Hull University, Hull HU6 7RX, UK.

<sup>(a)</sup> [pos@cs.nott.ac.uk](mailto:pos@cs.nott.ac.uk), <sup>(b)</sup> [g.sherman@2008.hull.ac.uk](mailto:g.sherman@2008.hull.ac.uk), <sup>(c)</sup> [uxa@cs.nott.ac.uk](mailto:uxa@cs.nott.ac.uk), <sup>(d)</sup> [d.menachof@hull.ac.uk](mailto:d.menachof@hull.ac.uk)

## ABSTRACT

When planning to change operations at ports there are two key stake holders with very different interests involved in the decision making processes. Port operators are attentive to their standards, a smooth service flow and economic viability while border agencies are concerned about national security. The time taken for security checks often interferes with the compliance to service standards that port operators would like to achieve.

Decision support tools as for example Cost-Benefit Analysis or Multi Criteria Analysis are useful helpers to better understand the impact of changes to a system. They allow investigating future scenarios and helping to find solutions that are acceptable for all parties involved in port operations.

In this paper we evaluate two different modelling methods, namely scenario analysis and discrete event simulation. These are useful for driving the decision support tools (i.e. they provide the inputs the decision support tools require). Our aims are, on the one hand, to guide the reader through the modelling processes and, on the other hand, to demonstrate what kind of decision support information one can obtain from the different modelling methods presented.

**Keywords:** port operation, service standards, cargo screening, scenario analysis, simulation, cost benefit analysis, multi criteria analysis

## 1. INTRODUCTION

Businesses are interested in the trade-off between the cost of risk mitigation and the expected losses of disruptions (Kleindorfer and Saad 2005). Airports and seaports face an additional complexity when conducting such risk analysis. In these cases there are two key stake holders with different interests involved in the decision processes concerning the port operation or expansion (Bichou 2004).

On the one hand we have port operators which are service providers and as such interested in a smooth flow of port operations as they have to provide certain service standards (e.g. service times) and on the other hand we have the border agency which represent national security interests that need to be considered. Checks have to be conducted to detect threats such as

weapons, smuggling and sometimes even stowaways. If the security checks take too long they can compromise the service standard targets to be achieved by the port operators. Besides these two conflicting interest there is also the cost factor for security that needs to be kept in mind. Security checks require expensive equipment and well trained staff. However, the consequences for the public of undetected threats passing the border can be severe. It is therefore in the interest of all involved parties to find the right balance between service, security, and costs.

But how can we decide the level of security required to guarantee a certain threshold of detection of threats while still being economically viable and not severely disrupting the process flow? A tool frequently used by business and government officials to support the decision making is Cost-Benefit Analysis (CBA) (Hanley and Spash 1993). While CBA is useful in our case to find the right balance between security and costs it struggles to provide decision support for the consideration of service quality. This is due to the fact that service quality is difficult to be expressed in monetary terms. Multi Criteria Analysis (MCA) is a tool that allows taking a mixture of monetary and non monetary inputs into account. It can use the results of a CBA as monetary input and service quality estimators as non monetary input and produce some tables and graphs to show the relation between cost/benefits of different options (DCLG 2009).

Different modelling methods can be used to conduct CBA. Damodaran (2007) lists Scenario Analysis (SA), Decision Trees (DT) and Monte Carlo Simulation (MCS) as being the most common ones in the field of Risk Analysis. Less frequently used in Risk Analysis but often used in Operational Research to investigate different operational practices is Discrete Event Simulation (DES) (Turner and Williams 2005; Wilson 2005). Depending on the world view the modeller adopts DES models can either be process oriented or object oriented (Burns and Morgeson 1988). Besides being useful for estimating factors required for CBA, DES models also allow to investigation how well service standards are reached in the system under investigation as it considers delays that one experiences while moving through the system (Laughery et al 1998). It is therefore well suited, in conjunction with CBA, to

provide all the inputs required for a MCA. However, sometimes it is even possible to find a solution that does not require any investment. It might be feasible to achieve the goal simply by changing certain working routines. In such cases DES might be able to provide the information required for making a better informed decision without the need to conduct a full MCA.

In previous work (Sherman et al 2010) we compared the efficiency of different methods for conducting CBA by using the same case study. This strategy allowed us to contrast all modelling methods with a focus on the methods themselves, avoiding any distortions caused by differences in the chosen case studies. In this paper we continue our investigation but this time we focus on the two competing DES approaches and what information they can provide to assist our analysis.

The remainder of the paper is structured as follows. In Section 2 we introduce our case study system, the Port of Calais. In Section 3 we show in detail how to conduct a CBA using SA for our case study system. In Section 4 we discuss the additional features DES has to offer and we explain how to implement a model of the case study system using different world views. Once we discussed the specific features of the different implementations, we demonstrate by an experiment how to use DES to provide decision support. In Section 5 we summarise the findings from Sherman et al (2010) and this paper in form of a table that shows real world phenomena that can be modelled with the different modelling methods, the data requirements, and the decision support information that is provided. Finally we suggest some future research activities.

## 2. CASE STUDY

Our case study involves the cargo screening facilities of the ferry Port of Calais (France). In this area of the port there are two security zones, one operated by French authorities and one operated by the UK Border Agency (UKBA). The diagram in Figure 1 shows the process flow in this area.

According to the UKBA, between April 2007 and April 2008 about 900,000 lorries passed the border and approximately 0.4% of the lorries had additional human freight (UKBA 2008). These clandestines as they are called by the UKBA are people who are trying to enter the UK illegally - i.e. without having proper papers and

documents.

The search for clandestines is organised in three major steps, one by France and two by the UKBA. On the French site all arriving lorries are screened, using passive millimetre wave scanners for soft sided lorries and heartbeat detectors for hard sided lorries. If lorries are classified as suspicious after the screening further investigations are undertaken. For soft sided lorries there is a second test with CO2 probes and if the result is positive the respective lorry is opened. For hard sided lorries there is no second test and they are opened immediately.

Although 100% of lorries are screened at the French site, not all clandestines are found. This shows that the sensor efficiency in the field is less than 100%. Unfortunately it is not known for any of the sensors how much less their efficiency is and field tests cannot be undertaken as it would be unethical to deliberately lock someone in the back of a lorry. Another problem with estimating the sensor efficiency is that the sensor data has to be interpreted by human operators, who might misinterpret them. Again, no data exist about the efficiency of operators.

On the British site only a certain percentage of lorries (currently 33%) is searched at the British sheds. Here a mixture of measures is used for the inspection, e.g. CO2 probes, dogs, and opening lorries. Once the lorries passed the British sheds they will park in the berth to wait for the ferry. In the berth there are mobile units operating that search as many of the parked lorries as possible before the ferry arrives, using the same mixture of measures than in the sheds. As shown in Table 1 only about 50% of the clandestines detected were found by the French, about 30% in the sheds and 20% by the mobile units in the berth. The overall number of clandestines that are not found by the authorities is of course unknown.

Table 1: Statistics from Calais

Statistic	Value
Total number of lorries entering Calais harbour	900,000
Total number of positive lorries found	3474
Total number of positive lorries found on French site	1,800
Total number of positive lorries found on UK site	1,674
... In UK Sheds	890
... In UK Berth	784

The key question is: What measures should we

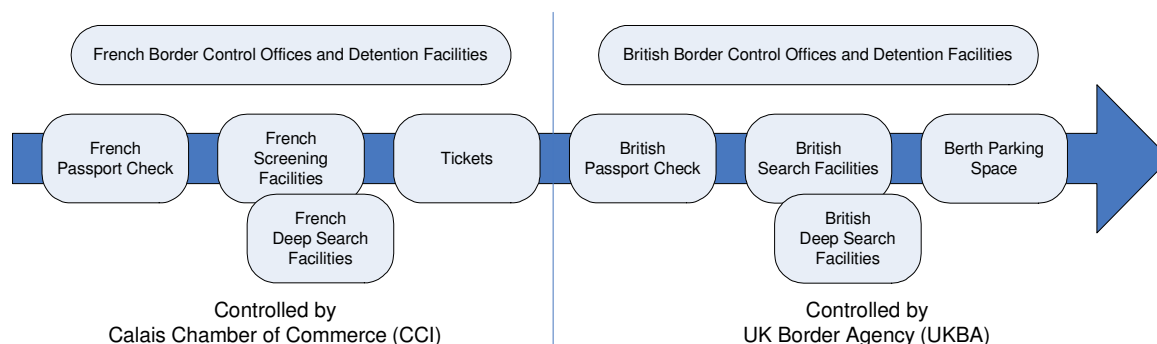


Figure 1: Border control operations at Calais

employ to reduce the overall number of clandestines that make it through to the UK? One way to improve detection rates could be to intensify the search activities. As we can see, clandestines are found at all stages of the cargo screening process and we can be sure that not all clandestines will be found in the end. However, when increasing search activities we also have to consider the disruptions this might cause to traffic flow. As mentioned in Section 1 we have different stakeholders with different interests involved in the decision making process. The two key objectives on which to base the decision are as follows: minimise costs (for the tax payer) and maximise service quality (by minimising waiting times and disruptions)

### 3. USING CBA & SA FOR DECISION SUPPORT

CBA seeks to value the expected impacts of an option in monetary terms (DCLG 2009). It involves comparing the Total Expected Costs (TEC) of each option against the total expected benefits, to see whether the benefits outweigh the costs, and by how much (Nas 1996). The aim is to determine the efficiency of the interventions relative to each other and the status quo. In our case total expected costs comprise the investments we have to make to increase the search activities. This might include things like increasing staff level, staff training, getting new sensors with better technology or building new search facilities. The total expected benefits will be the money that is saved for each clandestine that does not make it into the UK. Clandestines that made it to the UK are very likely to work illegally and therefore causing some income tax losses. Furthermore, they will not pay their contributions to health insurance and pensions. Therefore, the government will have to look after them once they are ill or old.

Uncertainty in the CBA parameters is often evaluated using a sensitivity analysis, which shows how the results are affected by changes in the parameters. In our case we have one parameter for which it is impossible to collect any data: the number of positive lorries that make it into the UK. A positive lorry is a lorry that has at least one clandestine on board. Usually clandestines attempt to cross the border in small groups. Conducting a sensitivity analysis cannot solve this problem but it can show the impact of changing the key parameter on the decision variable and can consequently provide some additional information to aid the decision making process and improve confidence in the decision made in the end.

SA is the process of analysing possible future events by considering alternative possible outcomes (Refsgaarda et al 2007). To define our scenarios we consider the following two factors: Traffic Growth (TG) and Positive Lorry Growth (PLG). For each of the factors we define three scenarios. Table 2 shows the factors and scenarios investigated and an estimate of the likelihood for each of the scenarios to occur in the future. A justification for the scenario definitions can be found in the following two paragraphs.

Table 2: Two factors with three scenarios each and their probability of occurrence

Factor 1	TG	p(TG)
Scenario 1	0%	0.25
Scenario 2	10%	0.50
Scenario 3	20%	0.25
Factor 2	PLG	p(PLG)
Scenario 1	-50%	0.33
Scenario 2	0%	0.33
Scenario 3	25%	0.33

Our TG scenarios are based on estimates by the port authorities who are planning to build a new terminal in Calais in 2020 to cope with all the additional traffic expected. According to DHB (2008) between 2010 and 2020 the traffic in the Port of Dover is expected to double. We assume that this is also applicable to the Port of Calais and have therefore chosen the 10% per annum growth scenario as the most likely one, while the other two are equally likely.

Our PLG scenarios are reflecting possible political changes. The number of people trying to enter the UK illegally depends very much on the economical and political conditions in their home countries. This is difficult to predict. If the situation stabilises then we expect no changes in the number of attempts to illegally enter the UK. However, as our worst case scenario we assume a 25% growth. Another factor that needs to be considered is that trafficking attempts very much depend on the tolerance of the French government to let clandestines stay nearby the ferry port while waiting for the opportunity to get on one of the lorries. However, it is currently under discussion that the French authority might close the camps where clandestines stay which would reduce the number of potential illegal immigrants drastically. Therefore we added a scenario where the number of attempts is drastically reduced by 50%. As there is currently no indication of which of these scenarios is most likely to occur we have assigned them the equal probabilities of occurrence. We will assume that any changes in clandestine numbers will proportionally affect successful and unsuccessful clandestines.

The question that needs to be answered here is how the UKBA should respond to these scenarios. We assume that there are three possible responses: not changing the search activities, increasing the search activities by 10% or increasing the search activities by 20%. For the CBA Search Growth (SG) is our primary decision variable.

The cost for increasing the search activities in Calais is difficult to estimate, as there is a mixture of fixed and variable cost and operations are often jointly performed by French, British and private contractors. However, if we concentrate on UKBA's costs, we can arrive at some reasonable estimates, if we assume that any increase in searches would result in a percentage increase in staff and infrastructure cost. Thus we estimate that a 10% increase in search activity would cost £5M and a 20% increase £10M.



Now we need to estimate the benefits we would expect when increasing the search activities. First we need to derive a figure for the number of Positive Lorries Missed (PLM) and how much each of these lorries cost the tax payer. A best guess of “successful” clandestines is approximately 50 per month (600 per year). With an average of four clandestines on each positive lorry an estimated 150 positive lorries are missed each year. It is estimated that each clandestine reaching the UK costs the government approx. £20,000 per year. Moreover, UKBA estimates that the average duration of a stay of a clandestine in the UK is five years, so the total cost of each clandestine slipping through the search in Calais is £100,000, resulting in £400,000 per PLM.

It is probably a fair assumption that an increase in searches will yield a decrease in the number of positive lorries and an increase in traffic will yield an increase in PLM. In absence of further information we assume linear relationships between the two parameters. Table 3 shows the number of PLM assuming there is no PLG. Equation 1 has been used to produce the table.

$$PLM(TG,SG)=PLM*(1+TG)/(1+SG) \quad (1)$$

Table 3: PLM for (PLG=0)

PLG 0%	SG 0%	SG +10%	SG +20%
TG 0%	150.00	136.36	125.00
TG 10%	165.00	150.00	137.50
TG 20%	180.00	163.64	150.00

With this information we can now calculate the Economic Cost (EC) for all the different scenarios (see Table 4) using Equation 2:

$$EC(TG,SG,PLG)=PLM(TG,SG)*(1+PLG) \quad (2)$$

Table 4: EC for different SG options

SG 0%	PLG -50%	PLG 0%	PLG 25%
TG 0%	£30,000,000	£60,000,000	£75,000,000
TG 10%	£33,000,000	£66,000,000	£82,500,000
TG 20%	£36,000,000	£72,000,000	£90,000,000
SG 10%	PLG -50%	PLG 0%	PLG 25%
TG 0%	£27,272,727	£54,545,455	£68,181,818
TG 10%	£30,000,000	£60,000,000	£75,000,000
TG 20%	£32,727,273	£65,454,545	£81,818,182
SG 20%	PLG -50%	PLG 0%	PLG 25%
TG 0%	£25,000,000	£50,000,000	£62,500,000
TG 10%	£27,500,000	£55,000,000	£68,750,000
TG 20%	£30,000,000	£60,000,000	£75,000,000

To be able to calculate the benefit we need to know the combined probabilities of each scenario’s likelihood to occur (listed in Table 2). We get this by multiplying the probabilities of the individual scenarios as shown in Equation 3. The results of these calculations can be found in Table 5.

$$p(TG,PLG)=p(TG)*p(PLG) \quad (3)$$

Table 5: Combined probabilities

	PLG -50%	PLG 0%	PLG 25%
TG 0%	0.0833	0.0833	0.0833
TG 10%	0.1667	0.1667	0.1667
TG 20%	0.0833	0.0833	0.0833

Now we multiply the EC from Table 4 with the probabilities from Table 5 to receive the TEC for each SG option, using Equation 4. The results are shown in Table 6. The final step is to calculate the Net Benefit (NB) by using SG=0 as the base case. The NB can be calculated using Equation 5 (where C = cost for SG).

$$TEC(SG)=\sum(EC(SG,TG,PLG)*p(TG,PLG)) \quad (4)$$

$$NB(SG)=TEC(SG=0)-TEC(SG)-C(SG) \quad (5)$$

Table 6: CBA for different SG options

Option	1	2	3
SG	0%	10%	20%
TEC	£60,500,000	£55,000,000	£50,416,667
C	£0	£5,000,000	£10,000,000
NB	£0	£500,000	£83,333

The results of the CBA suggest that there is a small benefit in choosing option 2. However, we need to keep in mind that the calculations are based on a lot of assumptions. Therefore, a small difference in the NB might just be random noise. In particular we have one factor that we do not know anything about - PLM. In order to learn more about the impact of this factor we can conduct a sensitivity analysis. Running CBA for several different PLMs reveals that for a higher PLM option 3 would give us the most benefits but for a lower PLM option 1 would be the best choice. The sensitivity analysis shows how important it is to get a good estimate of the PLM.

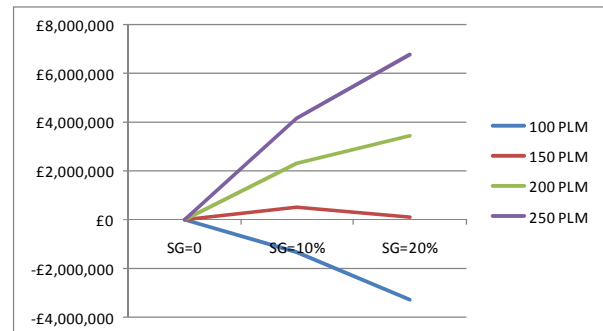


Figure 2: Sensitivity analysis results

If nothing else, the process of thinking through scenarios is a very useful exercise. In our case SA helped us to define the scenarios of interest. The sensitivity analysis later helped us to understand the importance of PLM when choosing an option. Overall SA only allows investigating a small number of factors but it seems to be a useful tool for structuring the overall investigation and get a first estimate regarding the benefits that one could gain from the different option.



#### 4. USING DES FOR DECISION SUPPORT

The main benefit of DES models is that time and space can be taken into account which allows us for the first time to assess service quality (in terms of waiting time) and consider real world boundaries (e.g. space limitations for queues). As we said before our goal is to find a balance between service quality and security. CBA on its own or in conjunction with any of the methods introduced before does not provide us with information about the impact of changes on service quality. Another benefit we get by using DES modelling is that it simplifies adding more operational details about the real system and better supports the utilisation of real world data which both make the models themselves and results more credible.

In the following two sub sections we describe two different DES modelling approaches: Process Oriented DES (PO DES) and Object Oriented DES (OO DES). The differences between these methods will be described later in the subsequent sections. Here we look at commonalities. For both modelling approaches we need additional data regarding arrival rates, cycle times (time lorries spend in a shed for screening), space availability between stations for queuing, and resources for conducting the searches.

In order to be able to emulate the real arrival process of lorries in Calais we created hourly arrival rate distributions for every day of the week from a year worth of hourly number of arrival records that we received from UKBA. These distributions allow us to represent the real arrival process, including quiet and peak times. In cases where this level of detail is not relevant we use an exponential distribution for modelling the arrival process and the average arrival time calculated from the data we collected as a parameter for this distribution.

The cycle times are based on data that we collected through observations and from interviews with security staff. In order to represent the variability that occurs in the real system we use different triangular distribution for each sensor types. Triangular distributions are continuous distributions bounded on both sides. In absence of a large sample of empirical data a triangular distribution is commonly used as a first approximation for the real distribution (XJTEK 2005). Every time a lorry arrives at a shed a value is drawn from a distribution (depending on the sensor that will be used for screening) that determines the time the lorry will spend in the shed.

We did not put any queue size limits into our case study simulation model. However, we have a run time display variable for each individual queue that displays the maximum queue length of that queue. In this way we can observe which queue is over its limit without interrupting the overall simulation run. If necessary, queue length restrictions could easily be added. In fact, in one of our experiments we restrict the capacity of UK shed queue and let lorries pass without searching them if the queue in front of the shed is getting too long. This

strategy improves the service quality but has a negative impact on security. Simulation allows us to see how big the impact of this strategy is in the different scenarios.

We also added some more details to the UK berth operation representation. We now consider the (hourly) arrival of the ferry boat. When the ferry arrives all search activities in the berth area are interrupted and all lorries currently in the berth area are allowed to board the ferry (as long as there is enough space on the ferry), regardless if they have been checked or not. Again, this strategy improves the service quality but has a negative impact on security. This is an optional feature of the DES model that can either be switched on or off.

Finally, in DES we can consider resources in a more detailed way. While previously resources have only been playing a role as a fixed cost factor (cost for SG), we can now take into account how different resource quantities influence the process flow and subsequently the service quality. These quantities can vary between different scenarios, but also between different times of day (e.g. peak and quiet times).

Clearly, as we can see from the above, DES helps to improve the decision making process. It allows besides the monetary estimates to get information on service quality and gain further insight into system operations. This additional insight can be used for decision making but also for optimising system operations. DES also allows you to conduct a sensitivity analysis to find high impact factors that need to be estimated more carefully. Of course this does not come without costs. DES models usually take much longer to build and need additional empirical data.

##### 4.1. DES using a process oriented world view

As before, we use the standard procedure for calculating the TEC for the different SG options that enables us to conduct a basic CBA. For this the PO DES delivers the number of Positive Lorries Found (PLF) which allows us to calculate the number of PLM under the assumption that there is a linear relationship between these two measures. This number can then be used as an input for the CBA.

On a first view the PO DES model looks very similar to the MCS model presented in Sherman et al (2010) and in fact it is an extended version of this model. In addition to the routing information defined in the MCS model we have added arrival rates, cycle times for screening the lorries and available resources. The data required for implementing these additions into the model has been provided by UKBA.

We now have a stochastic and dynamic simulation model that allows us to observe the evolution of the system over time. This is a big benefit compared to the static models we used previously as it allows us for the first time to consider our second objective in the analysis - to provide high quality service. One of the key factors for providing high quality service is to keep average waiting times below a specified threshold (service standard). By adding arrival rates, cycle times and available resources the model is able to produce

realistic waiting time distributions, which gives us an indication of the service quality we are achieving with different parameter settings.

Another useful output that PO DES provides is resource utilisation. This information allows us to fine-tune our CBA as we are able to better estimate SG costs. So far we have assumed that SG has a fixed cost which linear correlated with TG. In reality however, the cost for SG might depend on the utilisation of the existing resources. If in the current situation facility and staff utilisation is low then SG can perhaps be implemented without any additional costs. PO DES allows to test at what level of TG additional resources are required (throughput analysis).

The PO DES model also allows us to analyse queue dynamics. A useful statistic to collect in this context is the "maximum queue sizes" which enables us to find bottlenecks in our system. With this information we can then pinpoint where in the system we should add additional resources to achieve maximum impact. Removing bottlenecks improves the system flow and consequently service quality.

Another interesting feature of our simulation model is that it allows us to represent temporal limited interventions and see what their effect is on system flow and detection rates of positive lorries. These procedures could be officers speeding up when queues are getting longer (missing more positive lorries) or changing search rates over time (less search at peak times) or stopping search completely once a certain queue length has been reached. PO DES can help us to find out, which of these strategies is best.

However, the process oriented modelling approach as described above has some limitations with regards to flexibility. We are using proportions for defining the vehicle routing (and detection rates), based on historic data. This assumes that even if we have changes in the system these proportions would always remain the same. While this is acceptable for many situations (in particular for smaller changes) there are occasions where this assumption does not hold. For example, if we change the proportion of vehicles searched in the berth from 5% to 50% we cannot simply assume that the detection rate grows proportionally. While it might be easy to spot some positive lorries as they might have some signs of tampering (so the detection rate for these is very high and these where the ones reported in the historic data), it will get more difficult once these have been searched and a growth in search rate does not yield an equivalent growth in detection rate any more. This needs to be kept in mind when using a PO DES approach.

Furthermore, the assumption of a linear relationship between number of PLF and number of PLM which is one of our key assumptions for the CBA is quite weak in connection with PO DES as this relationship can be influenced by some of the interventions. For example, the temporal limited intervention of stopping search completely once a certain queue length has been reached influences the overall number of clandestines in

the system (when calculating this value by adding up the number of PLF and number of PLM), although this number should not be influenced by any strategic interventions. This needs to be considered in the analysis of the results..

#### **4.2. DES using an object oriented world view**

OO DES has a very different world view compare to PO DES. Here we model the system by focusing on the representation of the individual objects in the system (e.g. lorries, clandestines, equipment, or staff) rather than using predefined proportions for the routing of entities based on historic data. In fact we only use historic data as inputs (arrival rates, search rates, staffing) and for validation purposes (numbers of PLF at the different stages). The system definition is now based on layout information and assumptions on sensor capabilities. Taking this new perspective means that we transfer all the "intelligence" from the process definition into the object definition and therefore change our modelling perspective from top down to bottom up.

Unlike in DT, MCS, and PO DES we do not use a probabilistic framework for directing lorries (and deciding which of the lorries are positive). Instead we aim to recreate processes as they appear in the real system. At the entrance of the port we mark a number of lorry entities as positive (to simulate clandestines getting on board to lorries). This gives us complete control over the number of positive lorries entering the system. As the lorries go through the system they will be checked at the different stages (French side, UK sheds, UK berth). For these checks we use sensors that have a specific probability of detecting true and false positives. Only lorries that have been marked positive earlier can be detected by the sensors as true positives. The marked lorries that are not detected at the end (port exit) are the ones that will make it through to the UK. Officers can decide how to use the detectors (e.g. speed up search time if queues are getting to long, change search rates, etc.) depending on environmental conditions.

One of the advantages of this simulation method is that it is much easier to manipulate object and system parameters, like for example the sensor detection rates and search rates. When we change these parameters we do not have to worry about linear dependencies any more. In fact, we can find out relationships between variables by looking at the evolution of the system over time. However, the most interesting thing here is that due to the fact that we model sensors as objects with each having individual detection rates the number of PLM becomes an output of the simulation. Yet, we have to keep in mind that this output is still a function of unknowns:  $PLM = f(\text{lorries marked positive, sensor detection rates ...})$ . But overall, this is very useful information as we can now do some what-if analysis and see the trends of how changes in the system setup impact on number of PLM. We do not rely on the implicit assumption of a linear relationship between

PLM and SG any more. While this is not directly solving our problem of estimating how many positive lorries we miss it gives us some additional information about system behaviour that might help us to decide in one way or another.

### 4.3. Experimentation with the OO DES model

In this sub section we want to show how we can use our DES model to test alternative scenarios. We have implemented our model in AnyLogic™ 6.6, a Java™ based multi-paradigm simulation software. Figure 3 shows a screenshot of a section of the simulation model (berth area) during execution.

To set up our OO DES we tried to reproduce the base scenario (as defined in Table 1) as closely as possible by calibrating our model to produce the correct values for the number of PLF at the different stages of the search process. We can do this by varying the number of positive lorries entering the port, the sensor detection rates, and the berth search rate. The results of the calibration exercise are presented in Table 7 (Scenario 1). To get somewhere close to the real PLF values at the different stages we had to increase the number of positive lorries entering the port. Hence, also the PLM value is much higher than the best guess we used in our CBA. The detection rates for the UK sheds and the UK berth had to be much higher than the ones on the French side, in order to match the true rates. We assume that in the real system this is due to the fact that UKBA uses some intelligence when choosing the lorries to be screened. Therefore their success rate is

much higher. In particular in the berth officers can drive around and pick suspicious lorries without any time pressure.

Our scenarios are defined by TG and SG. All other parameters (grey) depend on these values (with the exception of the queue size restriction). All dependencies are explained in Section 3. For this experiment we assume that there is no change in the number of people trying to get into the UK (PLG=0). In Table 7 we only show the changes in the scenario setup. Empty fields mean that there is no change in the set-up for that specific parameter.

For this experiment we have defined a service standard that needs to be achieved. We have a threshold time in system that should not be exceeded by more than 5% of the lorries that go through the system. We have one intervention that allows us to influence the process flow. We can define a threshold for the maximum queue size in front of the UK sheds (queue size restriction). If this threshold is exceeded lorries are let pass without screening. While this intervention improves the flow there is a risk that more positive lorries are missed as less lorries are inspected.

The first three scenarios deal with TG. There is no problem in regards to compliance with service standards. Resource utilisation does not change as the number of searches does not change. However, the number of PLF is decreasing while the number of PLM is going up. The next two scenarios deal with SG. The increase in search activities causes some delays and at a

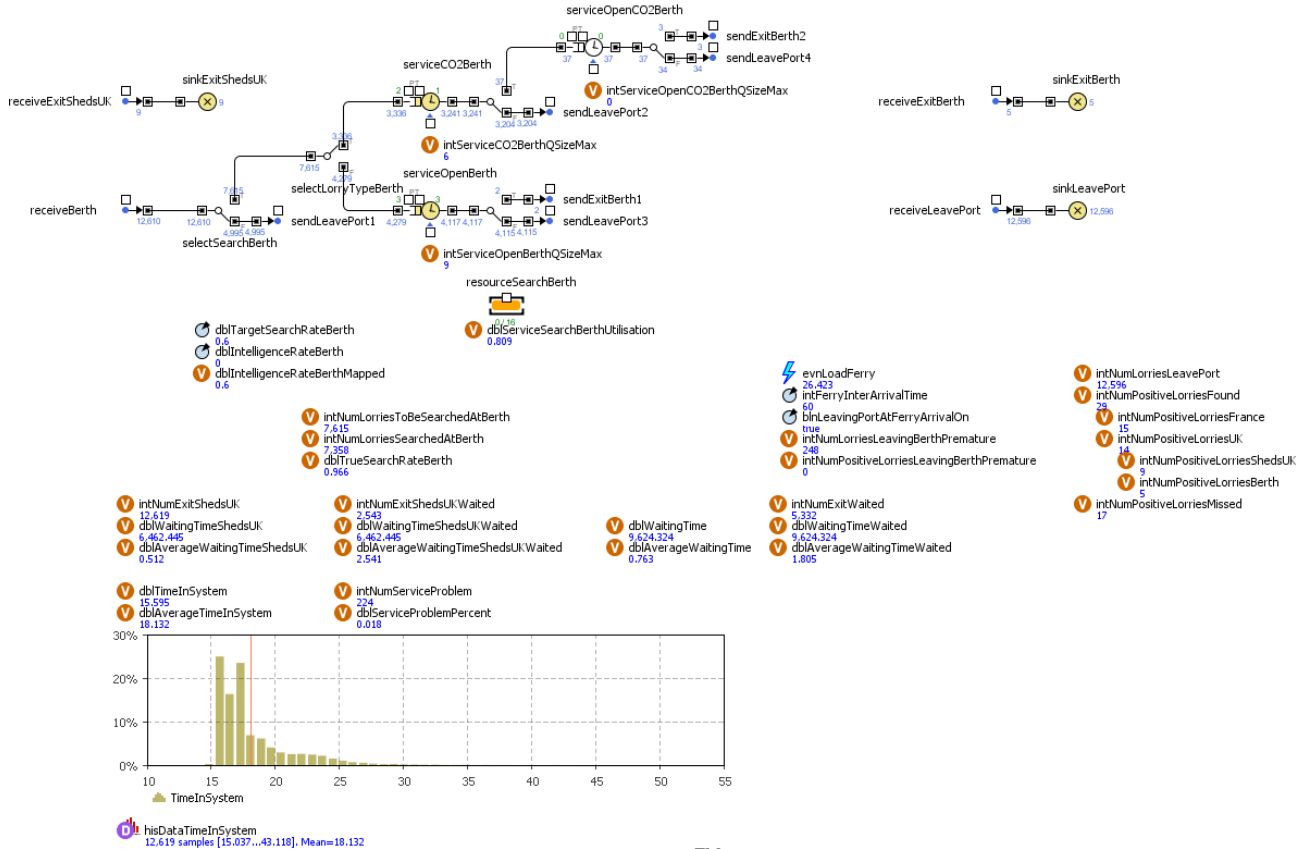


Table 7: OO DES simulation experiment set-ups and results (10 replications)

Scenarios		1	2	3	4	5	6	7
Traffic Growth (TG)		0%	10%	20%	0%			
Search Growth (SG)		0%			10%	20%		
Lorries	Arrivals	900000	990000	1080000	900000			
	Soft-sided	0.44						
	Positive	0.00550	0.00500	0.00458	0.00550			
Search rate	UK Sheds	0.330	0.300	0.275	0.363	0.396		
	UK Berth	0.600	0.545	0.500	0.660	0.720		
Detection Rates	France	0.41						
	UK Sheds	0.80						
	UK Berth	0.95						
Queue size restriction	UK Sheds	off					10	9
Results		1	2	3	4	5	6	7
Waiting times (avg) <sup>(1)</sup>	France	0.858	1.019	1.268	0.863	0.859	0.860	0.863
	UK Sheds	2.612	2.474	2.321	3.452	5.046	3.940	3.763
	Overall	1.831	1.783	1.856	2.439	3.620	2.901	2.788
Time in system (avg)		18.099	18.085	18.155	18.517	19.274	18.893	18.834
Service problem		0.019	0.019	0.020	0.036	0.068	0.052	0.049
Resource utilisation	UK Sheds	0.676	0.676	0.677	0.744	0.812	0.803	0.801
	UK Berth	0.808	0.808	0.809	0.868	0.915	0.914	0.914
Positive lorries	France	1774.9	1765.5	1745.9	1780.5	1774.3	1757.5	1769.7
	UK Sheds	900.8	814.0	733.8	981.2	1078.0	1061.2	1042.8
	UK Berth	699.9	658.4	630.7	715.9	743.0	746.5	746.8
	Missed	1590.1	1697.2	1797.0	1480.7	1365.7	1361.7	1358.1

SG of 20% the system does not comply with service standards any more. On the other hand, as expected, an increase in search activities improves the number of PLF and reduces the number of PLM. Scenario 5 indicates that service standards cannot be achieved with the current staff/facilities and that an investment has to be made in order to reduce the number of PLM and comply with service standards. However, the last scenario shows that there is also a strategic solution possible that does not require any investment. By managing the queues in front of the UK sheds it is possible to reduce the number of service problems to fewer than 5% (compliant with service standards) while still keeping the number of PLM at a very low level. Therefore, when applying this intervention no investment is required.

As we have found a solution for our problem that does not require balancing costs and benefits there is no need to conduct a MCA. However, for other scenarios where we are not so lucky we might want to consider using MCA. A good guide to MCA with a worked example is provided by DCLG (2009).

## 5. CONCLUSIONS

To help the managers and analysts to decide which method best to use for supporting their decision processes concerning port operation and expansion it is important to be clear about the real world phenomena that can be considered with the different methods and the decision support information that can be obtained by applying them. Table 8 list these for all the methods discussed in Sherman et al (2010) and in this paper.

The methods can be divided in two categories: static and dynamic. For static methods the passage of time is not important (Rubinstein and Kroese 2008).

SA, DT, and MCS belong to this category. The lack of a concept of time in these methods makes it impossible to analyse service quality of a system as all performance measures for such analysis rely on the passage of time. On the contrary, dynamic methods consider the passage of time. DES belongs to this category and allows evaluate a system's compliance with service standards. Besides, it provides other useful information about the dynamics of the system that can be used for optimising processes and improving flows.

DES can be implemented in different ways, either as PO DES or OO DES. In our experience PO DES seems to be easier to implement but less flexible (easier to manipulate). OO DES seem to be more flexible but also more difficult to implement. However, it has some advantages. In our case OO DES was the only tool that also provided an estimate of the number of PLM, which helps us to better evaluate the effect of different interventions.

For evaluating the trade-off between security and cost CBA is a valuable tool. If we want to balance security, cost and service then MCA is a better choice. While CBA relies on monetary inputs, MCA allows using monetary and non-monetary inputs. This is useful as service quality is difficult to capture by monetary measures. However, sometimes none of these tools is required for the analysis as the answer might come directly from the model as we have shown in Section 4.3.

A natural extension of the OO DES modelling approach would be to add "intelligent" objects that have a memory, can adapt, learn, move around, and respond to specific situations. These could be used to model officers that dynamically adapt their search strategies based on their experiences, but also clandestines

Table 8: Comparison of modelling methods regarding real world phenomena that can be considered in the model (black) and decision support information that can be obtained from the model (red)

SA	DT	MCS	PO DES	OO DES
Scenarios (factors and the decision variable)	Scenarios (factors and the decision variable)	Scenarios (factors and the decision variable)	Scenarios (factors and the decision variable)	Scenarios (factors and the decision variable)
Linear relationships	Linear relationships	Linear relationships	Linear relationships	-
TEC	TEC	TEC	TEC	TEC
PLM, PLF	PLM, PLF	PLM, PLF	PLM, PLF	PLM, PLF
	System structure	System layout	System layout	System layout
	Existing resources	Existing resources	Existing resources	Existing resources
		System variability	System variability	System variability
			Service time distributions	Service time distributions
			Resource utilisation	Resource utilisation
			Dynamic system constraints (e.g. peak times)	Dynamic system constraints (e.g. peak times)
			System throughput	System throughput
			Waiting times (service quality)	Waiting times (service quality)
			Time in system	Time in system
			Bottleneck analysis	Bottleneck analysis
			Dynamic decisions by system	Dynamic decisions by objects
				Sensor detection rates
				Number of pos. lorries entering the system

learning from failed attempts and improving their strategies when trying again. We are currently working on implementing such “intelligent” objects.

## REFERENCES

- Bichou, K., 2004. The ISPS code and the cost of port compliance: An initial logistics and supply chain framework for port security assessment and management. *Maritime Economics and Logistics*, 6(4), 322-348.
- Burns, J.R. and Morgeson, J.D., 1988. An object-oriented world-view for intelligent, discrete, next-event simulation. *Management Science*, 34(12), 1425-1440.
- Damodaran, A., 2007. *Strategic Risk Taking: A Framework for Risk Management*. New Jersey: Wharton School Publishing.
- Department for Communities and Local Government, 2009. *Multi-Criteria Analysis: A Manual*. London: DCLG. Available from: [http://eprints.lse.ac.uk/12761/1/Multi-criteria\\_Analysis.pdf](http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf) [accessed 19 July 2011].
- Dover Harbour Board, 2008. *Planning for the next generation - Third round consultation document*. Dover: DHB. Available from [http://www.doverport.co.uk/\\_assets/client/images/collateral/dover\\_consultation\\_web.pdf](http://www.doverport.co.uk/_assets/client/images/collateral/dover_consultation_web.pdf) [accessed 20 July 2011]
- Hanley, N. and Spash, C.L., 1993. *Cost-Benefit Analysis and the Environment*. Hampshire: Edward Elgar Publishing Ltd.
- Kleindorfer, P.R. and Saad, G.H., 2005. Managing disruption risks in supply chains. *Production and Operations Management*, 14 (1), 53-98.
- Laughery, R., Plott, B. and Scott-Nash, S., 1998. Simulation of Service Systems. In: J Banks, ed. *Handbook of Simulation*. Hoboken, NJ: John Wiley & Sons.
- Nas, T.F., 1996. *Cost-Benefit Analysis: Theory and Application*. Thousand Oaks, CA: Sage Publishing.
- Refsgaard, J.C., Sluijs, J.P., Højberg, A.L. and Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling & Software*, 22(11), 1543-1556.
- Rubinstein R.Y. and Kroese, D.P., 2008. *Simulation and the Monte Carlo Method (2nd Edition)*. New York: John Wiley & Sons.
- Sherman, G., Siebers, P.O., Aickelin, U. and Menachof, D., 2010. Scenario Analysis, Decision Trees and Simulation for Cost Benefit Analysis of the Cargo Screening Process. *Proceedings of the International Workshop of Applied Modelling and Simulation (WAMS)*. May 5-7, Buizos, Brasil.
- Turner, K. and Williams, G., 2005. Modelling complexity in the automotive industry supply chain. *Journal of Manufacturing Technology Management*, 16(4), 447-458.
- UK Border Agency, 2008. *Freight Search Figures provided by the UK Border Agency* [unpublished]
- Wilson, D.L., 2005. Use of modeling and simulation to support airport security. *IEEE Aerospace and Electronic Systems Magazine*, 208, 3-8.
- XJTEK, 2005. *AnyLogic User's Guide*. St. Petersburg, Russian Federation: XJ Technologies.

# IMPROVEMENT OF THE FORECAST OF ECONOMIC PROCESSES PARAMETERS

Yuri L. Menshikov

*Dnepropetrovsk University, Gagarina av.72, 49010,  
Dnepropetrovsk, Ukraine*

*e-mail: Menshikov2003@list.ru*

## ABSTRACT

In work the problem of synthesis of mathematical model of economic process is examined in deterministic statement. It is supposed that the amount of measurements by each variable minimally and coincides with number of variable in model. Such problem can also be named as fast identification of parameters of mathematical model of economic processes. The some possible variants of statement of such problem are considered. The retrospective test calculation on real measurements were executed for comparison with known methods.

Keywords: economic processes, identification parameters, different statements, regularization method

## 1. INTRODUCTION

The classical problem of parameters identification of linear stationary many-dimensional model consists in synthesis of linear connection between the chosen characteristics of process  $q_1, q_2, \dots, q_n$  (Grop 1979, Seidg and Melsa 1974). For simplicity we shall consider only problem of construction of linear model:

$$q_1 = z_1 q_2 + z_2 q_3 + \dots + z_{n-1} q_n + z_n, \quad (1)$$

where  $z_1, z_2, \dots, z_n$  – unknown coefficients of stationary mathematical model.

Let's denote  $z^T = (z_1, z_2, \dots, z_n)$ ,  $(\cdot)^T$  – the mark of transposition. It is supposed that for everyone variable  $q_i$  ( $i=1, 2, \dots, n$ ) we have  $m$  of measurements  $q_{ik}$  ( $k=1, 2, \dots, m$ ),  $n < m$ . Let's designate  $q_i^T = (q_{i1}, q_{i2}, \dots, q_{im})$ .

This problem can be reduced to the solution of the redefined linear non-uniform system of the algebraic equations which are being executed, as a rule, by method of least squares (Grop 1979, Seidg and Melsa 1974). It is supposed that statistical characteristics for all variables are given. However, the similar information can be received only on the basis of numerous experiments during long time (infinite time). For a big interval of time the hypothesis about a constancy of properties of the considered process (for

example, constancy of coefficients of linear model) can not be fair. So it makes the process of parameters identification of linear mathematical model groundless.

Last years some new approaches to the solution of a problem of parameters identification of mathematical model have been offered in which the statistical characteristics of variables are not used (Kunchevich, Lichak and Nikitenko 1988, Polajk and Nazin 2006, Gubarev and Tiginov 2006). Concerning an errors of measurements the hypothesis is accepted that they are limited on absolute size to some given constant. Such initial preconditions in the greater degree correspond to a real situation. In work (Kunchevich, Lichak and Nikitenko 1988), for example, the problem of solution of system of the linear algebraic equations is considered at which the right part and matrix of system are approximately given. Naturally the result of the solution of a problem in such statement will be the some set of the possible solutions. It is shown that if all matrixes of system are non-degenerate the set is limited and correspond to the convex polyhedrons.

In article (Polajk and Nazin 2006) the problem of obtaining of the guaranteed multiplicity characteristics for researched variable process is formulated. If in family of the approached matrixes of linear system have not singular matrixes the effective algorithm of the solution of the put problem is offered.

In the given work the problem of synthesis of parameters of multivariate regress is considered in which the information about statistical properties of measurements are not used also (Kunchevich, Lichak and Nikitenko 1988, Polajk and Nazin 2006, Gubarev and Tiginov 2006). However in this case it is supposed that the singular matrixes and close to them belong to possible matrixes of linear system with guarantee (Menshikov 2004). The number of measurements is minimal and equals to number of variables of researched process. It allows attribute such algorithm to fast algorithms of identification of parameters.

We shall present the problem of synthesis of linear mathematical model with  $n$  variables relatively  $q_1$  for number of measurements  $m = n$ , as a problem of the solution of system (Menshikov 2004)

$$A_p(q_2, q_3, \dots, q_n) z = q_1, \quad (2)$$

where the operator  $A_p(q_2, q_3, \dots, q_n)z$  is determined as follows

$$A_p(q_2, q_3, \dots, q_n)z = z_1q_2 + z_2q_3 + \dots + z_{n-1}q_n + z_ne,$$

$e$  is the unit vector of dimension  $n$ .

As the measurements of variables are received experimentally it is assumed that each measurement  $q_{ij}$ ,  $1 \leq i, j \leq n$  has some error the maximal size of which is known:

$$|q_{ij} - q_{ij}^{ex}| \leq \delta_i, \quad 1 \leq j \leq n, i = 1, 2, \dots, n, \quad (3)$$

where  $q_{ij}^{ex}$  is exact measurements of variable  $q_i$ .

The similar information of measurement errors, as a rule, is known a priori. The statistical characteristics of errors of measurements are unknown.

Let us denote vector  $\overline{p}$  as vector from space  $R^n \oplus R^n \oplus R^n \oplus \dots \oplus R^n = R^{n(n-1)}$ :

$$p^T = (q_{21}, \dots, q_{2n}, q_{31}, \dots, q_{3n}, \dots, q_{n1}, \dots, q_{nn}),$$

where  $R^n$  is Euclidean vector space.

Each vector  $q_i$  can accept meanings in some closed area  $D_i \subset R^n$  by virtue of inequalities (3). Vectors  $p$  can accept meanings in some closed area  $D = D_2 \oplus D_3 \oplus D_4 \oplus \dots \oplus D_n \subset R^{n(n-1)}$ . The

certain operator  $A_p$  associates with each vector  $p$  from area  $D$ . The class of operators  $\{A_p\} = K_A$  will correspond to the set  $D \subset R^{n(n-1)}$ .

Shall we rewrite (2) as

$$A_p z = u_{\delta_1}, \quad (4)$$

where

$$u_{\delta_1} = q_1; u_{\delta_1} \in U = R^n; z \in Z = R^n; \|u_{\delta_1} - u_1^{ex}\| \leq \delta_1, u_1^{ex}$$

– exact right part of (4);  $\|A^{ex} - A_p\|_{Z \rightarrow U} \leq h$ ,  $A^{ex}$  –

exact operator in (4);  $\|\cdot\|$  is the norm of a vector in Euclidean space  $R^n$ .

Let us consider now the set of the solutions of the equation (4) with the fixed operator  $A_p \in K_A$ :

$$Q_{\delta_1, p} = \{z: \|A_p z - u_{\delta_1}\| \leq \delta_1\}.$$

The set  $Q_{\delta_1, p}$  is limited if  $\Delta = \det A_p \neq 0$  and unlimited if  $\Delta = \det A_p = 0$ .

## 2. PROBLEM STATEMENTS

Any vector  $z$  from set  $Q_{\delta_1, p}$  is the good mathematical model of process so this vector after action of the operator  $A_p$  vector  $A_p z$  coincides with the given vector  $q_1$  with accuracy of measurement  $\delta_1$ . For choice of particular model from set  $Q_{\delta_1, p}$  it is necessary to use additional conditions. If such conditions are absent then it is possible to accept as the solution (4) the element  $z_p = Q_{\delta_1, p}$  for which the equality is carried out (Menshikov and Nakonechnij 2003)

$$\|z_p\|^2 = \inf_{z \in Q_{\delta_1, p}} \|z\|^2. \quad (5)$$

The vector  $z_p = Q_{\delta_1, p}$  is possible to interpret as a maximum steady element to the change of the factors not taken into account (most stable part), as the influence of these factors will increase norm of a vector  $z_p$  (Menshikov and Nakonechnij 2003). Such a property of the solution  $z_p$  is especially important if one takes into account that the vector  $z_p$  further will be used for

forecasting real processes (parameter  $q_1$ ).

Consider now the set  $Q^* = \bigcup_{p \in D} Q_{\delta_1, p}$  (Menshikov 2004).

Let us consider an extreme problem

$$\|z^*\|^2 = \inf_{p \in D} \inf_{z \in Q_{\delta_1, p}} \|z\|^2. \quad (6)$$

The vector  $z^* = Q^*$  is an estimation from below of possible solutions of the equation (4). The similar problem in classical identification is not examined. The statement of the following extreme problem is possible also:

$$\|z_{sup}^*\|^2 = \sup_{p \in D} \inf_{z \in Q_{\delta_1, p}} \|z\|^2. \quad (7)$$

The vector  $z_{sup}^* \in Q^*$  has the greatest norm among the solutions of a problem of synthesis on sets  $Q_{\delta_1, p}$ . The similar problem in the literature is not considered either. Models  $z^*, z_{sup}^*$  can be used for short-term forecasting

of change of variable  $q_1$  as on the one hand models

$z^*, z_{sup}^*$  are received by an rapid way and on the other hand these models are steadiest to the change of the factors not taken into account.

Except (6), (7) it is possible to examine the following statements of problems:

$$\|z_{0,0,\dots,1}\|^2 = \inf_{q_2 \in D_2} \inf_{q_3 \in D_3} \dots \inf_{q_{n-1} \in D_{n-1}} \sup_{q_n \in D_n} \inf_{z \in Q_{\delta_1, p}} \|z\|^2 \quad (8)$$

$$\|z_{0,0,\dots,1,1}\|^2 = \inf_{q_2 \in D_2} \inf_{q_3 \in D_3} \dots \sup_{q_{n-1} \in D_{n-1}} \sup_{q_n \in D_n} \inf_{z \in Q_{\delta_1, p}} \|z\|^2 \quad (9)$$

.....

$$\|z_{0,1,\dots,1,1}\|^2 = \inf_{q_2 \in D_2} \sup_{q_3 \in D_3} \dots \sup_{q_{n-1} \in D_{n-1}} \sup_{q_n \in D_n} \inf_{z \in Q_{\delta_1, p}} \|z\|^2 \quad (10)$$

In some cases it is expedient to consider the following problems of identification of parameters:

$$\|z^{0,0,\dots,0}\|^2 = \inf_{z \in Q_{\delta_1, p^{0,0,\dots,0}}} \|z\|^2, \quad (11)$$

$$\|z^{0,0,\dots,1}\|^2 = \inf_{z \in Q_{\delta_1, p^{0,0,\dots,1}}} \|z\|^2, \quad (12)$$

.....

$$\|z^{1,1,\dots,1}\|^2 = \inf_{z \in Q_{\delta_1, p^{1,1,\dots,1}}} \|z\|^2, \quad (13)$$

where vector  $p^{0,0,\dots,0}$  has the minimal possible size of all components of vector  $p$ ,  $p^{0,0,\dots,1}$  has the minimal possible size of components  $q_1, q_2, \dots, q_{n-1}$  and has the maximal size of  $q_n$ ; . . . ; vector  $p^{1,1,\dots,1}$  has the maximal possible size of all components of vector  $p$ .

It is possible to consider the following extreme problem

$$\|A_{p^{opt}} z_{\delta_1}^{pl} - u_{\delta_1}\|^2 = \inf_{z_a \in Q^*} \sup_{A_p \in K_A} \|A_p z_a - u_{\delta_1}\|^2, \quad (14)$$

where  $z_a$  is the solution of extreme problem

$$\|z_a\|^2 = \inf_{z \in Q_{\delta_1, a}} \|z\|^2. \quad (15)$$

Let's called solution  $z_{\delta_1}^{pl}$  as **more plausible mathematical model**.

Use of such model with the purpose of the forecast allows to receive the characteristic  $q_1$  with the least maximal deviation from experiment with possible

variations of variables  $q_2, q_3, \dots, q_n$  within the given errors.

**Theorem.** Solution  $z_{\delta_1}^{pl}$  of extreme problem (14) exists, uniquely and is steady to small changes of the initial data if the vector  $z_{\delta_1}^{pl}$  is defined uniquely from condition (14) (Menshikov 2004).

### 3. METHODS OF SOLUTION

One of possible ways of the solution of extreme problems (6) - (10) is use in accounts of special mathematical models of researched processes (Menshikov 2005, Menshikov 2006). If the special mathematical models exist then the solution of extreme problems (6) - (10) can be replaced with more simple extreme problems with the precisely given operator such as problems (11) - (13). Further solution is carried out by regularization method of Tikhonov with a choice of regularization parameter by discrepancy method (Tikhonov and Arsenin 1979).

For approximate solution of extreme problem (14) the interval of change of each components of vector  $p$  is divided by a uniform grid  $p_m$ . The number of the operators in set will be final  $K_A = \{A_1, A_2, \dots, A_N\} = \{A_i\}$ . For each operator  $A_i$  the solution  $z_i$  is defined. Further the approached solution of an extreme problem (14) is being defined by simple sorting.

$$\|A_{p^{opt}} z_{\delta_1}^{pl} - u_{\delta_1}\|^2 = \inf_{z_a \in Q^*} \sup_{A_p \in K_A} \|A_p z_a - u_{\delta_1}\|^2 =$$

$$= \min_j \max_i \{ \|A_i z_j - u_{\delta_1}\|^2 \}, 1 \leq i \leq m, 1 \leq j \leq m.$$

where  $z_j$  satisfies the condition

$$\|A_j z_j - u_{\delta_1}\|^2 = \delta^2.$$

### 4. TEST CALCULATION

The problem of construction econometric model was chosen for test calculation on the data which are given in work (Malenkov 1975, tabl.1). In the beginning for first four years was constructed econometric model by a method of the least squares in the assumption that the errors of measurements satisfies to the normal law of distribution. The minimization was carried out on first variable  $q_1$  (on a vertical). As a result of accounts the following mathematical model was received:

$$q_1 = 0.018q_2 - 2.35q_3 + 0.0076q_4 + 22.27. \quad (16)$$



**Table 1** Import, manufacture, change of stocks and consumption in France (millard. franks, in the prices 1959).

Number (Years)	Import $q_{1i}$	Manu- facture $q_{2i}$	change of stocks $q_{3i}$	Consum ption $q_{4i}$
1 (1949)	15.9	149.3	4.2	108.1
2 (1950)	16.4	161.2	4.1	114.8
3 (1951)	19.0	171.5	3.1	123.2
4 (1952)	19.1	175.5	3.1	126.9
5 (1953)	18.8	180.8	1.1	132.1
6 (1954)	20.4	190.7	2.2	137.7
7 (1955)	22.7	202.1	2.1	146.0
8 (1956)	26.5	212.4	5.6	154.1
9 (1957)	28.1	226.1	5.0	162.3
10 (1958)	27.6	231.9	5.1	164.3

Further in the assumption, that the errors of measurements do not surpass 5 %, the mathematical models  $z_{17}, z_{18}, z_{19}, z_{20}$  as result of the solution of the following extreme problems were constructed:

$$\|z_{17}\|^2 = \inf_{z \in Q_{\partial_1, p}^{0,0,0}} \|z\|^2, \quad (17)$$

$$\|z_{18}\|^2 = \inf_{z \in Q_{\partial_1, p}^{0,0,1}} \|z\|^2, \quad (18)$$

$$\|z_{19}\|^2 = \inf_{z \in Q_{\partial_1, p}^{0,1,1}} \|z\|^2, \quad (19)$$

$$\|z_{20}\|^2 = \inf_{z \in Q_{\partial_1, p}^{1,1,1}} \|z\|^2, \quad (20)$$

where

$p^{0,0,0} = (q_2^0, q_3^0, q_4^0)^T$ ,  $p^{0,0,1} = (q_2^0, q_3^0, q_4^1)^T$ ,  
 $p^{0,1,1} = (q_2^0, q_3^1, q_4^1)^T$ ,  $p^{1,1,1} = (q_2^1, q_3^1, q_4^1)^T$ ,  
 $q_k^0$  is minimal possible size of  $q_k$ ,  $q_k^1$  is maximal possible size of  $q_k$ . The following econometric models correspond to solutions  $z_{17}, z_{18}, z_{19}, z_{20}$  :

$$q_1 = -0.099q_2 - 1.137q_3 + 0.24q_4 + 7.03, \quad (17a)$$

$$q_1 = -0.09q_2 - 1.14q_3 + 0.27q_4 + 7.03, \quad (18a)$$

$$q_1 = -0.1q_2 - 1.14q_3 + 0.27q_4 + 7.03, \quad (19a)$$

$$q_1 = -0.09q_2 - 1.03q_3 + 0.24q_4 + 7.03. \quad (20a)$$

The forecast of size  $q_1$  on these models for 1953, 1954 is given in the table 2. For comparison with real measurements (table 1, the second column) is possible to make a conclusion that best short-term forecast gives model (17a) among models (16), (17a) - (20a).

**Table 2** Forecast of economic parameters for 1953; 1954.

N	Tab.1	mod. (16)	mod. (17a)	mod (18a)	Mod. (19a)
5	18.8	23,9	19,7	24,84	23,12
6	20.4	21,6	18,8	24,17	22,38

The results of calculations on real measurements by regularization method (Tikhonov and Arsenin 1979) have shown that mathematical models  $z_{17}, z_{18}, z_{19}, z_{20}$  describe the real situation more exactly than classical econometrical models  $z_{16}$ .

It is necessary to note that the quality of the forecast for the long period is worse, than for the short period. It is expected effect as the offered algorithms were designed for short-term forecast.

Choice in practical problems the certain mathematical model is being determined of the specificity of a concrete problem and final goal of use of mathematical model. However the best model for the forecast can not be determined a priori.

## CONCLUSION

The offered approach to a problem of identification of parameters of static linear mathematical model allows to expand a class of the possible solutions (mathematical models) up to maximal possible. Some variants of statement of such problems of parameters identification are considered. There are no basic restrictions for distribution of such approach to nonlinear problems of identification, in author opinion.

## REFERENCES

- Grop, D. 1979. Methods of identification of systems. *World, Moscow*.  
 Seidg, A.P., Melsa, Dj. L. 1974. Identification of control systems. *Science, Moscow*.

- Kunchevich, V.M., Lichak, M.M., Nikitenko, F.S. 1988. The solution of system of the linear equations at presence of uncertainty in both parts. *Cybernetics*, Kiev, 4, 42-49.
- Polajk, B.T., Nazin, S.A. 2006. Evaluation of parameters in linear many-dimensional systems with interval uncertainty, *Problems of Control and Informatics*, Kiev, 4, 103-115.
- Gubarev, V.F., Tigunov, P.A. 2006. About features identification of many-dimensional of continuous systems on the data with the limited uncertainty, *Problems of Control and Informatics*, Kiev, 231-246.
- Menshikov, Yu.L. 2004. About one variant of an econometrics problem. *Proceedings of Inter. Conference Problems of Solution making under Uncertainties (PDMU-2004)*, May 25-30, Abstracts, pp.164-165, Ternopil, (Ukraine).
- Menshikov Yu.L., Nakonecniij A.G. (2003) Principle of the maximal stability in inverse problems with a minimum of the a priori information. *Proceedings of Inter. Conf. Problems of Decision making under Uncertainties (PDMU-2003)*, pp.80-82. September 8-12, Abstracts, Kiev-Alushta, (Ukraine).
- Menshikov, Yu.L. 2005. Econometric models for special purposes. *Proceeding of Int. Conf. Problems of Decision Making under Uncertainties (PDMU-2005)*, Abst., pp.192-193, Berdyansk, (Ukraine).
- Menshikov, Yu.L. 2006. The Fast Identification of Parameters. *Proceedings of 11-th Int. Conf. Mathematical Methods in Electromagnetic Theory, MMET'06*, pp.267-269, Kharkiv, (Ukraine).
- Tikhonov, A.N., Arsenin, V.Ya. 1979. *Methods of solutions of incorrect problems*, Moscow, Science.
- Malenkov, E. 1975. *Statistics methods in econometrics*. Moscow, Statistics.

#### **AUTHORS BIOGRAPHY**

**Yuri Menshikov.** He received the degrees of Ph. D. in Applied Mathematics at Kharkov Polytechnic University, Ukraine on 1979. He is working in Dnepropetrovsk University (Mechanics and Mathematics Faculty) from 1970. Dr. Yuri Menshikov is reviewer of conference Mathematical Modelling and he is member of Germany Society for Industrial and Applied Mathematics (GAMM). Research interests of Yuri Menshikov include system control, differential equations, variation methods, inverse problems. He is the author and coauthor of two monographs and than 270 scientific papers in international journals and conference proceedings.

# SOLVING FULLY FUZZY LINEAR PROGRAMMING PROBLEM WITH EQUALITY AND INEQUALITY CONSTRAINTS

Ahmad Jafarnejad<sup>1(a)</sup>, Mahnaz Hosseinzadeh<sup>2(b)</sup>, Hamed Mohammadi Kangarani<sup>3(c)</sup>

<sup>(a)</sup> Full Professor, Faculty of Management, University of Tehran, Tehran, Iran

<sup>(b)</sup> PhD Candidate in Operation Research, Faculty of Management, University of Tehran, Tehran, Iran

<sup>(c)</sup> Psychiatrist (MD), specialist in fuzzy logic in personality traits, University of social welfare and rehabilitation sciences, Tehran, Iran

<sup>(a)</sup>[jafarnjd@ut.ac.ir](mailto:jafarnjd@ut.ac.ir), <sup>(b)</sup>[mhosseinzadeh@ut.ac.ir](mailto:mhosseinzadeh@ut.ac.ir), <sup>(c)</sup>[hamed.mohammadikangarani@uswr.ac.ir](mailto:hamed.mohammadikangarani@uswr.ac.ir)

## ABSTRACT

Fuzzy linear programming problems are models in which all parameters as well as the variables are represented by fuzzy numbers and are known as FFLP problems. In this paper a new method is proposed to find the fuzzy optimal solution of FFLP problems including both equality and inequality constraints. Numerical examples are solved to show capability of the proposed method in real life situations that the inherent fuzziness of a decision problem demands a fuzzy decision to be taken. The main advantage of the proposed method is simplicity of computations.

Keywords: Fully Fuzzy Linear Programming Problems, Triangular Fuzzy numbers, fuzzy ranking.

## 1. INTRODUCTION

The first application of FST to decision-making processes was presented by Bellman and Zadeh (1965) and the concept of fuzzy mathematical programming on general level was first proposed by Tanaka et al. (1973). The first formulation of fuzzy linear programming (FLP) is proposed by Zimmermann (1978). Various fuzzy linear programming techniques are surveyed in literature which are classified into two main classes: fuzzy linear programming (Verdegey 1982, Warners 1987a, Warners 1987b, Zimmerman 1976, Chanas 1983, Verdegey 1984a, Verdegey 1984b) and possibilistic linear programming (Carlsson and Corhonen 1986, Ramik and Rimanek 1985, Tanaka Ichihashi and Asai 1984, Lai and Hwang 1992, Rommelfanger Hanuscheck and Wolf 1989, Buckley 1988). For the sake of simplicity, usually these techniques consider only crisp solutions of the fuzzy problems (Stanculescu Fortemps Install and Wertz 2003).

Generally speaking, in fuzzy linear programming problems, the coefficients of decision variables are fuzzy numbers while decision variables are crisp ones. This

means that in an uncertain environment, a crisp decision is made to meet some decision criteria (Tanaka Guo and Zimmermann 2000), thus the decision-making process is constrained to crisp decisions that hide the fuzzy aspect of the problem. Buckley and Feuring (2000), have developed a kind of fuzzy linear programming problems in which all the parameters as well as the variables are represented by fuzzy numbers which is known as FFLP problems. Then some other authors have proposed different methods to solve FFLP problems under two categories: FFLP problems with inequality constraints (Maleki Tata and Mashinchi 2000, Buckley and Feuring 2000, Hashemi Modarres and Nasrabadi 2006, Allahviranloo Lotfi Kiasary Kiani and Alizadeh 2008) and FFLP problems with equality constraints (Dehghan Hashemi and Ghatee 2006, Lotfi Allahviranloo Jondabehe and Alizadeh 2009, Kumar Kaur and Singh 2011). As pointed out by Buckley and Feuring (2000), searching for the optimal solutions of FFLP problems is very difficult. They used evolutionary algorithm to find the solution.

The paper is organized as follows: in section 2, some basic definitions of fuzzy set theory and arithmetic between two triangular fuzzy numbers are reviewed; In Section 3 formulation of FFLP problems are discussed; In Section 4 a new method is proposed for solving FFLP presented in section3; To illustrate the proposed method, numerical examples are solved in Section 5; Conclusion is drawn in Section 6

## 2. PRELIMINARIES

### Definition 2.1.

A fuzzy number  $\tilde{A} = (m_A, w_A, \acute{w}_A)$  is a LR type if and only if:

$$A(x) = \begin{cases} L\left(\frac{m_A - x}{w_A}\right) & -\infty < x \leq m_A \\ R\left(\frac{x - m_A}{w_A}\right) & m_A \leq x < +\infty \end{cases} \quad w_A, w_A \geq 0$$

Where  $m_A$  is the center (core) and  $w_A$  and  $w_A$  are the left and right bandwidths of  $\tilde{A}$  respectively. This is a parametric form of fuzzy number  $\tilde{A}$ , so we can show it as a triangular shape as follow:

$$\tilde{A} \equiv (m, w_A, w_A)_{LR}$$

**Definition 2.2.**

A LR type Fuzzy number is said to be a triangular fuzzy number if and only if its membership function is as follow:

$$A(x) = \begin{cases} 1 - \frac{m_A - x}{w_A} & m_A - w_A < x \leq m_A \\ 1 - \frac{x - m_A}{w_A} & m_A \leq x < m_A + w_A \\ 0 & \text{otherwise} \end{cases}$$

The graphical display of a triangular fuzzy number  $A$  is shown in figure (1).

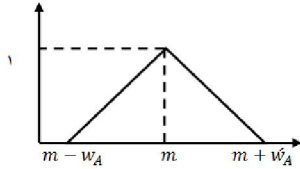


Figure 1: Graphical display of a triangular fuzzy number

**Definition 2.3.**

Given two triangular fuzzy numbers  $\tilde{A} = (a, w_A, w_A)_{LR}$  and  $\tilde{B} = (b, w_B, w_B)_{LR}$ ,  $\tilde{B} < \tilde{A}$  if and only if:

$$b < a \quad \text{and} \quad w_B + w_B \geq w_A + w_A$$

And  $\tilde{B} = \tilde{A}$  if and only if

$$b = a \quad \text{and} \quad w_B + w_B = w_A + w_A$$

**Definition 2.4.**

A triangular fuzzy number  $\tilde{A} = (a, w_A, w_A)_{LR}$  is said to be non-negative fuzzy number if and only if  $a - w_A \geq 0$

**Definition 2.5.**

Given two fuzzy numbers  $\tilde{A} = (a, w_A, w_A)_{LR}$  and  $\tilde{B} = (b, w_B, w_B)_{LR}$  with continuous nondecreasing function over  $[0, \infty)$ , fuzzy Arithmetic operations are defined as follows:

$$\tilde{A} + \tilde{B} = (a + b, w_A + w_B, w_A + w_B)_{LR}$$

$$\tilde{A} - \tilde{B} = (a - b, w_A + w_B, w_A + w_B)_{LR}$$

$$(a, w_A, w_A)_{LR} \cdot (b, w_B, w_B)_{LR} \approx (a \cdot b, aw_B + bw_A, aw_B + bw_A)_{LR} \quad \text{for } \tilde{A} > 0 \text{ and } \tilde{B} > 0$$

$$(a, w_A, w_A)_{LR} \cdot (b, w_B, w_B)_{LR} \approx (a \cdot b, -aw_B + bw_A, -aw_B + bw_A)_{LR} \quad \text{for } \tilde{A} < 0 \text{ and } \tilde{B} > 0$$

**3. FULL FUZZY NUMBER LINEAR PROGRAMMING PROBLEMS (FFLP)**

A Full fuzzy number linear programming problem type (FFLP) is defined as:

$$\text{Max (or Min)} \tilde{Z} = (\tilde{C}^T \otimes \tilde{x})$$

$$\text{Subject to: } \tilde{A} \otimes \tilde{x} \begin{matrix} \leq \\ (=) \\ \geq \end{matrix} \tilde{b} \quad (3.1)$$

$$\tilde{x} \geq 0$$

Where  $\tilde{x} = (x_m, w, w)$ ,  $\tilde{C} = (c, w_c, w_c)$ ,  $\tilde{A} = (A, w_A, w_A)$ ,  $b = (b, w_b, w_b)$ ,  $x_m = \text{core}(\tilde{x})$ ,  $c = \text{core}(\tilde{C})$ ,  $A = \text{core}(\tilde{A})$ ,  $b = \text{core}(\tilde{b})$  and  $w, w, w_c, w_c, w_A, w_A, w_b, w_b$  are the left and right bandwidths of  $\tilde{x}$ ,  $\tilde{C}$ ,  $\tilde{A}$  and  $\tilde{b}$  respectively.  $\tilde{x} \in F^n$ ,  $\tilde{b} \in F^m$ ,  $\tilde{A} = [a_{ij}]^{M \times n} \in F^{M \times n}$ ,  $\tilde{C} \in F^n$ .

Therefore the above FFLP problem can be written as:

$$\text{Max (or Min)} \tilde{Z} = \sum_{j=1}^n (c_j, w_c, w_c) \otimes (x_j, w_j, w_j)$$

Subject to:

$$\sum_{j=1}^n (a_{ij}, w_{a_{ij}}, w_{a_{ij}}) \otimes (x_j, w_j, w_j) \begin{pmatrix} \leq \\ (=) \\ \geq \end{pmatrix} (b_i, w_{b_i}, w_{b_i})$$

$$\forall i = 1, 2, \dots, M, j = 1, \dots, n$$

$$(x_j, w_j, w_j) \text{ non-negative } j = 1, \dots, n \quad (3.2)$$

**Definition 3.1.**

We say that fuzzy vector  $\tilde{x} = (x_{jm}, w_j, w_j)$  is a fuzzy feasible solution to the problem (3.2) if  $\tilde{x}$  satisfies the constraints of the problem.

**Definition 3.2.**

A fuzzy feasible solution  $\tilde{x}^*$  is a fuzzy optimal solution for (3.2), if for all fuzzy feasible solution  $\tilde{x}$  for (3.2), we have  $\tilde{c} \otimes \tilde{x}^* \begin{pmatrix} \leq \\ (=) \\ \geq \end{pmatrix} \tilde{c} \otimes \tilde{x}$

#### 4. PROPOSED METHOD TO FIND THE FUZZY OPTIMAL SOLUTION OF THE FFLP PROBLEM.

In this section, a new method is proposed to find the fuzzy optimal solution of the FFLP problems:

First, using the arithmetic operations presented in definition (2.5), the problem (3.2) is written in the form of problem (4.1):

$$\begin{aligned} \text{Max (Min)} Z &= [\sum_{j=1}^n c_j x_j, \sum_{j \in C_p} (c_j w_j + x_j w_{c_j}) + \\ &\sum_{j \in C_{\bar{p}}} (-c_j \dot{w}_j + x_j w_{c_j}), \sum_{j \in C_p} (c_j \dot{w}_j + x_j \dot{w}_{c_j}) + \\ &\sum_{j \in C_{\bar{p}}} (-c_j w_j + x_j \dot{w}_{c_j})] \\ \text{s. t. } &[\sum_{i=1}^M \sum_{j=1}^n a_{ij} x_j, \sum_{a_{ij} \in A_p} (a_{ij} w_j + x_j w_{a_{ij}}) + \\ &\sum_{a_{ij} \in A_{\bar{p}}} (-a_{ij} \dot{w}_j + x_j w_{a_{ij}}), \sum_{a_{ij} \in A_p} (a_{ij} \dot{w}_j + x_j \dot{w}_{a_{ij}}) + \\ &\sum_{a_{ij} \in A_{\bar{p}}} (-a_{ij} w_j + x_j \dot{w}_{a_{ij}})] \begin{pmatrix} \leq \\ (=) \\ \geq \end{pmatrix} (b_i, w_{b_i}, \dot{w}_{b_i}) \\ \forall i &= 1, 2, \dots, M, j = 1, \dots, n \end{aligned} \quad (4.1)$$

$(x_j, w_j, \dot{w}_j)$  non – negative  $j = 1, \dots, n$

Where:

$$C_p = \{j | \tilde{c}_j \text{ is non negative}\}$$

$$C_{\bar{p}} = \{j | \tilde{c}_j \text{ is negative}\}$$

$$A_p = \{j | \tilde{a}_{ij} \text{ is non negative}\}$$

$$A_{\bar{p}} = \{j | \tilde{a}_{ij} \text{ is negative}\}$$

The objective function denotes a triangular fuzzy number which its center, left side and right side bandwidth are as follows:

$$Z_m = \sum_{j=1}^n c_j x_j$$

$$w_Z = \sum_{j \in C_p} (c_j w_j + x_j w_{c_j}) + \sum_{j \in C_{\bar{p}}} (-c_j \dot{w}_j + x_j w_{c_j})$$

$$\dot{w}_Z = \sum_{j \in C_p} (c_j \dot{w}_j + x_j \dot{w}_{c_j}) + \sum_{j \in C_{\bar{p}}} (-c_j w_j + x_j \dot{w}_{c_j})$$

Furthermore the left hand sides of the constraints are triangular fuzzy numbers as well. The components are as follows:

$$(AX)_m = \sum_{i=1}^M \sum_{j=1}^n a_{ij} x_j$$

$$\begin{aligned} (AX)_w &= \\ &\sum_{a_{ij} \in A_p} (a_{ij} w_j + x_j w_{a_{ij}}) + \sum_{a_{ij} \in A_{\bar{p}}} (-a_{ij} \dot{w}_j + x_j w_{a_{ij}}) \end{aligned}$$

$$\begin{aligned} (AX)_{\dot{w}} &= \\ &\sum_{a_{ij} \in A_p} (a_{ij} \dot{w}_j + x_j \dot{w}_{a_{ij}}) + \sum_{a_{ij} \in A_{\bar{p}}} (-a_{ij} w_j + x_j \dot{w}_{a_{ij}}) \end{aligned}$$

As mentioned in the definition (2.3), a triangular fuzzy number will be larger than another triangular fuzzy number, if its center is larger and its bandwidths are smaller than the other one. So, to maximize (minimize) the value of the objective function that is a triangular fuzzy number, we can maximize (minimize) the mean and minimize (maximize) the bandwidths. In addition, constraints show a comparison of two triangular fuzzy number and we can use the definition (2.3) for such a comparison.

So in the second stage the problem (4.1) is converted to the following crisp LP problem using the fuzzy ranking method defined in definition (2.3). :

$$\text{Max (Min)} Z_m = \sum_{j=1}^n c_j x_j$$

$$\begin{aligned} \text{Min (Max)} (w_Z + \dot{w}_Z) &= \sum_{j \in C_p} (c_j w_j + x_j w_{c_j}) + \\ &(c_j \dot{w}_j + x_j \dot{w}_{c_j}) + \sum_{j \in C_{\bar{p}}} (-c_j \dot{w}_j + x_j w_{c_j}) + (-c_j w_j + \\ &x_j \dot{w}_{c_j}) \end{aligned}$$

$$\text{s. t. } \sum_{i=1}^M \sum_{j=1}^n a_{ij} x_j \begin{pmatrix} \leq \\ (=) \\ \geq \end{pmatrix} b_i$$

$$\begin{aligned} &\sum_{a_{ij} \in A_p} (a_{ij} w_j + x_j w_{a_{ij}}) + (a_{ij} \dot{w}_j + x_j \dot{w}_{a_{ij}}) + \\ &\sum_{a_{ij} \in A_{\bar{p}}} (-a_{ij} \dot{w}_j + x_j w_{a_{ij}}) + \\ &(-a_{ij} w_j + x_j \dot{w}_{a_{ij}}) \begin{pmatrix} \geq \\ (=) \\ \leq \end{pmatrix} w_{b_i} + \dot{w}_{b_i} \end{aligned}$$

$$\forall i = 1, 2, \dots, M, j = 1, \dots, n \quad (4-2)$$

Where:

$$C_p = \{j | \tilde{c}_j \text{ is non negative}\}$$

$$C_{\bar{p}} = \{j | \tilde{c}_j \text{ is negative}\}$$

$$A_p = \{j | \tilde{a}_{ij} \text{ is non negative}\}$$

$$A_{\bar{p}} = \{j | \tilde{a}_{ij} \text{ is negative}\}$$

In the next stage we find the optimal solution  $x_j, w_j$  and  $\dot{w}_j$  by solving the problem (4.2). Then we put the values  $x_j, w_j, \dot{w}_j$  in  $\tilde{x} = (x_{jm}, w_j, \dot{w}_j)$  and find the fuzzy optimal solution. Finally, we calculate the fuzzy optimal value of the objective function by putting  $\tilde{x}$  in  $\sum_{j=1}^n (c_j, w_{c_j}, \dot{w}_{c_j}) \otimes (x_j, w_j, \dot{w}_j)$ .

## 5. EXAMPLES

In this Section two FFLP examples including equality and inequality constraints as well as some negative fuzzy coefficients are solved to show the capability of the proposed method in different situations.

### Example 5-1:

To illustrate the efficiency of the method we consider an example used by Kumar Kaur and Singh (2011). Choosing this example is intentionally to compare the efficiency of our method with the efficiency of their method. This example includes equality constraints with  $\tilde{a}_{ij} \geq 0, i = 1, \dots, M$  and  $j = 1, \dots, n$ .

In the form we have presented our equations and models in our paper, the Kumar example can be rewritten as below:

$$\text{Max } Z_1 = (2,1,1) \otimes (x_{1m}, w_1, \dot{w}_1) + (3,1,1) \otimes (x_{2m}, w_2, \dot{w}_2)$$

$$\text{s. t. } (1,1,1) \otimes (x_{1m}, w_1, \dot{w}_1) + (2,1,1) \otimes (x_{2m}, w_2, \dot{w}_2) = (10,8,14)$$

$$(2,1,1) \otimes (x_{1m}, w_1, \dot{w}_1) + (1,1,1) \otimes (x_{2m}, w_2, \dot{w}_2) = (8,7,13)$$

According to the proposed method the above LP is converted to:

$$\text{Max } 2x_{1m} + 3x_{2m}$$

$$\text{Min } 2x_{1m} + 2x_{2m} + 2w_1 + 2\dot{w}_1 + 3w_2 + 3\dot{w}_2$$

Subject to:

$$x_{1m} + 2x_{2m} = 10$$

$$x_{1m} + x_{2m} + 1w_1 + 2w_2 = 8$$

$$x_{1m} + x_{2m} + 1\dot{w}_1 + 2\dot{w}_2 = 14$$

$$2x_{1m} + x_{2m} = 8$$

$$x_{1m} + x_{2m} + 2w_1 + w_2 = 7$$

$$x_{1m} + x_{2m} + 2\dot{w}_1 + \dot{w}_2 = 13$$

$$x_{1m} - w_1 \geq 0, \quad x_{2m} - w_2 \geq 0$$

The solution of the problem is:

$$(x_{1m}, w_1, \dot{w}_1) = (2, 0, 0)$$

$$(x_{2m}, w_2, \dot{w}_2) = (4, 1, 0)$$

$$\tilde{z}^* = (16, 9, 6)$$

The optimal solution of the Kumar approach is:

$$(x_{1m}, w_1, \dot{w}_1) = (2, 1, 1)$$

$$(x_{2m}, w_2, \dot{w}_2) = (4, 2, 2)$$

$$\tilde{z}^* = (16, 12, 17)$$

Based on the definition (2.3), our solution finds more precise value for decision variable and better (larger value) for objective function. In a fuzzy environment a more precise decision is more helpful to decision makers.

### Example 5-2:

Let us consider the following FFLP and solve it by the proposed method. This example includes inequality constraints in which:

$$\exists i \text{ and } j, \tilde{a}_{ij} \leq 0.$$

$$i = 1, \dots, M, \quad j = 1, \dots, n$$

The problem is:

$$\text{Max } Z_1 = (6,5,3) \otimes (x_{1m}, w_1, \dot{w}_1) + (3,1,5) \otimes (x_{2m}, w_2, \dot{w}_2)$$

$$\text{s. t. } (3,1,1) \otimes (x_{1m}, w_1, \dot{w}_1) + (2,1,1) \otimes (x_{2m}, w_2, \dot{w}_2) \lesseqgtr (16,10,14)$$

$$(1,2,1) \otimes (x_{1m}, w_1, \dot{w}_1) + (3,2,1) \otimes (x_{2m}, w_2, \dot{w}_2) \lesseqgtr (17,16,13)$$

The above LP is converted to:

$$\text{Max } 6x_{1m} + 3x_{2m}$$

$$\text{Min } 8x_{1m} + 6x_{2m} + 6w_1 + 6\dot{w}_1 + 3w_2 + 3\dot{w}_2$$

Subject to:

$$3x_{1m} + 2x_{2m} \leq 16$$

$$2x_{1m} + 2x_{2m} + 3w_1 + 2w_2 + 3\dot{w}_1 + 2\dot{w}_2 \geq 24$$

$$x_{1m} + 3x_{2m} \leq 17$$

$$3x_{1m} + 3x_{2m} - w_1 + 3w_2 - \dot{w}_1 + 3\dot{w}_2 \geq 29$$

$$x_{1m} - w_1 \geq 0, \quad x_{2m} - w_2 \geq 0$$

The solution of the problem is:

$$(x_{1m}, w_1, \dot{w}_1) = (5.33, 0, 1.27)$$

$$(x_{2m}, w_2, \dot{w}_2) = (0, 0, 4.76)$$

$$\tilde{z}^* = (31.98, 26.65, 37.89)$$

As it can be seen, the proposed approach can deal with Fuzzy linear programming problems which contain inequality constraints and some negative fuzzy coefficient as well.

## 6. CONCLUSION

In this paper a new method is proposed to find the fuzzy optimal solution of FFLP problems; using the proposed method the fuzzy solution of FFLP problems, occurring in real life situation, can be easily obtained. With the aid of this, decision makers are able to obtain fuzzy decisions to reflect the inherent fuzziness of a decision problem, so the need to do sensitivity analysis after obtaining a crisp solution decreases. To illustrate the proposed method FFLP examples including equality and inequality constraints as well as some negative fuzzy coefficients are solved and result shows the capability of the method in different mentioned situations. The main advantage of the proposed method is simplicity of computations.

## REFERENCES

- Allahviranloo, T., Lotfi, F.H., Kiasary, M.K., Kiani, N.A., Alizadeh, L., 2008, Solving full fuzzy linear programming problem by the ranking function, *Applied Mathematical Science*, 2, 19–32.
- Buckley, J.J., 1988, Possibilistic Linear programming with triangular fuzzy numbers, *Fuzzy Sets and Systems*, 26, 159–174.
- Buckley, J.J., Feuring, T., 2000, Evolutionary algorithm solution to fuzzy problems: fuzzy linear programming, *Fuzzy Sets and Systems*, 109, 35–53.
- Carlsson, C., Corhonen, P., 1986, A parametric approach to fuzzy linear programming, *Fuzzy Sets and Systems*, 20, 17–30.
- Chanas, S., (1983), The use of parametric programming in FLP, *Fuzzy sets and Systems*, 11, 243–251.
- Dehghan, M., Hashemi, B., Ghathe, M., 2006, Computational methods for solving fully fuzzy linear systems, *Applied Mathematics and Computations*, 179, 328–343.
- Hashemi, S.M., Modarres, M., Nasrabadi, E., Nasrabadi, M.M., 2006, Fully fuzzified linear programming, solution and duality, *Journal of Intelligent & Fuzzy Systems*, 17, 253–261.
- Kumar, A., Kaur, J., Singh, P., 2011, A new method for solving fully fuzzy linear programming problems, *Applied Mathematical Modeling*, 35, 817–823.
- Lai, Y.J., Hwang, C.L., 1992, A new approach to some possibilistic linear programming problems, *Fuzzy Sets and Systems*, 49, 121–133.
- Lotfi, F.H., Allahviranloo, T., Jondabeha, M.A., Alizadeh, L., (2009), Solving a fully fuzzy linear programming using lexicography method and fuzzy approximate solution, *Applied Mathematical Modeling*, 33, 3151–3156.
- Maleki, H.R., Tata, M., Mashinchi, M., 2000, Linear programming with fuzzy variables, *Fuzzy Sets and Systems*, 109, 21–33.
- Ramik, J., Rimanek, J., 1985, Inequality relation between fuzzy numbers and its use in fuzzy optimization, *Fuzzy Sets and Systems*, 16, 123–138.
- Rommelfanger, H., Hanuscheck, R., Wolf, J., 1989, Linear programming with fuzzy objectives, *Fuzzy Sets and Systems*, 29, 31–48.
- Stanculescu, C., Fortemps, Ph., Install, M., Wertz, V., 2003, Multiobjective fuzzy linear programming problems with fuzzy decision variables, *European Journal of Operational Research*, 149, 654–675.
- Tanaka, H., Okuda, T., Asai, K., 1973, On fuzzy mathematical programming, *Journal of Cybernetics Systems*, 3, 37–46.
- Tanaka, H., Ichihashi, H., Asai, K., 1984, A formulation of fuzzy Linear Programming Problems based on comparison of fuzzy numbers, *Control and cybernetics*, 13, 186–194.
- Tanaka, H., Guo, P., Zimmermann, H.J., 2000, Possibility distributions of fuzzy decision variables obtained from possibilistic linear programming problems, *Fuzzy Sets and Systems*, 113, 323–332.
- Verdegay, J.L., 1982, Fuzzy mathematical programming, In fuzzy Information and Decision Processes, Gupta MM, Sanchez E (eds), North-Holland, 231–236.
- Verdegay, J.L., 1984a, A dual approach to solve the fuzzy Linear Programming Problem, *Fuzzy Sets and Systems*, 14, 131–141.
- Verdegay, J.L., 1984b, Application of fuzzy optimization in operational research, *Control and Cybernetics*, 13, 229–239.
- Warners, B., 1987a, Interactive Multiple Objective Programming subject to flexible constraint, *European Journal Of Operational Research*, 31, 342–349.
- Warners, B., 1987b, An Interactive Fuzzy Programming system, *Fuzzy Sets and Systems*, 23, 131–147.
- Zadeh, L.A., 1965, Fuzzy sets, *Information and Control*, 8, 338–353.
- Zimmermann, H.J., 1978, Fuzzy programming and linear programming with several objective functions, *Fuzzy Sets and Systems*, 1, 45–55.
- Zimmerman, H.J., 1976, Describing an optimization of fuzzy systems, *International Journal of General System*, 2, 209–216.

# RESTORING AQUATIC ECOSYSTEMS ON THE BASIS OF THE GMAA DSS

Antonio Jiménez <sup>(a)</sup>, Alfonso Mateos <sup>(b)</sup>

Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid

<sup>(a)</sup>[ajimenez@fi.upm.es](mailto:ajimenez@fi.upm.es), <sup>(b)</sup>[amateos@fi.upm.es](mailto:amateos@fi.upm.es)

## ABSTRACT

The Generic Multi-Attribute Analysis System is a decision support system based on an additive multi-attribute utility model that is intended to allay many of the operational difficulties involved in the multicriteria decision-making process. In this paper we illustrate the application of this decision support system to the restoration of aquatic ecosystems in two contamination scenarios simultaneously taking into account several conflicting objectives, like environmental, social and economic impacts. In the first scenario, the contamination is by radionuclides produced, for instance, by a nuclear plant accident / disaster, like the Chernobyl or, more recently, Fukushima disasters. In the second scenario, intervention strategies are selected to combat eutrophication and the sharp decline in the bird population in Ringkøbing Fjord (Denmark).

Keywords: decision support system, multi-attribute utility theory, restoration of aquatic ecosystems

## 1. INTRODUCTION

Most of the real decision-making problems that we face nowadays are complex in the sense that they are usually plagued with uncertainty, and several conflicting objectives have to be taken into account simultaneously.

Decision support systems (DSS) play a key role in these situations helping decision-makers (DMs) to structure and gain a better understanding of the problem and finally make a decision.

DSS are becoming increasingly popular in environmental management (Stam, Salewicz and Aronson 1998; Tecle, Shrestha and Duckstein 1998; Ito, Xu, Jinno, Hojiri and Kawamura 2001; Poch, Comas, Rodríguez-Roda, Sánchez-Marré and Cortés 2004).

There are a number of examples where multi-attribute utility theory (MAUT) has been used for environmental management problems can be found, e.g., in the field of forest management (Ananda and Herath 2009), natural resource management (Mendoza and Martins 2006)), different fields of water management (Joubert, Stewart, Eberhard 2003), river management (Reichert, Borsuk, Hostmann, Schweizer, Sporri, Tockner, and Truffer 2007; Corsair, Ruch, Zheng, Hobbs and Koonce 2009), landscape ecology (Geneletti 2005), evaluation of farming systems (Prato and Herath 2007), and site

selection for hazardous waste management (Merkhofer, Conway and Anderson 1997).

In this paper we will illustrate the application of the generic multi-attribute analysis (GMAA) system to the restoration of aquatic ecosystems contaminated by radionuclides and to the selection of intervention strategies against eutrophication and the sharp decline in the bird population in Ringkøbing Fjord (Denmark).

We have divided the paper into a further three sections. Section 2 outlines the GMAA system and its main characteristics are introduced, and Sections 3 and 4 address the complex decision-making problems pointed out above.

## 2. THE GENERIC MULTI-ATTRIBUTE ANALYSIS DSS

The GMAA system is a PC-based DSS based on an additive multi-attribute utility model that is intended to allay many of the operational difficulties involved in the Decision Analysis (DA) cycle (Jiménez, Ríos Insua and Mateos 2003, 2006, Jiménez and Mateos 2011).

The GMAA system accounts for uncertainty about the alternative performances and for incomplete information about the DM's preferences, leading to classes of utility functions and weight intervals. This is less demanding for a single DM and also makes the system suitable for group decision support, where individual conflicting views in a group of DMs can be captured through imprecise answers.

An additive multi-attribute utility function is used to evaluate the alternatives. This is considered to be a valid approach in most practical situations for the reasons described in (Raiffa 1982; Stewart 1996).

The GMAA provides several types of sensitivity analysis (SA). For instance, it computes the *potentially optimal alternatives* among the *non-dominated alternatives*.

*Monte Carlo simulation techniques* enable simultaneous weight changes and generate results that can be easily analyzed statistically to provide more insight into the multi-attribute model recommendations (Mateos, Jiménez and Ríos Insua 2006).

## 3. RESTORING AQUATIC ECOSYSTEMS CONTAMINATED BY RADIONUCLIDES

The restoration of radionuclide contaminated aquatic ecosystem has been studied in depth as a part of several



EU projects in which we have participated: *MOIRA* (1997-98), *COMETES* (1999-2001), *EVANET-HYDRA* (2001-04) and *EURANOS* (2004-08).

A synthetic, flexible and user-friendly computerized DSS, *MOIRA*, was implemented as a part of these projects. The *MOIRA* system included a multi-attribute analyses module for the global assessment of the effectiveness of the intervention strategies. This module was the origin of the *GMAA* system, which was finally built into the final versions of the *MOIRA* system.

The selection of intervention strategies was based on environmental models for predicting the migration of radionuclides through freshwater and coastal ecosystems and the effects of feasible countermeasures on contamination levels. Other social and economic criteria were also taken into account.

The *MOIRA* system was tested on several real scenarios contaminated as a consequence of the Chernobyl accident, like lake Øvre Heimdalsvatn (Jiménez et al., 2003), located in Oppland county (Norway); lake Kozhanovskoe (Ríos Insua, Mateos, Jiménez 2004), located in the region of Bryansk (Russia); and lake Svyatoye in Belarus (Ríos Insua, Gallego, Jiménez, Mateos 2006).

Specifically, lake Kozhanovskoe was heavily contaminated with  $^{137}\text{Cs}$  after the Chernobyl accident in 1986. In 1998, it was classed as a radio-ecological reserve, and fishery was officially forbidden because of the high levels of fish contamination with  $^{137}\text{Cs}$ . The population around the lake, which lay in the population evacuation zone, was evacuated, as the levels of contamination with  $^{137}\text{Cs}$  were rather high ( $^{137}\text{Cs}$  fallout on the lake was about  $600000 \text{ Bq/m}^2$ ). However, many residents continued to live in villages near the lake, and fish caught in lake Kozhanovskoe were a predominant source of food for local residents even 10 years on from the Chernobyl accident.

An objective hierarchy was built for this decision-making problem, which intended to provide the grounds for the description and evaluation of the hypothetical restoration alternatives for the scenario in question, see Figure 1.

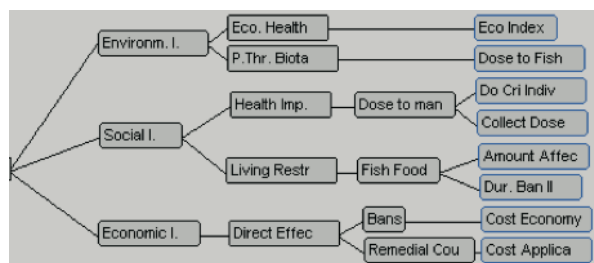


Figure 1. Objective Hierarchy

*Environmental Impact* (Environm. I.) is one of the main objectives of the decision analysis. It was divided into *Lake Ecosystem Index* (Eco Index), a simple and rational approach for measuring the ecological status of a lake, and *Radiation Dose to Biota* (Dose to Fish). *Social Impact* (Social I.) was handled by two sub-

objectives: *Minimizing Impact on Health* (Health Imp.) and *Living Restrictions* (Living restr).

Regarding the health impact, we focused on the *Dose to Critical Individuals* (Do Cri Indiv), who should never receive radiation levels above thresholds for early health effects, and *Collective Dose* (Collect Dose), which was linearly related to the increase in the risk of developing serious latent effects, mainly cancers. As regards living restrictions, other impacts were taken into consideration. These include countermeasures affecting the direct consumption of fish for food or its processing in the food industry, drinking water and water used by the food industry, the use of water for crop irrigation and the recreational uses of water bodies. For all these objectives, the attributes were the *Amount of Fish Affected* by restrictions (Amount Affec), as well as the *Duration of Restrictions* (Dur. Ban II).

Finally, *Economic Impact* (Economic I.) focused on *Direct Effects* (Direct Eff), which included the costs generated by the different bans or restrictions ON normal living conditions, which can be subdivided into *Costs to the Economy* (Cost Economy) and *Application Costs* (Cost Applica), i.e., costs of chemical and physical remedial countermeasures.

The following six strategies were considered for evaluation:

- *No Action*. Natural evolution of the situation without intervention.
- *Potash Treatment*. Reduction of aquatic organism uptake by potash treatment of aquatic ecosystems contaminated by radiocesium.
- *Fertilization*. Tonnes of fertilizer added to the lake to increase biomass.
- *Lake Liming*. Reduction of radionuclide remobilisation from sediments.
- *Sediment Removal*.  $6 \text{ km}^2$  of sediments removed from the lake down to depth of 5 cm.
- *Automatic Food Bans*. Automatic fish consumption ban when  $^{137}\text{Cs}$  content in fish is greater than  $1000 \text{ Bq/kg}$ .

The impacts of the intervention strategies were then established in terms of the attributes associated with the lowest-level objectives (see Table 2 in Gallego, Jiménez, Mateos, Sazykina, Ríos-Insua, Widengård 2001).

DM preferences were elicited according to the *DA* cycle. An imprecise component utility function was assessed for each attribute, representing DM preferences concerning the respective possible attribute impacts. Figure 2 shows the class of utility functions for *Dose to Fish*.

On the other hand, objective weights representing their relative importance were elicited along the branches of the objectives hierarchy. Then, the attribute weights used in the additive multi-attribute utility model were assessed by multiplying the elicited weights in the path from the overall objective to the respective attributes (see Figure 3). These attribute weights are

indicators of the influence of the individual criteria on the decision.

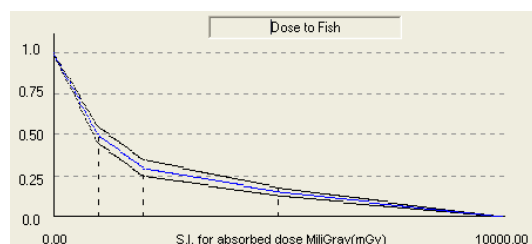


Figure 2. Component Utility Functions for *Dose to Fish*

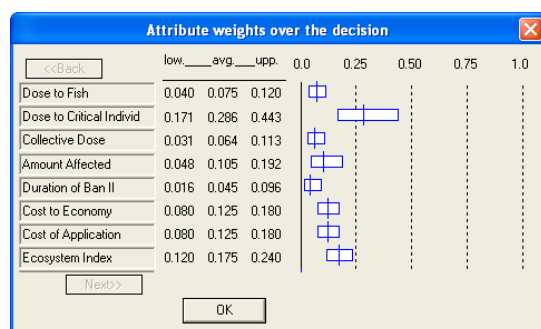


Figure 3. Relative Importance of Attributes

The additive multi-attribute utility model, which demands precise values, was then used to assess, on the one hand, average overall utilities, on which the ranking of alternatives is based and, on the other, minimum and maximum overall utilities, which give further insight into the robustness of this ranking.

Figure 4 shows the ranking of the intervention strategies for lake Kozhanovskoe, where the vertical white lines on each bar represent average utilities. The best-ranked intervention strategy was *Automatic Food Bans* with an average overall utility of 0.8245, followed by *Lake Liming* (0.5794) and *Potash Treatment* (0.5592), whereas the worst ranked option was *Sediment Removal* with a utility of 0.4663.

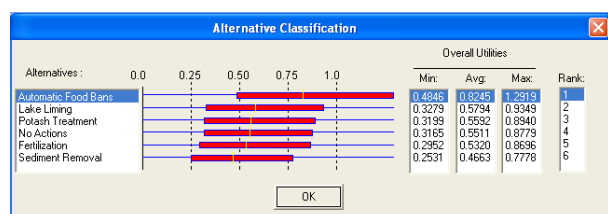


Figure 4. Ranking of Intervention Strategies

Although *Automatic Food Bans* seemed to outperform the other intervention strategies on the basis of the average overall utilities. But, looking at the utility intervals, the robustness of this ranking is questionable because there is a big overlap between the output intervals, raising doubts about recommending this strategy. Consequently, a SA should be carried out to provide further insight into the recommendations.

First, only two strategies were non-dominated and potentially optimal, i.e., they were not dominated by

any other strategy and best-ranked for at least one combination of the imprecise parameters, i.e., weights, component utilities and strategy impacts. Thus, the SA focused the analysis on these strategies, *Automatic Food Bans* and *Lake Liming*. Then, Monte Carlo simulation techniques were applied. Attribute weights were randomly assigned values taking into account the weight intervals provided by the DMs in weight elicitation (see Figure 3). In the 10000 trials performed, *Automatic Food Bans* outperformed *Lake Liming*, i.e., it was best ranked.

Moreover, if attribute weights were generated completely at random, which would mean that there is no knowledge whatsoever of the relative importance of the attributes, then *Automatic Food Bans* would outperform *Lake Liming* by more than 60% (see mean values in Figure 5).

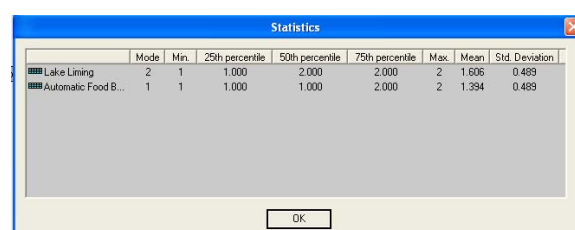


Figure 5. Multiple Bloxplot Statistics

#### 4. SELECTING INTERVENTION FOR THE RESTORATION OF RINGKØBING FJORD

The selection of intervention strategies against eutrophication and the drastic decrease in the bird population in Ringkøbing Fjord was studied at length in (Brhyn, Jiménez, Mateos, Ríos Insua 2009; Jiménez, Mateos, Brhyn, 2011),

Ringkøbing Fjord is a large and shallow brackish lagoon on the west coast of Denmark. It has an area of 300 km<sup>2</sup>, a volume of 0.57 km<sup>3</sup>, a maximum depth of 5.1 m and a mean depth of 1.9 m. The lagoon receives large (2 km<sup>3</sup>yr<sup>-1</sup>) freshwater inputs from the catchment, as well as saltwater inputs through a sluice that connects the lagoon to the sea (see Figure 6).

Ringkøbing Fjord has gone through two environmental regime shifts during the last decades (Håkanson, Bryhn and Eklund 2007), which has stirred up public sentiment in the area, mainly because of the disappearance of waterfowl.

The following nine intervention strategies were considered for analysis:

- *S1: 10% P abatement*. Reducing the *P* input by 10%.
- *S2: 33% P abatement*. Reducing the *P* input by 33%.
- *S3: 10% N+P abatement*. Reducing the nutrient input by 10%.
- *S4: 33% N+P abatement*. Reducing the nutrient input by 33%.
- *S5: Sluice*. Building a pumping station or another sluice between the lagoon and the sea to increase the saltwater inflow.
- *S6: Salt 7.2*. Reducing the salinity to 7.2‰.

- *S7: 10% P abatement + Sluice*. A combination of intervention strategies S1 and S5.
- *S8: 33% P abatement + Sluice*. A combination of intervention strategies S2 and S5.
- *S9: No action*. Natural evolution of the situation without intervention

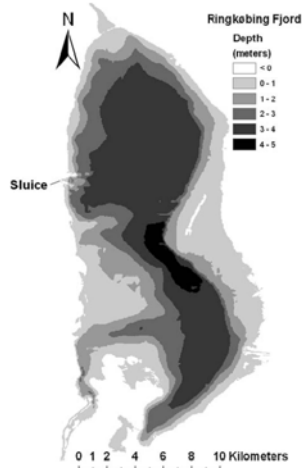


Figure 6: Ringkøbing Fjord

Intervention strategies were evaluated considering their environmental, social and economic impacts (see Figure 7). There were two attributes stemming from the *environmental impact*, *natural TRIX deviation* (N TRIX Dev) and *number of birds* (N. of Birds). The degree of eutrophication in a coastal area can be expressed as a *TRIX* (TRophic state Index) deviation from the background value. The attribute associated with this lowest-level objective represented the average *TRIX* deviation against previous years over a 20-year period.

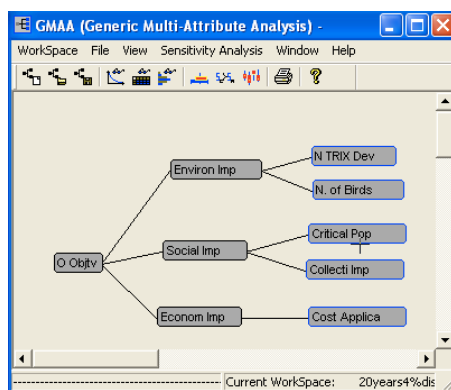


Figure 7. Objective Hierarchy for Ringkøbing Fjord

Another environmental impact we took into account was related to the sharp fall in bird-days over the year in recent decades. The associated attribute accounted for the number of birds representing the average number of Bewick's swans and pintails living in the lagoon in a year for the time period under consideration.

Regarding the *social impact* we made a distinction between the *social impact for critical population* (Critical Pop), i.e., people living around the lagoon that

may be affected by the application of intervention strategies, and *collective social impact* (Collecti Imp). Both subjective attributes account for aspects like sentiment, possible employment associated with strategy application, crop image...

Finally, the *economic impact* was computed by the average costs concerning the intervention strategy application (Cost Applica), i.e., nutrient abatement costs and/or construction and maintenance costs for facilities.

Note that while the models or experts initially provided precise performances, imprecision was introduced by means of an attribute deviation of 10% to evaluate the robustness of the evaluation (see Table 4 in Jiménez, Mateos, Brhyn, 2011).

Next, the DMs' preferences were quantified. This implies assessing, on the one hand, component utilities in attributes that represent the DMs' preferences concerning the possible intervention strategy impacts, and, on the other, local weights, which represent the relative importance of criteria in the objective hierarchy.

Regarding the utility function corresponding for the *natural TRIX deviation* attribute, the background *TRIX* of 4.3–5.1 was estimated. This was used as a baseline for optimal conditions in this study. On the other hand, *TRIX* in 1990–1993 was 6.3. This was identified as the least preferred value. Taking into account that the attribute represents the average *TRIX* deviation against previous years over a 20-year period and the above information, attribute values between [0, 0.4] were assigned a utility 1. Then, the utility function was linearly decreasing in the range [0.4, 1.6], and for any average deviation greater than 1.6, the associated utility was 0 (see Figure 8).

Regarding the *number of birds*, the component utility function is straightforward. The best attribute value is assumed to be 100,000 birds (Bewick's swans and pintails) and the worst one is 0 (no birds). Consequently, the utility function is increasing. For a description of the component utilities function corresponding the remaining attributes, see (Jiménez, Mateos, Brhyn 2011).

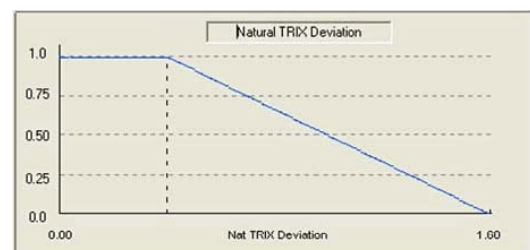


Figure 8: Component Utility Function For *Natural TRIX Deviation*

On the other hand, the ecocentric, anthropocentric and tax-refuser perspectives were used to elicit different weight sets. A pure ecocentrist would assign a weight of 100 to environmental impacts. However, a pure anthropocentrist would assign no weight to the environmental impact but distribute all the weights across the social and economic impacts. Finally, a

persistent tax-refuser would assign total weight to the economic impact and no weight to social or environmental impact. Shrivastava (1995) and Rauschmayer (2001) give reasons why different perspectives may be needed can be found.

Finally, Monte Carlo simulation techniques were applied. They can be useful for analyzing the intervention strategies from the ecocentric and anthropocentric perspectives. The attribute weights were selected at random, and the computer-simulated ranking of attribute importance from different perspectives was stored to efficiently explore the results of many weight combinations. Figure 8 shows the resulting multiple boxplots.

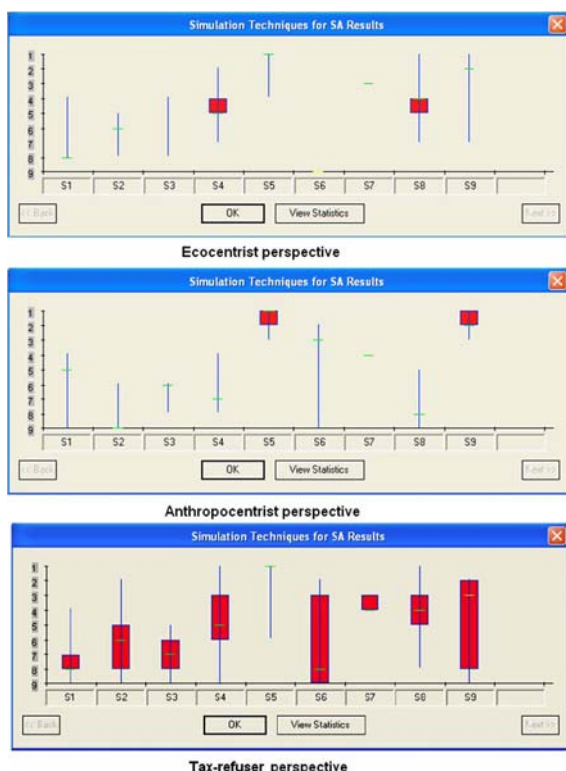


Figure 8: Strategy Evaluation From Different Perspectives

Looking at the multiple box plots for the ecocentric and anthropocentric perspectives, we find that S5: *Sluice* and S9: *No action* are ranked highest in both boxplots. S8: 33% *P abatement* + *Sluice* is ranked highest from the ecocentric viewpoint, but its best ranking from the anthropocentric perspective is fifth. Finally, S6: *Salt7.2*, with a best ranking of second from the anthropocentric viewpoint, is ranked as the worst strategy from the ecocentric perspective. S5: *Sluice* and S9: *No action* look better than the others. Moreover, the average rankings for both are 1.011 and 2.489 from the ecocentric perspective, respectively, and 1.531 and 1.605 from the anthropocentric viewpoint. These results are even consistent with the tax-refuser perspective, in which S5 is better ranked (average ranking 1.246) than S9 (average ranking 4.621). Thus, we arrived at the conclusion that S5: *Sluice* was the intervention strategy

to be recommended.

Moreover, if we assume that there is no knowledge whatsoever of the relative importance of the attributes, i.e., weights for the attributes are generated completely at random, S5: *Sluice* was again the best intervention strategy throughout the simulation.

The same methodology was applied for different interest rates (0, 2, 4, 6, and 8%), and we arrived at the same conclusion that S5: *Sluice* was the intervention strategy to be recommended.

## 5. CONCLUSIONS

Most environmental decision-making problems have multiple conflicting objectives and are usually plagued with uncertainty, being impossible to predict with certainty what the consequences of each strategy under consideration will be.

The GMAA system is a decision support system that to allay many of the operational difficulties involved in a decision-making problem by helping decision makers to structure and achieve a better understanding of the problem to make a final decision.

Two examples illustrate the application of the GMAA system and prove that sensitivity analysis tools can be useful to provide further insight into the recommendations.

## ACKNOWLEDGMENTS

The paper was supported by Madrid Regional Government project S-2009/ESP-1685 and the Spanish Ministry of Education and Science project TIN 2008-06796-C04-02.

## REFERENCES

- Ananda, J. Herath, G., 2009. A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics*, 68, 2535–2548.
- Bryhn, A. Jiménez, A. Mateos, A., Ríos-Insua, S., 2009. Multi-attribute analysis of trophic state and waterfowl management in Ringkøbing fjord, Denmark. *Journal of Environmental Management* 90 (8), 2568–2577.
- Corsair, H.J., Ruch, J.B., Zheng, P.Q., Hobbs, B.F., Koonce, J.F., 2009. Multicriteria Decision Analysis of Stream Restoration: Potential and Examples. *Group Decision and Negotiation*, 18, 387–417.
- Gallego, E., Jiménez, A., Mateos, A., Sazykina, T., Ríos-Insua, S., Windergård, M., 2001. Application of multiattribute analysis methodologies to the evaluation of the effectiveness of remedial strategies with the MOIRA System. In: *Implementing Computerised Methodologies to Evaluate the Effectiveness of Countermeasures for Restoring Radionuclide Contaminated Fresh Water Ecosystems*. Rome: ENEA, 111–130.
- Geneletti, D., 2005. Formalising expert opinion through multi-attribute value functions: an application in



- landscape ecology. *Journal of Environmental Management*, 76, 255–262.
- Håkanson, L., Bryhn, A.C., Eklund, J.M., 2007. Modelling phosphorus and suspended particulate matter in Ringkøbing fjord in order to understand regime shifts. *Journal of Marine Systems*, 68, 65–90.
- Ito, K., Xu, Z.X., Jinno, K., Hojiri, T., Kawamura, A., 2001. Decision support system for surface water planning in river basins. *Journal of Water Resources Planning and Management*, 127, 272–276.
- Jiménez, A., Ríos-Insua, S., Mateos, A., 2003. A Decision support system for multiattribute utility evaluation based on imprecise assignments. *Decision Support Systems*, 36 (1), 65–79.
- Jiménez, A., Ríos-Insua, S., Mateos, A., 2006. A generic multi-attribute analysis system. *Computers and Operations Research*, 33 (4), 1081–1101.
- Jiménez, A., Mateos, A., 2011. GMAA: A DSS based on the decision analysis methodology. Application survey and further developments, In: C. Jao (ed.). *Efficient Decision Support Systems: Practice and Challenges – From Current to Future*. InTech Open Access Publisher, p. 26 (to appear).
- Jiménez, A., Mateos, A., Brhyn, A., 2011. Selecting intervention strategies against eutrophication and the drastic decrease in bird abundance in Ringkøbing fjord. *TOP An Official Journal of the Spanish Society of Statistics and Operations Research* (to appear, DOI: 10.1007/s11750-010-0136-x).
- Joubert, A., Stewart, T.J., Eberhard, R., 2003. Evaluation of water supply augmentation and water demand management options for the city of Cape Town. *Journal of Multi-Criteria Decision Analysis*, 12, 17–25.
- Mateos, A., Jiménez, A., Ríos-Insua, S., 2006. Monte Carlo simulation techniques for group decision-making with incomplete information. *European Journal of Operations Research*, 174 (3), 1842–1864.
- Mendoza, G.A., Martins, H., 2006. Multi-Criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230, 1–22.
- Merkhofer, M.W., Conway, R., Anderson, R.G., 1997. Multiattribute utility analysis as a framework for public participation in siting a hazardous waste management facility. *Environmental Management*, 21, 831–839.
- Poch, M., Comas, J., Rodríguez-Roda, I., Sánchez-Marré, M., Cortés, U., 2004. Designing and building real environmental decision support systems. *Environmental Modeling and Software*, 19 (9), 857–873.
- Prato, T., Herath, G., 2007. Multiple-criteria decision analysis for integrated catchment management. *Ecological Economics*, 63, 627–63.
- Raiffa, H., 1982. *The Art and Science of Negotiation*. Cambridge: Harvard University Press.
- Rauschmayer, F., 2001. Reflections on ethics and MCA in environmental decisions. *Journal of Multi-Criteria Decision Analysis*, 10, 65–74.
- Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Sporri, C., Tockner, K., Truffer, B., 2007. Concepts of decision support for river rehabilitation. *Environmental Modelling and Software*, 22, 188–201.
- Ríos-Insua, S., Jiménez, A., Mateos, A., 2004. A Time-dependent decision support system for multiattribute decision-making. *Integrated Computer-Aided Engineering*, 11 (1), 63–75.
- Ríos-Insua, S., Gallego, E., Jiménez, A., Mateos, A., 2006. A multi-attribute decision support system for selecting environmental intervention strategies for radionuclide contaminated freshwater ecosystems. *Ecological Modelling*, 196, 195–208.
- Shrivastava, P., 1995. Ecocentric management for a risk society. *Academy of Management Review*, 20, 118–137.
- Stam, A., Salewicz, K.A., Aronson, J.E., 1998. An interactive reservoir system for lake Kariba. *European Journal of Operational Research*, 107, 119–136.
- Stewart, T.J., 1996. Robustness of additive value function method in MCDM. *Journal of Multi-Criteria Decision Analysis*, 5, 301–309.
- Teale, A., Shrestha, B.P., Duckstein, L., 1998. A multiobjective decision support system for multiresource forest management. *Group Decision and Negotiation*, 7, 23–40.

## AUTHORS BIOGRAPHY

**Antonio Jiménez Martín** received his degree in a Computer Science from Universidad Politécnica de Madrid (UPM). He is currently an Associate Professor of Operations Research and Simulation at the UPM's School of Computing Science. His research interest is decision and risk analysis and management, and he is involved in the development and implementation of decision support systems based on multi-attribute utility theory with applications to environment restoration and e-democracy. His research has been published mainly in COR, DSS, EJOR, JMDA, GDN, JORS, Ecological Modelling... and he has co-authored three books.

**Alfonso Mateos Caballero** is an Associate Professor of Statistics and Operations Research at the School of Computing Science (Universidad Politécnica de Madrid). His research interest is decision analysis and risk analysis and management, and he is currently involved in the development intelligent decision support systems based on influence diagrams and multi-attribute utility theory with applications to the environment, medicine and e-democracy. His articles have been published in several academic journals including: DSS, EJOR, JORS, Computational Optimization and Applications, GDN, JMDA, and Ecological Modelling.

# PLANNING HIGHWAYS RESURFACING USING COMPUTER SIMULATION

Mohamed Marzouk<sup>(a)</sup>, Marwa Fouad<sup>(b)</sup>, Moheeb El-Said<sup>(c)</sup>

<sup>(a)</sup> Associate Professor, Structural Engineering Department, Faculty of Engineering, Cairo University, Egypt

<sup>(b)</sup> Graduate Student, Structural Engineering Department, Faculty of Engineering, Cairo University, Egypt

<sup>(c)</sup> Professor, Structural Engineering Department, Faculty of Engineering, Cairo University, Egypt

<sup>(a)</sup> [mm\\_marzouk@yahoo.com](mailto:mm_marzouk@yahoo.com), <sup>(b)</sup> [eng.marwafouad\\_2006@yahoo.com](mailto:eng.marwafouad_2006@yahoo.com), <sup>(c)</sup> [elsaid1204@yahoo.com](mailto:elsaid1204@yahoo.com)

## ABSTRACT

Resurfacing highway project is one of the complex projects that are characterized by repetitive operations, difficult construction environment, and different construction tools. Such characteristics lead to uncertainties with respect to estimated duration and resurfacing cost. This paper presents a framework that is used for planning of highways resurfacing using computer simulation and optimization. The proposed framework has two features; 1) planning of repairing highways, and 2) selecting work zone length that achieves minimum total cost and total duration. A numerical example is presented to demonstrate the use of the proposed framework.

Keywords: highways resurfacing, computer simulation, optimization, generic algorithms

## 1. INTRODUCTION

Work zones often cause traffic congestion on high volume roads. As traffic volumes increase so does work zone-related traffic congestion. Negative impacts on road users can be minimized by bundling interventions on several interconnected road sections instead of treating each road section separately. Negative impacts on road users can be quantified in user costs. The optimum work zone is the one that results in the minimum overall agency and user costs. The minimization of these costs is often considered in planning. In order to achieve this goal the interventions on each asset type (pavement, bridges, tunnels, hardware, etc.) must be bundled into optimum packages. Hajdin and Lindenmann (2007) presented a method that enables road agencies to determine optimum work zones and intervention packages. The method allows the consideration of both budget constraints and distance constraints, including maximum permissible work zone length or minimum distance between work zones. The mathematical formulation of this optimization problem is a binary program that can be solved by existing techniques (i.e., the branch-and-bound method).

Pavements on two-lane two-way highways are usually resurfaced by closing one lane at a time. Vehicles then travel in the remaining lane along the

work zone, alternating directions within each control cycle. Several alternatives can be evaluated. Each alternative is defined by the number of closed lanes and fractions of traffic diverted to alternate routes. Chen et al. (2005) defined an algorithm, referred to as SAUASD (simulated annealing for uniform alternatives with a single detour), to find the best single alternative within a resurfacing project. It searches through possible mixed alternatives and their diverted fractions, to minimize total cost, including agency cost (resurfacing cost and idling cost) and user cost (user delay cost and accident cost). Thus, traffic management plans are developed with uniform or mixed alternatives within a two-lane highway resurfacing project.

Wang et al. (2002) proposed a hybrid genetic algorithm-microscopic traffic simulation methodology for scheduling of pavement maintenance activities involving lane closures aiming at minimizing network traffic delay. Genetic algorithm is implemented as an optimization tool for the generation and selection of maintenance schedules. For each possible schedule, a microscopic traffic simulation model is used to simulate traffic operations in the network to estimate the delay caused by lane closures. This concept has been implemented using genetic algorithm written in the C language and simulation carried out by PARAMICS. This hybrid methodology combines two different approaches: genetic algorithm (GA) and microscopic traffic simulation.

In each GA generation, a population of schedule is generated. Each schedule refers to the spatial and temporal scenario of lane closures in a network over a 24-hour period. Each lane closure schedule is simulated several times, each with a different random number seed. This is to average out any bias in the results (total traffic delay) caused by the random number seed in the simulation input. At the end of each simulation run, the total travel time is fed back to the GA program for delay calculation. GA then evaluates the relative merit of each lane closure schedule by comparing the average total traffic delays, and uses better schedules to produce next generation of solutions. The process continues until a pre-defined generation size is met. Then, the schedule with least total traffic delay is recommended. To arrive at a near optimal lane closure schedule, the GA search

process needs to evolve through many generations, and each generation contains a population of chromosomes. For each unique chromosome, several simulation runs need to be carried out with different random number seeds. Therefore, hundreds of simulation runs are anticipated. The impact on traffic due to construction projects is not completely inevitable. However, for a project that requires multiple work crews, the impact on traffic can be improved by means of the appropriate schedule. This is because of the “simultaneous operations”. Different types of construction projects set up work zones on roads. Especially in urban areas, lane-closures as a result of work zones have a considerable impact on local traffic. However, for a construction project that consists of several work zones and several work crews, the traffic impact may be improved by appropriate scheduling. Lee (2009) proposed a scheduling model based on the route-changing behavior of road users. The proposed model calculates traffic delay of vehicles by microscopic simulation, and applies team ant colony optimization to search for a near-optimal schedule.

## 2. WORK ZONE SAFETY

Work zone safety has been a research focus for many years and improving the safety in highway work zones is a high-priority task for traffic engineers. Yingfeng and Yong (2009) explained evaluating the effectiveness of the Temporary Traffic Control (TTC) methods used in highway work zones which would help traffic engineers to identify traffic control deficiencies, and thus, make continuous improvement in safety. In this study, the effectiveness of several TTC methods (including flagger/officer, stop sign/signal, flasher, no passing zone, and center/edge lines) in mitigating work zone crash severity and preventing common human errors from causing severe work zone crashes was quantified using a logistic regression technique. The findings may provide valuable knowledge for traffic engineers in understanding the effects of the TTC methods on the severity or involvement of certain human errors in work zone crashes. They may also provide insights on safety implications of the work zone environment associated with each evaluated TTC method. According to the logistic regression analyses, the presence of a flagger or officer directing traffic could reduce the odds of having fatalities in a severe crash; having flashers or center/ edge lines in work zones could reduce the odds. However, based on the available crash data, the statistics did not support close associations between the usages of stop signs/signals and no-passing-zone control in work zones and fatality involvement in severe crashes.

Regarding the effectiveness of TTC methods in preventing common human errors from causing severe crashes in work zones, the evaluation showed that flaggers/officers could considerably lower the odds of severe work zone crashes caused by human errors such as "disregarded traffic control," "inattentive driving," and "exceeded speed limit or too fast for condition."

No-passing zone control in work zones was effective in reducing the odds of having severe crashes caused by "disregarded traffic control". In addition, having center/edge lines in work zones could lower the odds of having severe work zone crashes caused by human errors such as "exceeded speed limit or too fast for condition" and "followed too closely." However, having stop signs/signals in work zones would dramatically increase the odds of having severe crashes caused by "followed too closely" human error. Logistic regression analyses were used to assess individual TTC methods so that quantified estimations of the effectiveness of each TTC could be obtained. The actual effectiveness of these methods may vary when used in combination with other traffic control devices and/or work zone conditions. This research can be extended in several ways. Fatal crash data from other sources could be added to increase the total number of fatal cases in order to improve the reliability of the analysis (Yingfeng and Yong 2009).

## 3. PROPOSED SYSTEM

This research presents a framework, named *Resurfacing\_Sim*, which helps contractors in planning of highway resurfacing operation (Fouad 2011). The developed framework performs planning and optimizing resurfacing highway and selects optimum length of work zone based on minimum total cost and total duration. These two functions are performed by two main components, simulation module and optimization module. The following sub-sections describe the developments made in these two modules.

### 3.1. Simulation Module

Developing a simulation model requires a sequence of tasks to represent the resurfacing operation, the relationships between these tasks. To simulate the process of highways resurfacing, the steps of simulation experiment are performed:

1. Define General data: General data include number of working hours per day, number of lanes, number of working days in week, number of work zones.
2. Define Zones data: Assigning labor crew types, number of crews, the number of utilized equipment, and production rates of the different crew types and equipment.
3. Define Tasks data: The contractor estimates the duration of involved tasks and materials costs. The framework provides the contractor with the probability distributions to estimate task duration; Normal, Uniform, Triangular, Part, and Pertpg distributions.
4. Run Simulation Module: After feeding input data, the Simulation Module estimates total duration and after that it calculates total cost for each zone. The contractor has to define two values of two simulation parameters; number of simulation runs and confidence level. These two parameters are required to account for probabilistic input data.

Converting traffic flow involves three main processes: 1) laying safety control devices, 2) breaking concrete platform and median, and 3) removing the broken concrete from work zone. Table 1 lists processes and tasks of converting traffic flow, whereas, Table 2 lists the types of resources that are needed to converting traffic flow. The input parameters that are used in the tasks are listed in Table 3. Figure 1 depicts the element of the network that captures the Converting Traffic Flow. Similarly, simulation networks are developed for the remaining processes including; modeling reconstruction of semi-rigid paving process and modeling reconstruction flexible paving and finishing process.

### 3.2. Optimization Module

The cost per lane-kilometer is the objective function for the optimization module. It consists of three components as shown in Equation 1. These components are; i) maintenance cost per lane-kilometer ( $C_m$ ), ii) delay cost per lane-kilometer ( $C_u$ ) including queue delay cost and moving delay cost, and iii) accident cost per lane-kilometer incurred by traffic passing the work zone ( $C_a$ ).

Table 1: Processes and Tasks of Converting Traffic Flow

Process	Task
Laying safety control devices	Safetydevices
Breaking concrete platform & median	BrkenPlatform
Removing broken concrete from work zone	Load
	Haul
	Dump
	Retune

Table 2: Resources used in Converting Traffic Flow Tasks

Resource Type	Resources
Labors	Safety crew
Equipment	Loader – Jack hammer – Trucks
Materials	Area of Work Zone – Volume of Waste Materials

Table 3: Input Parameters of Converting Traffic Flow Model

Parameter	Description
AWZ	Area of Work Zone
VOSOIL	Volume of waste material
NOJAC	Number of jack hummer
NTRUCK	Number of trucks
NOLOD	Number of loaders
SC	Number of safety crew

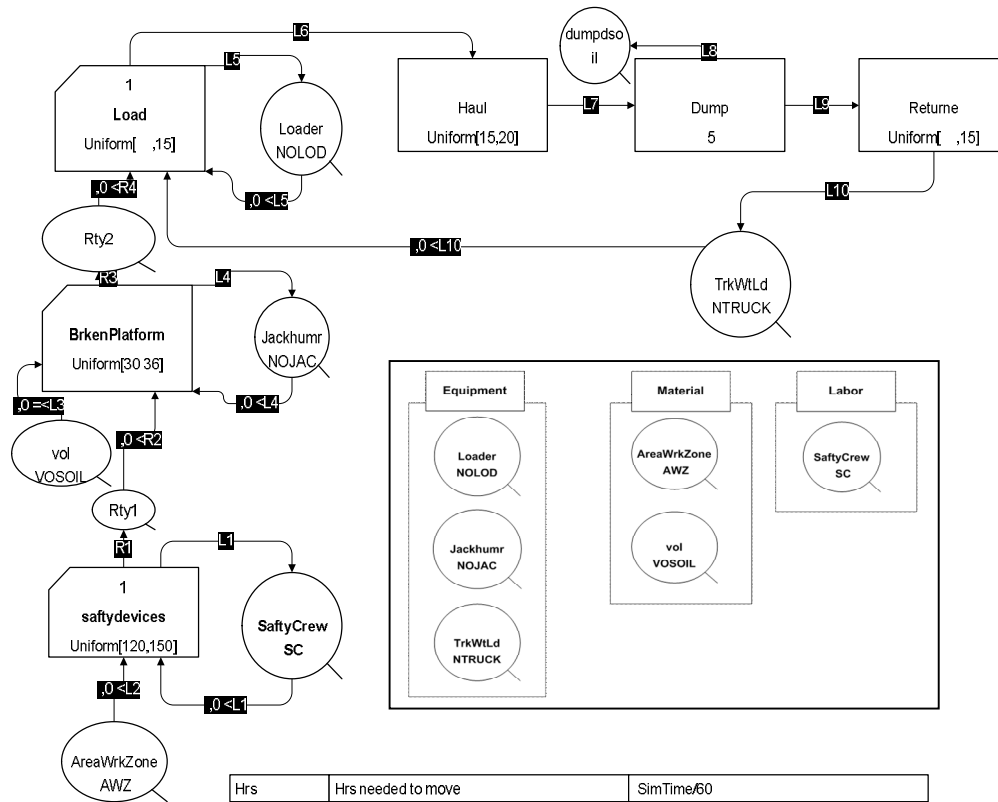


Figure 1: Converting Traffic Flow Simulation Network



$$C_T = C_m + C_u + C_a \quad (1)$$

The optimization variables are defined as any entities within studied system, where any change in this entity would seriously affect the observed optimization functions. Based on interviews with expert engineers and extensive analysis of resurfacing operation, optimization variables have been determined. Six optimization variables are considered:

1. *Hourly Flow Rate in Direction 1 ( $Q_1$ )*: Number of vehicle in the same direction with work zone.
2. *Hourly flow rate in Direction 2 ( $Q_2$ )*: Number of vehicle in opposite direction against work zone.
3. *Average maintenance time per lane-kilometer ( $Z_4$ )*: the required duration for maintenance for each lane per kilometer
4. *Work zone length ( $L$ )*: the optimum length for work zone that decreases delay in traffic time and decreases accidents.
5. *Average work zone speed ( $V$ )*: speed of vehicle at work zone.
6. *Average headway ( $H$ )*: the time of the distance between two vehicles as shown in Figure 2.

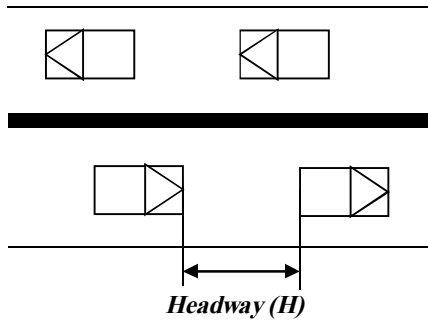


Figure 2: Average headway

#### 4. NUMERICAL EXAMPLE

In order to demonstrate the use of the developed framework in optimization highways resurfacing operation, an actual project example is considered. The example involves a highway with a length of 15 Km. Some of parameters are considered constant for the project such as: 1) fixed cost for setting up a work zone ( $z_1$ ), 2) average additional maintenance cost per work zone unit length ( $z_2$ ), 3) setup time ( $z_3$ ), 4) value of user time ( $v$ ), 5) road original speed in normal conditions ( $V_0$ ), 6) number of accidents per 100 million vehicle hours ( $n_a$ ), and 7) average cost per accident ( $v_a$ ). These parameters are listed in Table 4. The other parameters are considered optimization parameters which are listed in Table 5.

Several experiments have been carried out to test the performance of the optimization module against different values of crossover threshold (CO), mutation threshold (m), and number of generations (G). Figure 3 shows the change in Total cost at different mutation thresholds. Whereas, Figure 4 shows the change in total cost at different crossover thresholds. The results reveal

that solutions are too sensitive to both crossover and mutation values. For this road example, best solutions for minimum cost are obtained at CO=0.5 and m=0.1.

Table 4: The Value of Input Parameters

Parameter	Value
$Z_1$	80 L.E/Zone
$Z_2$	160 L.E/Lane. Km
$Z_3$	10 hr/Zone
$v$	12.7 L.E/Veh. Hr
$V_0$	80 Km/ hr
$n_a$	67 Accident/ 100mvh
$v_a$	17.6 L.E/ hr

Table 5: Optimization Parameters

Parameter	Minimum Value	Maximum Value
$Q_1$	3000 Veh/hr.	5000 Veh/hr.
$Q_2$	3000 Veh/hr.	5000 Veh/hr.
$Z_4$	10 Hr/Lane.Km	20 Hr/Lane.Km
$L$	1 Km	15 Km
$V$	10 Km/ hr	15 Km/hr.
$H$	2 sec	10 sec

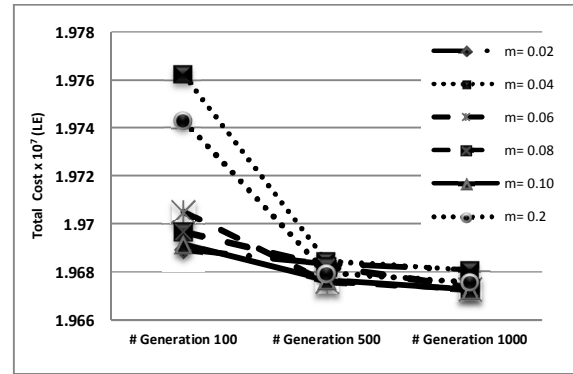


Figure 3: Total Cost vs. No. Of Generations at Different Mutation Thresholds (CO= 0.5)

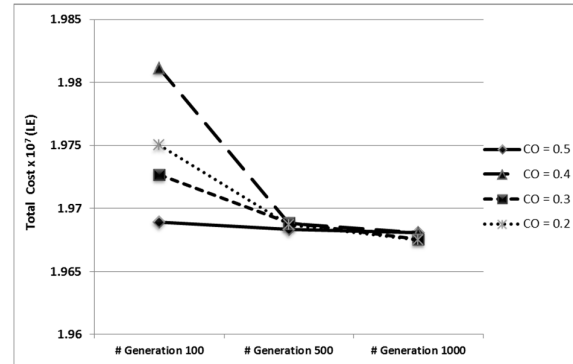


Figure 4: Total Cost vs. No. Of Generations at Different Crossover Thresholds (m = 0.02)

#### 5. SUMMARY

This paper presented a framework that is dedicated for highways resurfacing using computer simulation and optimization. The framework consists of two main

components; simulation module and optimization module. Simulation module is responsible for estimating total duration for each zone. The simulation module is implemented using the Microsoft Visual basic 6.0 programming language and it utilizes STROBOSCOPE (general purpose simulation language) as simulation engine. It estimates the duration for each resurfacing zone which is used to calculate the total duration and total cost for the overall highway resurfacing operation. The total cost is essentially the objective function of optimization module. This cost is calculated by summing up the direct cost (resurfacing operation cost), indirect cost, and the cost of the impact of work zone on road. The optimization module considers six optimization variables including; i) hourly flow rate in Direction 1, ii) hourly flow rate in Direction 2, iii) average maintenance time per lane-kilometer, iv) work zone length, v) average work zone speed, and vi) average headway. A numerical example was presented to demonstrate the main features of the proposed framework.

## REFERENCES

- Chen, C.H., Schonfeld, P., and Paracha, P. 2005. Work zone optimization for two-lane highway resurfacing projects with an alternate route. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1911, TRB, National Research Council, Washington, D.C., 51-66.
- Fouad, M. 2011. Optimizing highways resurfacing using computer simulation." Thesis (MSc), Cairo University, Cairo, Egypt.
- Hajdin, R., Lindenmann, H.P. 2007. Algorithm for the planning of optimum highway work zones. *Journal of Infrastructure Systems*, 13(3), 202-214.
- Lee, H.Y. 2009. Optimizing schedule for improving the traffic impact of work zone on roads. *Automation in Construction*, 18 (8), 1034-1044.
- Wang, Y., Cheu, R.L., Fwa, T.F., 2002, Highway maintenance scheduling using genetic algorithm with microscopic traffic simulation. In *Proceedings of the 81<sup>st</sup> Annual Meeting of the Transportation Research Board*, paper #02-2174.
- Yingfeng, L.A., Yong, B., 2009, Effectiveness of temporary traffic control measures in highway work zones. *Safety Science*, 47(3), 453-458.

## AUTHORS BIOGRAPHY

**MOHAMED MARZOUK** is Associate Professor at Department of Structural Engineering, Faculty of Engineering, Cairo University. He received his B.Sc. and M.Sc. in Civil Engineering from Cairo University in 1995 and 1997, respectively. He received his Ph.D. from Concordia University in 2002. His research interest includes simulation and optimization of construction processes, O-O simulation, fuzzy logic and its applications in construction, risk analysis, and decision analysis. His e-mail is <[mm\\_marzouk@yahoo.com](mailto:mm_marzouk@yahoo.com)>.

**MARWA FOUAD** is Graduate Student at Department of Structural Engineering, Faculty of Engineering, Cairo University. She received his B.Sc. and M.Sc. in structural engineering from Cairo University in 2006 and 2011, respectively. Her research interest includes simulation, decision support systems, and construction technology. Her email is <[eng.marwafouad\\_2006@yahoo.com](mailto:eng.marwafouad_2006@yahoo.com)>.

**MOHEEB EL-SAID** is Professor of construction engineering and management at Department of Structural Engineering, Faculty of Engineering, Cairo University. His research interest includes productivity, schedule risk assessment, life cycle costs, and simulation. His e-mail is <[elsaid1204@yahoo.com](mailto:elsaid1204@yahoo.com)>.

# 3 YEARS OF FUNDING INITIATIVE FOR COMPUTATIONAL MATHEMATICS IN AUSTRIA

P. Kersch<sup>(a)</sup>, A. Pogány<sup>(b)</sup>

<sup>(a)</sup> Austrian Research Promotion Agency (FFG) Departments Thematic Programmes, Sensengasse 1, 1090 Vienna, Austria, [www.ffg.at/tp](http://www.ffg.at/tp)

<sup>(b)</sup> Federal Ministry of Transport, Innovation and Technology (bmvit) Department III/I 5 Department for IT, nanotechnology, industrial technologies and space flight, Renngasse 5, 1010 Vienna, Austria, [www.bmvit.gv.at](http://www.bmvit.gv.at)

<sup>(a)</sup> [peter.kersch@ffg.at](mailto:peter.kersch@ffg.at), <sup>(b)</sup> [alexander.pogany@bmvit.gv.at](mailto:alexander.pogany@bmvit.gv.at)

## ABSTRACT

The funding initiative ModSim Computational Mathematics started in 2008. In 2010 there was the third call for project-proposals. This paper describes the results from the three calls and the comparison with respect to the temporal and financial outcome from the three calls. Also, the topics and contribution of different proposers are shown. Each year the number of proposals increased. The requested funding has increased with the announced available money for funding. At the same time, the size of the funded projects has decreased. Also the request for funding of big companies has increased from 37% to 76%, but their success-rate is lower than for the smaller companies. The sum of the funding asked for was around 19 Mio €. In total around 9 Mio € have been announced for the projects and 6.4 Mio € were decided for funding.

Keywords: computational mathematics, funding, Austria, economy

## 1. INTRODUCTION

In Austria, the quality of scientific research in mathematics is high (FWF 2007). On the other hand, the use of applied mathematics in companies was low (Ohler and Tiefenthaler 2007). Therefore the federal ministry for transport, innovation and technology (bmvit) decided to stimulate the challenging use of applied mathematics for commercial use. The funding initiative called “ModSim Computational Mathematics” (ModSim) was established by the bmvit. It provided 6 Mio € funding money for research projects. ModSim is owned by the bmvit and run by the Austrian Research Promotion Agency (FFG). The FFG performs awareness measures, contacts researchers and companies working in the field of or possibly benefit from ModSim. It also organises the evaluation process and monitors the funded projects. The funded research projects are between basic research and product development.

## 2. THE FUNDING INITIATIVE MODSIM COMPUTATIONAL MATHEMATICS

The goals of ModSim are:

- Intensify the challenging use of computational mathematics in the Austrian business and research setup
- Development of structures for research and development with the purpose of long-term transfer of knowledge between science and economy in the area of Computational Mathematics

These goals should be reached by funding research projects. The projects are preferably run by consortia consisting of researchers from research institutions and companies. But also projects of individual organisations are possible.

The European legal framework for the funding is the European-Union law “Community Framework for State Aid for Research and Development and Innovation” (Framework, 2006). Below that also Austrian law and the funding initiative FIT-IT (Research, Innovation and Technology for ICT) define the details of the funding scheme.

The projects fund research of the type ‘industrial research’ as defined in the Community Framework. The outcome of a ModSim-project must not exceed a research prototype. Such a research prototype only shows the functionality, but is not yet suitable for mass production. So the projects deal with issues which might come to the market within 3-8 years. The runtime of a project has a maximum limit of 3 years and a minimum time of 6 month.

The FFG organizes the evaluation process, signs funding-contracts and deals with the financial issues of the funded projects. Since 2008 three calls for proposals for research projects have been made.

### 2.1. Types of funded projects

There were two types of projects:

- Cooperative RTD: Research project in cooperation. It is fund between 54% and 76% of the total costs. The funding rate depends on the degree of contribution of the research

partner and the size of the involved companies. For Cooperative RTD – projects, the unfunded part of the research partners have to be paid by the companies. The funding scheme for Cooperative RTD was taken from the Austrian ICT-funding-initiative FIT-IT.

- Stimulation projects: may be run by a research group or by a company individually or in cooperation with partners. A stimulation project is fund with 50% of the total project costs. Financing of the unfunded costs is up to the consortium without rules from ModSim.

Stimulation projects provide more freedom with respect to the number of partners and the way of financing the unfunded part. Nevertheless, the funding rate for Stimulation-projects is smaller than for Cooperative-RTD-projects.

In the first two calls, also projects for the development of Human Resources were available. They were not used so often and have not been successfully at all. Therefore they were not available in the third call.

All proposals were evaluated by an international jury. The jury process is described in Kersch and Pogany (2009).

### 2.1.1. Applications and funding in ModSim

Table 1: Number of Applications in ModSim

	Coop. RTD	Stim.	Dvlp of HR	sum
2008 (1 <sup>st</sup> Call)	8	4	1	13
2009 (2 <sup>nd</sup> Call)	8	8	2	18
2010 (3 <sup>rd</sup> Call)	9	10	-	19
sum	25	22	3	50

It can be seen that during time, the number of applications has increased. The number of Stimulation proposals has more than doubled between the first and the 3<sup>rd</sup> call. In total the Cooperation RTD - project is the most wanted type of projects. On the other hand, Stimulation projects have gained more interest each year.

Table 2: Financial aspects of the applications and funded projects

	Available Money [Mio €]	Re- quested funding [Mio €]	Average costs of funded projects [1000 €]	Nr of partners of each Coop. RTD
2008	3	5.8	643	3.3
2009	4	7	576	3.6
2010	2	6.6	391	2.8
sum	9	19.4	-	-
average	3	6.5	537	3.2

Contrary as stated before, 9 Mio € were announced for funding instead of 6 Mio €. Both sums are true as not all available money was used for funding in the first

two calls. In Table 2 it is clearly visible that the amount of requested funding and the amount of partners in Cooperative RTD-projects have a similar relation to the available money. The relatively high amount of requested money in the 3<sup>rd</sup> call might be due to the higher number of Stimulation projects and due to the fact, that the initiative ModSim was better known in the Austrian research community.

Considering the number of proposals and the demand for funding, it can be said, that despite the increasing number of applications of Stimulation projects, the total sum of requested funding did not increase. This shows, that the project volume decreased during time.

Table 3: Requested Funding of different organisations [Mio €]

	Enter- prises	Research organisations	Uni- versities	Other
2008	2.5	1.5	1.8	0.01
2009	3.2	2.3	1.3	0.02
2010	3.1	1.4	1.7	0.03
sum	8.8	5.2	4.8	0.6

Table 3 shows clearly the high demand for funding coming from enterprises and research organisations. This shows that the funding scheme attracts the contribution of enterprises.

Looking at the contribution of small and medium-sized enterprises (SME) compared to all companies involved, table 4 shows their participation.

Table 4: Part of SME of total involvement of companies [Mio €]

	Requested funding by SME	Requested funding by big companies	Funding for SME	Funding for big companies
2008	1.6	0.9	0.8	0.2
2009	1.3	1.8	0.3	0.2
2010	1	2.1	0.3	0.6
average	1.3	1.6	0.5	0.3

Although the requested money of SMEs in the proposals has decreased during the calls, their acceptance rate with respect to money is always higher than for big sized companies.

The higher contribution of big-sized companies in the calls with time might be due to the long decision ways (and times) in big-sized companies (see table 4). Also successful SME from the first call(s) might not have submitted in the 2<sup>nd</sup> or 3<sup>rd</sup> call.

Table 5: Funding of different organisations [Mio €]

	Enter- prises	Research organisations	Uni- versities	Others
2008	1.1	0.6	0.5	0.09
2009	0.5	0.6	0.3	0.1
2010	0.9	0.5	0.5	0
sum	2.5	1.8	1.3	0.1

In 2009 the enterprises requested most of the funding, but received the smallest amount compared to the other years (compare tables 4 and 5). This might be due to the high competition in the year 2009. Also in 2009 the non-academic research organisations had the highest demand, but received more funding than the enterprises.

Looking at all years and the funding money with respect to the demand, the funding success of the companies was 28%, of the non-academic research organisations 35%, of the universities 27% and of the others 17%. So the monetary most successful type of organisations were the non-academic research organisations, followed by companies and universities.

In addition to the total public funding of 5.7 Mio € another 4.4 Mio € have been invested mainly by companies but also by non-academic research institutes and universities in the field of research in computational mathematics.

### 3. RESULTS FROM THE FUNDED PROJECTS

#### 3.1. Topics of the funded projects

The funded projects work in the fields of (alphabetic order):

- Automotive
- Building industry: 2 times
- Economy
- Energy production
- Hydraulic Equipment
- Logistic
- Logistic for public health
- Manufacturing technology: 3 times
- Material science
- Mechatronics
- Medicine
- Meteorology
- Optics
- Surface technology
- Timberwork
- Water supply

It is clearly visible that many different topics are addressed in the funded projects. Most successful were topics of manufacturing technologies (3 times). The kind of computational mathematics is as versatile as the topics: it ranges from numerical simulation, computational fluid dynamics, ab-initio calculations up to machine learning.

#### 3.2. Institutions running the funded projects

The motivation of companies acting as project leaders are twofold:

1. Use of the achievements in modelling and simulation in house or
2. Use of the methods as a service for other companies as users.

From the funded projects led by companies, 6 from 8 projects are designated to be used for in-house use. This shows a high demand for innovation. 50% of the leading companies using modelling in-house are big companies and 50% are SMEs. The two companies, which deliver the results of the projects as part of their own product/service are both SME.

For Stimulation projects, universities were most successful, whereas for Cooperative RTD - projects, the most successful leader were the research institutions. If a company was involved as a project partner, it was very successful with a success rate of 48%. This shows, that the cooperation of a company was very welcome in ModSim.

#### 3.3. Outcome of the funded projects

Up to now the first two projects have successfully finished their work. Within the finished projects several papers have been written, further collaboration could be established and the visibility of Computational Mathematics in Austria was increased by the work. As the finished projects cover just 4% of the total funding money, it is too early to conclude from the results from these projects to the effect of the funding.

### 4. COMPARISON WITH OTHER FUNDING INITIATIVES

In Finland there was a similar initiative between 2005 and 2009: The MASI Programme run by Tekes, the Finnish Funding Agency for Technology and Innovation. MASI aimed at:

- The widespread implementation in industry of modelling and simulation
- Innovation of modelling and simulation processes
- The creation of new commercial opportunities

The MASI programme had a total budget of 92 Mio €. (Holviala, 2009)

In comparison to ModSim, MASI was 15 times bigger and ran about twice as long. MASI also organised seminars, workshops, manager days and fair presentations. Whereas in ModSim, the public events were restricted to the publication of funding opportunities.

Another funding programme is located in Baden-Württemberg in Germany, called “High Performance Computing”, which concentrates on the topics life science, energy & environment and automotive. The runtime of the 7 funded projects is from 2009 until 2012 and 4 Mio € are available (HPC, 2011).

## 5. CONCLUSION AND OUTLOOK FOR MODSIM

### 5.1. Conclusion

Looking at the interim results of the running projects, it can be seen that new impact is given to the use of Computational Mathematics in some Austrian companies. Especially small and medium-sized companies have won capabilities, which were not so easy to access without funding. For some academic institutions, new fields of applications were found.

SMEs have been attracted by ModSim initially. To keep the contribution of SMEs high during time, additional special effort has to be made.

In the year 2009 the requested funding by enterprises and research organisation has increased, but their success rate was lower than in the year before and after.

### 5.2. Outlook

ModSim always was intended as an initiative with a definite end, the 3<sup>rd</sup> call was the last one. The already running projects run until their planned end.

The topic of simulation for industrial needs will now be taken up by a bigger initiative "Intelligente Produktion" (smart production: internet: <http://www.ffg.at/intelligente-produktion>, in german). This initiative aims at keeping the production site Austria competitive. Within this context, modelling and simulation is a proper tool to enhance the development time for example. Concerning funding, the already known companies and research institutions of ModSim will be able also to submit their proposals in the field of smart production.

### ACKNOWLEDGMENTS

The work described in this paper is financed by the bmvit via the funding scheme "ModSim Computational Mathematics".

### REFERENCES

- Framework, 2006: Official Journal of the European Union, 30.12.2006; 2006/C 323/01, "Community Framework for State Aid for Research and Development and Innovation"; <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:323:0001:0026:en:PDF>; 2011/06/17
- FWF 2007: FWF - Der Wettbewerb der Nationen – oder wie weit die österreichische Forschung von der Weltspitze entfernt ist. Eine Analyse der internationalen Wettbewerbsfähigkeit wissenschaftlicher Forschung Österreichs in den Natur- und Sozialwissenschaften, Wien, 2007 (in german)
- Holviala, 2009: „MASI Programme 2005-2009“ in „Tekes Programme Report 3/2010 – Final Report“

edited by Niina Holviala, VTT Technical Research Centre of Finland, Libris Oy, Helsinki 2010, ISBN 978-952-457-498-3

HPC, 2011: <http://www.bwstiftung.de/forschung/laufende-programme-und-projekte/information-kommunikation/high-performance-computing.html>; 2011/06/17, (in german)

Kersch P, Pogany A 2009: Public Funding of Modelling and Simulation in Austria, in: International Workshop on Modelling & Applied Simulation (MAS) at the conference "International Mediterranean and Latin American Modeling Multiconference" (I3M), September 23-25 2009, Puerto de la Cruz, Tenerife, Spain, Editors: O. Balci, I. Castilla, F. Longo, M. Massei; p 15-17

Ohler F, Tiefenthaler B 2007: DI Fritz Ohler, DI Brigitte Tiefenthaler, Technopolis Forschungs- und Beratungs- GmbH, 2007, „Modellierung und Simulation – Analyse und Förderung eines emergierenden Forschungsgebietes“ (in german)

### AUTHORS BIOGRAPHY

**Peter Kersch** is working at the Thematic Programmes of the Austrian Research Promotion Agency (FFG) since 2006. He is involved in the funding schemes for ICT, Computational Mathematics and smart production. Before joining the FFG he worked at the Institute for Solid State Physics and Materials Science Dresden (IFW Dresden), where he finished his PhD-thesis in the area of experimental solid state physics.

**Alexander Pogány** is working as policy expert for the Federal Ministry for Transport, Innovation and Technology within the area of nano- and industrial technologies. Before that he worked as validation expert in quality control for Baxter Bioscience. He holds a Master in Microbiology.

# DYNAMIC PHENOMENA AND QUALITY DEFECTS IN LASER CUTTING

Dieter Schuöcker<sup>(a)</sup>, Joachim Aichinger<sup>(b)</sup>, Georg Schuöcker<sup>(c)</sup>

<sup>(a)</sup>Upper Austrian Laser Center

<sup>(b)</sup>Upper Austrian Laser Center

<sup>(c)</sup>Svoent Entwicklungs GmbH & CO KG

<sup>(a)</sup>[dschuoecker@ooelz.at](mailto:dschuoecker@ooelz.at), <sup>(b)</sup>[jaichinger@ooelz.at](mailto:jaichinger@ooelz.at), <sup>(c)</sup>[georg.schuoecker@svoboda.at](mailto:georg.schuoecker@svoboda.at)

## ABSTRACT

If relatively thick workpieces are cut with a laser severe quality defects appear as enhanced roughness and adherent slag. The latter quality degradations must be caused by dynamic phenomena taking place in the liquid layer formed at the momentary end of the cut kerf. Therefore a theoretical analysis of the latter zone has been carried out that shows that surface tension plays a major role and causes an intermittent ejection of melt that is finally responsible for fluctuations of the volume of the molten body leading to roughness of the cuts. The latter intermittent melt removal can also be shown to cause adherent dross and slag. A mathematical evaluation of the latter considerations shows good coincidence with experiments.

Keywords: laser cutting, laser cut quality, modeling of laser cutting, dynamics of laser cutting

## 1. INTRODUCTION

Laser cutting is a most advanced technology for shaping sheet metals. Nevertheless in the case of thick workpieces certain quality defects can be observed as for instance cut edge roughness and irregularities and also adherent material as dross and slag (see Figure 1).

These effects can be caused by inappropriate parameters but also may have their origin in distortions going out from the various modules of the laser cutting system as the laser source, the motion system and the CNC control. In order to clarify the latter phenomena not only modeling of the cutting process itself, as usually done by many research groups, but also of the complete cutting system is ultimately necessary and has been tackled at the Upper Austrian Laser Center beginning with an improved model of the interaction process.

Laser cutting is governed to a large extent by the temperature distribution built up by a high power laser beam moving over the surface of the work piece. The latter temperature distribution results from the solution of the heat conduction equation. The solution of the latter is very complex since the thermal parameters involved depend on the temperature and so the differential equation is highly nonlinear. Nevertheless G. Herziger, one of the most prominent laser scientists

from the RWTH Aachen, has shown, that using average values for the relevant parameters as thermal conductivity, thermal diffusivity, specific heat and so on introduces only an error of a few percent into the solution of the heat conduction equation and therefore the temperature dependent relevant parameters can be replaced by average values. Therefore the heat conduction equation can be solved in an analytical way without the necessity of FEM-simulations, thus yielding analytical models for laser cutting. Models of that kind have been developed in the EU countries, in Russia, Japan and the USA.



Figure 1: Periodic Striations and Adherent Material at the Cut Edges, 8mm Steel

The distortions of workpiece quality must be related to dynamic phenomena appearing during laser cutting. So far the time dependent behavior of laser cutting has not been analyzed very often since most of the theoretical descriptions of laser cutting restrict themselves to steady state conditions. Nevertheless to be able to understand the dynamics of the process first the mechanism of continuous laser cutting, that means continuous cutting with constant high quality must be understood. Therefore in the first main part of the actual paper the mechanism of laser cutting and the geometry of the process will be explained. These considerations will show that at the momentary end of the cut kerf where the laser beam and a sharply focused jet of process gas hits the workpiece a thin liquid layer is



formed, that extends in vertical direction throughout the full depth of the workpiece. So far the phenomena appearing in this molten zone are not well understood and thus in the second main part of the paper the behavior of this region including dynamic phenomena will be analyzed in detail yielding an explanation for the formation of the well known striation pattern on the cut edges that leads to a certain roughness and also a model for the mechanism of the formation of adherent material at the cut edge. It will be shown that nice coincidence exists between theoretical outcomings and experimental results.

## 2. MECHANISM OF STEADY STATE LASER CUTTING

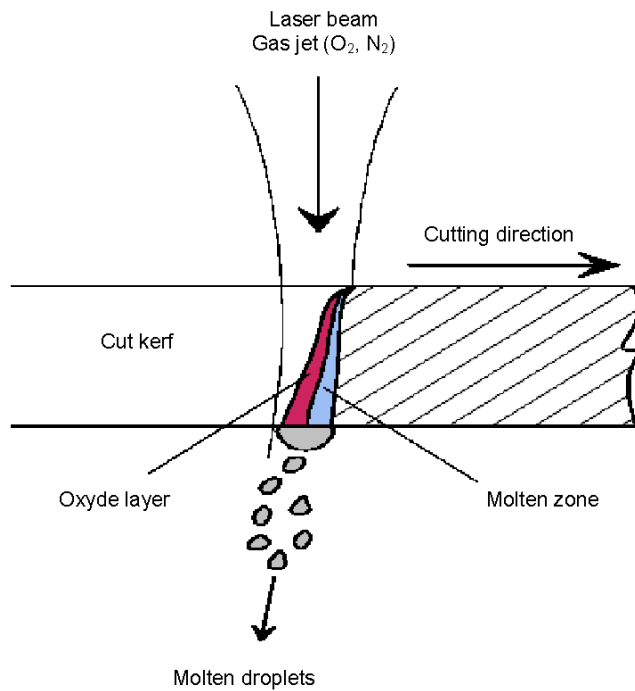


Figure 2: The Mechanism of Laser Cutting

Figure 2 shows a cross section of a work piece that has partly been cut whereas on the left hand side the kerf has been formed and on the right hand side the material is still uncut. At the momentary end of the cut kerf the laser beam and also a sharply focused jet of process gas preferably a mixture of oxygen and nitrogen usually deviating from the composition of air impinge on the workpiece. The laser beam is absorbed to some part by the momentary end of the cut and heats the material, finally leading to the formation of a thin molten layer that extends throughout the full depth of the workpiece. Moreover the oxygen content of the gas jet mentioned before, that hits the surface of the molten layer leads to the generation of additional heat by reaction. Due to the friction between the gas jet and the molten material, melt is ejected at the bottom of the workpiece thus leading to the material removal necessary for laser cutting. The liquid layer moves with

the laser beam and the gas jet in cutting direction and melts solid material at the interface between melt and solid and compensates thus the material losses due to ejection at the bottom of the workpiece as mentioned before. So all phenomena necessary for laser cutting take place in the molten zone as absorption of laser radiation, generation of reaction heat, heating and melting of the solid material and ejection of liquid material at the bottom of the workpiece and therefore the liquid layer can be regarded as the actual cutting tool. It moves through the workpiece and in the case of steady state cutting without any distortions of its volume, mass and temperature. A mathematical description of steady state laser cutting as described above must rely on balance equations as for the energy, for momentum and for mass since the internal structure and the mechanisms that take place in the liquid layer are unknown. In the next part of the actual paper it will be shown that there are time dependent phenomena associated to the liquid layer and playing an important role. With the description of these dynamic phenomena a possible explanation for the formation of a very rough surface structure and also of adherent material as dross and slag on the cut edges especially for the case of relatively thick workpieces can be found.

## 3. DYNAMIC BEHAVIOR OF THE MOLTEN LAYER

The molten layer that is located at the momentary end of the cut kerf as mentioned above is subject to friction with the cutting gas flow, that exerts a shear stress  $\tau$  on the liquid body. The latter depends on the viscosity of the cutting gas  $\eta$ , the density of the latter  $\rho$ , its speed  $v$  being equal to the speed of sound if a regular nozzle and not a laval nozzle is assumed, and the thickness of the workpiece  $d$  (Vicanek and Simon 1987):

$$\tau = \text{Sqrt}(\eta \rho v^3 / d) \quad (1)$$

To get the force acting on the liquid layer, the latter expression must be multiplied by the surface of the liquid layer given by the width of the kerf  $w$  and the thickness of the work piece. This force has a magnitude of  $2 \text{ N/m}^2$ . The latter force accelerates the molten material according to

$$\frac{d_{\text{melt}}}{dt} = \rho_m d w \frac{s}{2} \cdot \tau w d \quad (2)$$

where  $\rho_m$  is the density of the molten material. The use of  $s/2$  instead of  $s$  in the accelerated mass will be justified later.



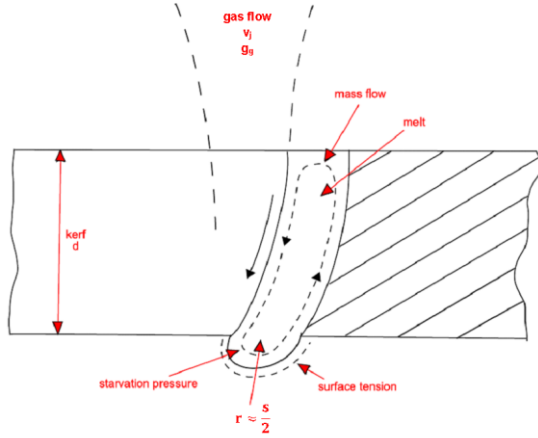


Figure 3: Cross Section of a Workpiece Partly Cut with a Laser, Left View Into the Kerf, Middle Molten Layer with Closed Loop Mass Flow, Right Part Still Uncut Material.

Due to the motion given by equation (2), the melt flows in a downward direction (Figure 3), but cannot leave the workpiece since the hydrostatic pressure in the melt is usually much smaller than the surface tension at the bottom of the liquid body. Therefore the melt flow must reverse there and flow in an upward direction, thus becoming a closed loop flow. The extension of both flows, downward and upward, is then half of the thickness of the liquid layer  $s/2$ . Since this mass flow is permanently accelerated by friction with the cutting gas, its speed  $v_{melt}$  rises and so does the starvation pressure  $p_{starv}$  built up at the bottom of the liquid body

$$p_{starv} = \rho_{melt} v_{melt}^2 / 2 \quad (3)$$

The latter pressure serves for the reversal of the melt flow direction.

The latter pressure adds to the hydrostatic pressure

$$p_{stat} = \rho_{melt} dg \quad (4)$$

The magnitude of the latter pressure is  $1000 \text{ N/m}^2$ , much smaller than the surface tension

$$p_{\sigma} = 2 \frac{\sigma}{r} \quad (5)$$

with  $r$  being the radius of curvature at the bottom of the liquid layer and approximately equal to  $s/2$ . An upper limit for the latter pressure is  $100.000 \text{ N/m}^2$ .

With permanently rising melt flow speed, eventually the starvation pressure will reach the surface tension, thus breaking up the surface skin and allowing the ejection of liquid material.

From the above equations, the time where this event takes place can be calculated:

$$T_{off} = \sqrt{\rho_{melt}} s \frac{1}{\tau} \sqrt{2 p_{\sigma} / s} \quad (6)$$

For steel with  $d=8 \text{ mm}$ ,  $v=2 \text{ m/s}$  and a laser power of  $3 \text{ kW}$ , equation 6 yields  $T_{off} = 0.0076 \text{ s}$ .

Due to the outflow of material in the opening in the bottom skin of the liquid layer the mass in the latter is reduced and thus also the pressure must decrease what means that after a short time the surface tension again prevails and the bottom skin closes and thus the outflow of material is interrupted, what means a new sequence as described here starts. The duration of the latter melt flow is given by the time needed by the melt to flow through the full vertical extension of the molten layer

$$T_{on} = \frac{d}{v_{melt}} \quad (7)$$

For the example treated before,  $T_{on} = 0.0028 \text{ s}$ .

With  $T_{off}$  and  $T_{on}$ , the frequency of the sequence of breaking and reestablishment of the surface tension is given by

$$f_{meltflow} = \frac{1}{T_{off} + T_{on}} \quad (8)$$

Since the mass and the volume of the liquid layer change during the above sequence it can be argued that also the width of the liquid layer changes and thus a pattern of periodic striations is generated on the cut edges. Equation (8) yields for the above example a striation frequency of  $96 \text{ Hz}$ , where experiments showed  $130 \text{ Hz}$ . Measurements for  $d=4 \text{ mm}$ ,  $v=3 \text{ m/min}$  yield a striation frequency of roughly  $380 \text{ Hz}$ , what confirms the theoretical tendency of the latter frequency to increase with decreasing workpiece thickness or the fact, that the roughness becomes smaller for thin workpieces.

It should be mentioned that various authors have developed models for the mechanism leading to periodic distortions on cut edges as for instance based on oscillations of the molten body (Schuöcker, 1986), waves on the melt surface (Chen and Yao, 1999), cyclic reaction (Ivarson, Powell, Kamalu, and Magnusson, 1994) and others. Fundamental work on the above topic has been carried out by Arata (Arata, Maruo, Miyamoto, and Takeuchi, 1979). All these models describe mechanisms that apply to specific situations e.g. Schuöcker treats cutting with and without oxygen content in the process gas whereas Ivarson and Ermolaev (Ermolaev, Kovalev, Orishich, and Fomin, 2006) apply only to reactive gas assisted laser cutting.

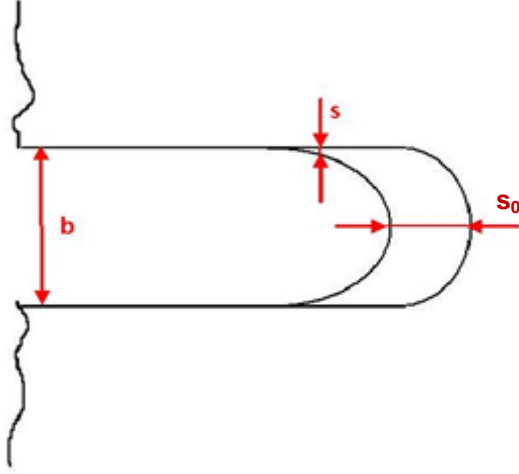


Figure 4: Cut Kerf with Molten Layer, seen from above

Following these considerations it can be argued that the condition for breaking up of the bottom surface skin at the molten layer near the kerf walls, the cut edges might be considerably different from the conditions in the case of the middle of the liquid layer as treated above, since at least the thickness of the liquid layer  $s$  is considerably smaller (see Figure 4). So it might happen that the bottom skin of the liquid layer breaks up earlier near the walls of the kerf than in the middle of the liquid layer (see Equation 6). In this case liquid material will also leave the molten body but in the case treated here it will flow along the walls of the kerf. In this case after the liquid material has left the molten body it may adhere at the kerf walls and this means the establishment of dross totally unwanted in laser cutting since it makes a post processing with big expenses necessary.

The latter considerations mean that the condition for breaking up the surface skin of the liquid layer takes place after a shorter time than for the central part. Nevertheless melt flow is in direct contact with the mass flow in the center of the liquid body and so the flow-speed will not change too much if one moves from the center of the liquid layer to the wall-near regions. With  $s_0$  being the thickness of the molten layer in the center of the melt, Eq. 6 writes then

$$T = \sqrt{\rho_{melt}} s_0 \frac{1}{\tau} \sqrt{2 p_{\sigma} / s} \quad (9)$$

So the breaking up of the surface tension is reached later near the walls with smaller  $s$ , that means usually the burst of liquid material out of the molten body will take place just in the center of the kerf. Nevertheless the liquid body is an unstable entity and may change its geometry especially in the center of the kerf in such a way that the point of equality between the starvation pressure and the surface tension is reached earlier in the wall-near zones of the liquid body and so the process will leave melt just at the walls and may then incline to form dross.

Coming back to the quasistationary performance of the melt body as mentioned in chapter 2, the average speed of melt ejection can now be determined from  $v_{melt}$ ,  $T_{open}$  and  $f_{meltflow}$ , allowing to calculate the mass loss per second of the molten body

$$\left( \frac{dm}{dt} \right)_{melt} = s w_k \rho_{melt} v_{melt} T_{on} f_{melt} \quad (10)$$

Again, the above example is used and yields for the above mass loss 0.00092 kg/s.

This can be compared to the mass molten per unit time and given by

$$\left( \frac{dm}{dt} \right)_{solid} = d w_k \rho_{melt} v \quad (11)$$

$v$  cutting speed. Again, for the above example, Equation (11) yields a mass gain of the liquid of 0.0008 kg/s, what agrees roughly to the mass loss of the liquid and justifies the assumption made for the thickness of the liquid layer.

The above considerations clearly show that the reasons for the intermittent mass ejection described here are a relatively weak friction between the process gas flow and the surface of the molten layer and a predominant surface tension causing a rough structure on the cut edge. Further on it has been shown that the main reason for adherent material as dross and slag is related to melt ejection not in the middle of the kerf but on the walls of the kerf. Based on these outcomes measures have been proposed in order to improve cutting quality, especially in view of a fine surface structure on the cut edge, leading to reduced roughness and the absence of dross and slag. The latter concept is subject to patent application (Schuöcker 2010).

## LIST OF REFERENCES

- Arata, Y, Maruo, H, Miyamoto, I, Takeuchi, S, 1979. Dynamic Behavior in Laser Gas Cutting of Mild Steel. *Transactions of JWRI* Vol.8, No.2: 15-26
- Chen, K, Yao, Y L, 1999. Striation Formation and Melt Removal in the Laser Cutting Process. *Journal of Manufacturing Processes* Vol.1/No.1: 43-53
- Ermaolaev, G V, Kovalev, O B, Orishick, A M, Fomin, V M, 2006. Mathematical modeling of striation formation in oxygen laser cutting of mild steel. *Journal of Physics D: Applied Physics* 39: 4236-4244
- Ivarson, A, Powell, J, Kamalu, J, Magnusson, C, 1994. The oxidation dynamics of laser cutting of mild steel and the generation of striations on the cut edge. *Journal of Materials Processing Technology* 40: 359-374
- Schuöcker, D, 1986. Dynamic Phenomena in Laser Cutting and Cut Quality. *Applied Physics B: Lasers and Optics* V40: 9-14

- Schuöcker, D, 2010. Laserschneidkopf: patent applied at Austrian patents office, 25.03.2011
- Vicanek, M, Simon, G, 1987. Momentum and heat transfer of an inert gas jet to the melt in laser cutting. *Journal of Physics D. Applied Physics* Vol. 20, No. 9: 1191-1196

## ACKNOWLEDGMENTS

The author is indebted to the Austrian research Promotion Agency (FFG) for funding the above work in the framework of the priority programme “Modeling and Simulating” especially to Dr. Peter Kersch for permanent encouraging. They also acknowledge the important contribution of experimental data by Svoent Company in St. Pölten / Austria and the technical assistance of Mrs. Carina Ebli.

## BIOGRAPHY

### Dieter Schuöcker

- |         |  |
|---------|--|
| 1941    | Born in Vienna   |
| 1968    | Graduation to Dr. techn. engineering   |
| 1979-81 | Voest-Alpine, Project leader for high power laser, development of laser cutting system “VALCUT”                                      |
| 1988    | Chairman of “7 <sup>th</sup> International Symposium on Gas Flow and Chemical Lasers (GCL)” in Vienna                                |
| 1993    | Head of the European-wide training program “EuroLaser Academy” in the framework of the COMETT and the LEONARDO initiative of the EU. |
| 1996    | Chair of nonconventional processing and laser technology of the Technical University of Vienna                                       |
| 2008    | Foundation of the Upper Austrian Lasercenter in Gmunden / Austria  |

### Joachim Aichinger

- |      |  |
|------|--|
| 1980 | Born in Linz                                   |
| 2010 | Dipl.-Ing. of Mechanical Engineering           |
| 2009 | Deputy head of the Upper Austrian Laser Center |

### Georg Schuöcker

- |      |                                      |
|------|--------------------------------------|
| 1976 | Born in Vienna                       |
| 2006 | Dr. techn. of mechanical engineering |
| 2010 | Svoent company, CEO                  |

# ANALYSIS OF A WAREHOUSE MANAGEMENT SYSTEM BY MEANS OF SIMULATION EXPERIMENTS

Aiello Giuseppe<sup>(a)</sup>, Enea Mario<sup>(b)</sup>, Muriana Cinzia<sup>(c)</sup>

<sup>(a)</sup> Dipartimento di Ingegneria Chimica, Gestionale, Informatica e Meccanica, University of Palermo,  
90128 Palermo – Italy

<sup>(b)</sup> Dipartimento di Ingegneria Chimica, Gestionale, Informatica e Meccanica, University of Palermo,  
90128 Palermo – Italy

<sup>(c)</sup> Dipartimento di Ingegneria Chimica, Gestionale, Informatica e Meccanica, University of Palermo,  
90128 Palermo – Italy

<sup>(a)</sup> [aiello@dtpm.unipa.it](mailto:aiello@dtpm.unipa.it), <sup>(b)</sup> [enea@unipa.it](mailto:enea@unipa.it), <sup>(c)</sup> [muriana@unipa.it](mailto:muriana@unipa.it)

## ABSTRACT

The supply chains for perishable products are nowadays affected by significant wastes and losses. Their management hence requires optimized approaches in order to remove such inefficiencies. In particular optimized warehouse management policies is a well established research topic which has been recently enriched with specific formulations for deteriorating stocks and shelf life based picking rules. In such context the proposed research aims at investigating the optimal warehouse management policy, taking into account the effects of uncertainty by means of simulation and approaching the effect of optimal picking rules. In the proposed approach, on the basis of a Weibull deterioration process, the optimal order quantity is calculated taking into account the deterioration cost, and the performance of the system is analyzed taking into account the inherent variability in the quality of the products entering the cold chain. The numerical application developed confirms the effectiveness of the model proposed.

Keywords: supply chain, perishable inventory, simulation

## 1. INTRODUCTION

Perishable inventory systems are characterized by a reduction of the value of the products with time, which ultimately results in the discard of the products stored upon the loss of the minimal quality level required by the consumer. Due to the necessity of preserving the quality attributes of the products until they reach the final market, the management of the supply chain for perishable products is a challenging task. In particular the supply chain lead time is an extremely critical parameter influenced by the operational policies and logistics variables enforced throughout the entire chain. In order to maximize the performance of the supply chain such policies should be established taking into account exogenous parameters such as the deterioration rate and the demand rate of the products. Supply chain

operations should hence be timely scheduled and properly managed in order to reach the best compromise between the cost of handling/transporting operations and the quality of the product delivered. In the practical context, significant losses of products are likely to occur due to inadequate management of the post harvest activities including warehousing, handling and transport operations. Such losses not only represent an industrial cost, but they also constitute an ethical and environmental concern which severely influences the sustainability of agro-industrial supply chains. Recent studies (Broekmeulen and Van Donselaar 2007), report that perishables stand for almost one third of the sales of the supermarket industry and approximately 15% of the perishables are lost due to spoilage.

Deterioration processes such as spoilage are the result of biochemical and biological phenomena such as respiration, lipid oxidation, microbiological proliferation, which ultimately determine the shelf life of the product. These phenomena are directly related to temperature, which in fact is the most significant environmental factor that influences the deterioration rate of harvested products. The relationship between product quality in terms of shelf life (SL) and temperature is studied extensively, e.g. by Doyle (1995) and Taoukis (1999). According to these studies, by knowing the time/temperature history of a product through the supply chain it is theoretically possible to predict its remaining shelf life (RSL) at any stage of the supply chain. The knowledge of the deterioration rate of a product is an information which should be considered when establishing an optimal inventory management system, as it influences for example the operative decisions about the replenishment policy, the optimal order quantity and the picking policy. Traditional warehouse management systems that are not based on such information are typically affected by the uncontrolled deterioration of products stored with time.

Optimal warehousing policies for perishable products is a well established research field which has lead to significant results concerning the efficiency of

logistic throughout the supply chain. Ketzenberg and Bloemhof (2009) report that the knowledge of the deterioration state in real-time allows dynamic decision making through the supply chain, so that products that are soon to expire can be distributed locally while less deteriorated products can be distributed to more distant locations. Another important result achievable by the evaluation of the RSL is related to possibility of implementing SL based logistics policies as LSFO (Least Shelf life First Out) rather than traditional FIFO (Firs In First Out) and LIFO (Least In First Out) rules thus improving supply chain performance, by reducing the fraction of deteriorated products wasted before sale. Finally it must be pointed out that the performance of the supply chain is influenced by some uncertain and uncontrollable factors such as the inherent variability of the deterioration state of the products entering the cold chain. This is due to the harvest operations, which may take several hours and are generally performed outdoor where environmental conditions cannot be modified. Even when the temperature throughout the supply chain is perfectly controlled, the variability in the level of initial maturation of the products results in the presence of a fraction of perished products. In this paper such uncertainties, are taken into account by means of a simulation model. A Simulation method is in fact generally preferred when the existence of uncertainties makes the task of mathematical programming highly complicated. Simulation is the process of constructing a model of a real system and conducting experiments with such a model, with the purpose of analyzing the behavior of the system and evaluating different strategies of operation. Simulation-based optimization is an active research area in the field of the stochastic optimization. Reviews of the research on simulation-based optimization developments can be found in Andradottir (1998) and Fu and Hu (1997). The effects of uncertainties and the large number of control variables that may be present in a supply chain, is the main reason why simulation is widely employed in several supply chain analysis. For example, Sezen and Kitapçı (2007), developed a spreadsheet simulation model for a single distribution channel and simulated three different scenarios reflecting various levels of demand fluctuations. Similarly, Banerjee et al. (2003) developed a supply chain simulation model with the aim to compare two different trans-shipment approaches. Zhao et al. (2002) investigated the complex relationships between forecast errors and early order commitments by simulating a simple supply chain system under uncertain demand. Finally Young et al (2004) addressed the problem of determining the safety stock level to use in the supply chain in order to meet a desired level of customer satisfaction using a simulation based optimization approach.

The purpose of the present paper is to analyze the performance of a cold chain considering the effects of uncertainties and exogenous parameters such as the demand rate and the kinetics of the deterioration phenomena. The approach proposed allows to

determine the best compromise between product availability and supply chain responsiveness by establishing the optimal set of supply chain operational parameters (stock levels) corresponding to the deterioration rate of the product and to the characteristics of the demand, taking into account holding/waste costs and product quality.

Finally a sensitivity analysis is conducted through an experimental plan in which the exogenous parameters as well as the RSL are varied in the two levels. A set of simulation runs is defined in which the warehouse replenishment cycle is modeled based on the optimal stock level in order to determine the operative variables.

## 2. SHELF LIFE MODELING AND DETERIORATION PROCESS

*Shelf life is defined as the time until a perishable product becomes unacceptable to consumers under a given storage condition* (Singh and Cadwallader 2004). The shelf life of a product can be measured directly by means of Accelerated laboratory tests, and subsequently evaluated by means of mathematical models. Mathematical Shelf Life models typically involve some input parameters that characterize the environmental conditions (as temperature, relative humidity, etc.) to determine the consequent decrease in the quality attributes. The modeling method based on the kinetic of reactions states that the quality decay of the product can be expressed through the Arrhenius law, by relating the reaction rate to the temperature based on the Activation Energy. In this case the following expression can be employed:

$$K = K_0 * \exp\left(-\frac{E_a}{RT}\right) \quad (1)$$

where  $k_0$  is the pre-exponential factor,  $E_a$  is the Activation Energy (temperature sensitivity) of the reaction that controls food quality loss,  $R$  is the universal gas constant and  $T$  is the absolute temperature (K), which may be constant or variable. Based on equation (1) the SL of a product exposed at variable temperature can be calculated on the basis of the following expression:

$$\ln SL = \ln SL_0 + \frac{E_a}{RT} \quad (2)$$

From (2) it is possible to determine the fraction of the SL consumed when a product is exposed to a constant temperature for a certain time interval (Giannakourou 2003). The formulation is:

$$f_c = \sum \frac{t_i}{SL_i} \quad (3)$$

where  $f_c$  represents the sum of the times at each constant temperature segment,  $t_i$ , divided by the  $SL_i$  at that temperature.

In order to preserve the quality of the product, all the logistics operations involved are carried out at pre-

established temperature, and a temperature control system is enforced throughout the supply chain which is thus properly defined “cold chain”. There are however some operations which cannot be included in the cold chain, like for example the harvesting operations which must be carried out outdoor. The deterioration phenomena which take place during these operations are therefore responsible of a different deterioration level of the products entering the cold-chain even when the harvesting phase is scheduled at the same maturation level.

When the products enter the cold chain they are kept at proper temperature which is assumed to be perfectly controlled, the deterioration is thus function of time only. Several authors have faced the problem of modeling the dependence of the deterioration rate from time. The most common deterioration models are linear, quadratic and Weibull deterioration models. The Weibull deterioration model is expressed as:

$$\theta(t) = \alpha \beta t^{\beta-1} \quad (4)$$

where  $\alpha$  is the scale parameter,  $\alpha > 0$ ;  $\beta$  is the shape parameter  $\beta > 0$ ;  $t$  is time of deterioration,  $t > 0$ . It is seen from equation (4) that the two-parameter Weibull distribution is appropriate for an item with decreasing rate of deterioration ( $\beta < 1$ ) only if the initial rate of deterioration is extremely high. Similarly, this distribution can also be used for an item with increasing rate of deterioration ( $1 < \beta < 2$ ) only if the initial rate is approximately zero. In the present paper the deterioration rate of products stored is determined by considering a Weibull distribution and RSL of products leaving the warehouse is calculated with the following equation:

$$RSL_{exit} = RSL * (1 - f_c) \quad (5)$$

The RSL at each stage of the supply chain is defined as the number of days a product is still available for consumption starting from the moment the product arrives at that stage of the supply chain. Van Donselaar and al. (2006), report that the RSL is a function of the SL, the distribution strategy (including e.g. decisions on direct delivery, cross-docking or delivery from stock at the retailer's distribution channel and the shipping frequency) and the inventory replenishment logic (e.g. push or pull).

### 3. STOCK ROTATION SYSTEM AND EOQ FOR DETERIORATING ITEMS

A warehouse management system is characterized by the stock management policies enforced and the operative parameters such as the average stock level, the duration of the replenishment cycle, etc.

The FIFO policy is the policy generally employed for managing perishable products and it is based on the assumption that the obsolescence of the products is related to their time of arrival. In this case the products that arrive first have the smaller RSL and then they

must sell first. However in agroindustrial supply chains, due to the variability of the quality of harvested products, not always their obsolescence is perfectly correlated to the time of arrival. In this case a quality attribute such as the RSL, should be more effectively employed to decide which product must be picked first. This would ensure that products that have the smaller quality attribute leave the warehouse first. Therefore, if the information about the RSL is available it would be possible to release the products with shortest RSL first, thus enforcing a LSFO picking rule. Taoukis and Giannakourou (1998), demonstrate that compared with FIFO policy, the LSFO would reduce rejected products and eliminate consumer dissatisfaction since the fraction of product with unacceptable quality consumed can be minimized.

Another important decision that must be taken when establishing an inventory management policy concerns the replenishment policy. Traditional well-known optimizing strategies such as the economic order quantity (EOQ) cannot be directly applied to perishable inventories unless the costs of deterioration is additionally considered. The traditional EOQ model aims at optimizing the holding and ordering costs neglecting the influence of the deterioration costs as well as the salvage value of products perished. For this reason the extension of the EOQ model to perishable inventories is a topic intensively treated in recent years. Tarun Jeet Singh and al. (2009), have built an EOQ model when the deterioration rate has a linear trend in the two cases in which shortage are allowed or not. Shibsankar Sana and Chaudhuri (2004), refer to an EOQ model including a quadratic deterioration rate. Ghosh and al. (2005), Begum and al. (2010), have built a model including deterioration costs in the case in which the deterioration rate follows a Weibull distribution. Manna and al. (2006), Nita and al. (2008), have determined an EOQ model by considering the salvage value of products and a deterioration cost when deterioration rate follows a Weibull distribution and delay in payments are permissible.

In the present paper the EOQ model is proposed derived by the model proposed by Nita et al (2008). The model is developed using the following notations:

- $C$  is the purchase cost per unit;
- $\gamma C$  is the salvage value, associated to deteriorated units during the cycle time, where ( $0 \leq \gamma < 1$ );
- $h$  is the inventory cost per unit per time;
- $A$  is the ordering cost per order;
- $T$  is the cycle time (a decision variable).

The following assumption are used:

- the demand rate of  $R$  units per time is assumed to be deterministic and constant;
- the system deals with a single item;
- the replenishment rate is infinite;

- the lead time is zero and shortages are not allowed;
- the deterioration rate of units follows the Weibull distribution function given by (4), where  $0 \leq \alpha < 1$ ,  $\beta \geq 1$ ,  $0 \leq t \leq T$ .
- the deteriorated units can neither be repaired nor be replaced during the cycle time.

Based on assumptions made and supposing that the decrease of inventory is only due to the demand rate and to the deterioration of products, the inventory level  $Q(t)$  is governed by the differential equation:

$$\frac{dQ(t)}{dt} + \theta(t) * Q(t) - R, \quad 0 \leq t \leq T \quad (6)$$

with the initial condition  $Q(0) = Q_{opt} = EOQ$  and the boundary condition  $Q(T) = 0$ .

Taking series expansion and ignoring second and higher power of  $\alpha$  (assuming  $\alpha$  to be very small), the solution of the differential equation (6) using the boundary condition  $Q(T) = 0$  is given by:

$$Q(t) = R \left[ T - t + \frac{\alpha T}{\beta+1} (T^\beta - (1+\beta)t^\beta) + \frac{\alpha\beta t^{\beta+1}}{\beta+1} \right] \quad (7)$$

that expresses the inventory level at each generic instant  $t$ . The number of units that deteriorate during a cycle can be calculated as:

$$D = Q - RT = \frac{\alpha RT^{\beta+1}}{\beta+1} \quad (8)$$

The cost of deterioration (CD) is:

$$CD = \frac{\alpha CRT^{\beta+1}}{\beta+1} \quad (9)$$

The salvage value (SV) is:

$$SV = \frac{\alpha \gamma CRT^{\beta+1}}{\beta+1} \quad (10)$$

The per cycle inventory holding cost (IHC) is:

$$IHC = h * \int_0^T Q(t)dt = hR \left[ \frac{T^2}{2} + \frac{\alpha\beta T^{\beta+2}}{(\beta+1)(\beta+2)} \right] \quad (11)$$

$$\text{The ordering cost is: } A \quad (12)$$

The total cost per cycle is:

$$TC(T) = CD - SV + IHC + A \quad (13)$$

and the Total Cost per Time Unit is:

$$TC_u(T) = \frac{[CD - SV + IHC + A]}{T} \quad (14)$$

By deriving (14) with respect to  $T$  and solving it the optimal cycle time is determined and by equation (7) the corresponding optimal order quantity is calculated.

#### 4. DESIGN OF SIMULATION EXPERIMENTS

In this paragraph the effect of the uncertainties on the warehouse management systems is addressed. According to Gong (2009) the uncertainty faced by warehouse systems can be classified in: sources outside the supply chain, sources in the supply chain but outside the warehouse, sources inside the warehouse, and sources within warehouse control systems. According to the variance structure of uncertainties, we classify uncertainty sources as unpredictable events like strikes, floods, and hurricanes, which usually are rare events, predictable events like demand seasonality, and internal variability like variance of order waiting time for batching. External uncertainty sources usually are more unpredictable, and will often bring higher variance to warehouse operations. On the other hand, inside uncertainty sources usually are more predictable and only bring low variance to warehouse operations.

In the present study the attention is focused on sources outside the warehouse but that affect warehouse management. They include predictable events like demand fluctuations and variability of RSL of products entering the warehouse.

In traditional warehouse management systems, where the picking policies used are based on arrival time of SKUs (FIFO) and no information management system is adopted, such uncertainties have a strong impact in terms of quality of products leaving the warehouse.

The purpose of this paper is to study the effect of these uncertainties by evaluating the RSL distribution of the products leaving the warehouse when the two policies LSFO and FIFO are applied. The study of warehouse system is carried out with the methodology of design of simulation experiments.

The simulation model is used to take into account the effect of the fluctuation in demand and variability of RSL of products entering the warehouse on the performance. Finally a sensitivity analysis is carried out consisting on a three factors experimental plan in which the RSL of products entering the warehouse, the demand rate and the deterioration rate vary on two levels. The response of the experimental plan, consisting in the average RSL of products leaving the warehouse has been determined. The results obtained show how the information about the RSL of the product can improve the operational and tactical management decisions thus increasing the quality of the products delivered.

An experimental plan is generally designed to estimate how changes in the input factors affect the results, or responses, of the experiment. While these methods were developed with physical experiments in mind (like agricultural or industrial applications), they can fairly easily be used in computer-simulation experiments as well, as described in more detail in Law and Kelton (2000). In fact, using them in simulation presents several opportunities for improvement that are difficult or impossible to use in physical experiments. In such situation the most suitable tool to know the system

behavior is the computer based-simulation experiment. The design of simulation experiments starts by building the experimental plan including all parameters that affect the system behavior, thus the model simulation is realized to represent the actual system, and finally each configuration of the experimental plan is executed through simulation model and average measures of interest are calculated.

An important result achievable from an experimental plan is related to the possibility of estimating the main effect of each factor in the plan, defined as the average difference in response when this factor moves from its low level to its high level. Furthermore the interaction between the factors if it seems to be present can be determined to know if the effect of one factor might depend in some way on the level of one or more other factors.

As mentioned above when an experimental plan must be executed the simulation can result very helpful to substitute the physical experiment thus resulting in a very inexpensive tool. Any computer based simulator aims to mirror the behavior of the real system that it represents. To achieve this goal, the different decision processes along the operational cycle, the replenishment cycle and the picking policy of a warehouse in this case, must be accurately reproduced. Keeping this in mind, the simulation model presented here is built through a discrete event simulator. Discrete event simulation concerns the modeling of a system as it evolves over time by a representation in which system variable changes instantaneously at separate points in time – the ones in which an event occurs.

A simulation model takes the form of a set of assumptions concerning the operations of the system. These assumptions are expressed in mathematical, logical, and symbolic relationships between the *entities*, or objects of interest, of the system. Some of these assumptions can comprise those situations in which one or more input in the model are random variable. In this case the outputs provided by the model can be considered only as an estimates of the true characteristic of the model. Once developed and validated a model can be used to investigate a wide variety of “what if” questions about the real-world system. Potential changes to the system, defined through the experimental plan, can be first simulated, in order to predict their impact on system performance.

Most experimental designs are based on an algebraic regression-model assumption about the way the input factors affect the outputs. It is assumed that the independent variables are continuous and controllable by experiments with negligible errors. Furthermore it is required to find a suitable approximation for the true functional relationship between independent variables and the response surface. If all variables are assumed to be measurable, the response surface can be expressed as follows:

$$y = f(x_1, x_2, \dots, x_k) \quad (15)$$

The goal is to optimize the response variable  $y$ . Usually a second-order model is utilized in response surface methodology as explained in Raissi (2009).

$$y = \beta_0 + \sum_{i=1}^k \beta_i g_i + \sum_{i=1}^k \beta_{ii} g_i^2 + \sum_{i=1}^k \sum_{j=1}^k \beta_{ij} g_i g_j + \varepsilon \quad (16)$$

where the  $\beta_j$  coefficients are unknown and must be estimated somehow, and  $\varepsilon$  is a random error term representing whatever inaccuracy such a model might have in approximating the actual simulation-model response  $y$ . The parameters of the model are estimated by making simulation runs at various input values according, for example, with the experimental plan, recording the corresponding responses, and then using standard least-squares regression to estimate the coefficients. In simulation, an estimated response-surface model can serve several different purposes. You use them as a proxy for the simulation, and very quickly explore many different input-factor-level combinations without having to run the simulation. And you could try to optimize (maximize or minimize, as appropriate) the fitted model to give you a sense of where the best input-factor-combinations might be. For more details see Kelton (2003).

When an experimental design is executed the quality control of the measures obtained is of fundamental importance to ensure a desired precision about average measures of variables of interest. This means that the average measures determined by each configuration of the experimental plan actually represent the estimate of the expected value (EV) for this measure. For this reason, depending on the precision desired for the measure under analysis, a certain number of replication of the model must be executed.

If the variables  $X_n$  of interest can be considered mutually independent and identically distributed (IID) with mean  $\mu$  and finite variance  $\sigma^2$ , we can use the simple mean  $\bar{X}_n$  to estimate the mean. Clearly, the classical case arises whenever we use independent replications to do estimation. In the classical case, the sample mean  $\bar{X}_n$  is a consistent estimator of the mean  $\mu$  by the law of large numbers (LLN). Then there is no bias and the MSE coincides with the variance of the sample mean,  $\overline{\sigma_n^2}$ , which is a simple function of the variance of a single observation  $X_n$ :

$$\overline{\sigma_n^2} = \text{MSE}(\bar{X}_n) = \frac{\sigma^2}{n} \quad (17)$$

Moreover, by the central limit theorem (CLT),  $X_n$  is asymptotically normally distributed as the sample size  $n$  increases, i.e.,

$$n^{1/2}[\bar{X}_n - \mu] \rightarrow N(0, \sigma^2) \text{ as } n \rightarrow \infty, \quad (18)$$

where  $N(a, b)$  is a normal random variable with mean  $a$  and variance  $b$ , and  $\rightarrow$  denotes convergence in



distribution. We thus use this large-sample theory to justify the approximation:

$$P(\bar{X}_n \leq x) \approx P\left(N\left(\mu, \frac{\sigma^2}{n}\right) \leq x\right) = P\left(N(0,1) \leq \frac{x-\mu}{\sqrt{\frac{\sigma^2}{n}}}\right) \quad (19)$$

Based on this normal approximation, a  $(1-\alpha)100\%$  confidence interval for  $\mu$  based on the sample mean  $\bar{X}_n$  is:

$$\left[\bar{X}_n - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right), \bar{X}_n + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right] \quad (20)$$

where

$$P\left(-z_{\frac{\alpha}{2}} \leq N(0,1) \leq +z_{\frac{\alpha}{2}}\right) = 1 - \alpha \quad (21)$$

and  $\alpha$  denotes the error (width of confidence interval divided by the estimated mean) admitted in the measure. The statistical precision is typically described by either the absolute width or the relative width of the confidence interval, denoted by  $w_a(\alpha)$  and  $w_r(\alpha)$ , respectively, which are:

$$w_a(\alpha) = \frac{2z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} \quad (22)$$

$$w_r(\alpha) = \frac{2z_{\frac{\alpha}{2}}\sigma}{\mu\sqrt{n}} \quad (23)$$

For specified absolute width or relative width of the confidence interval,  $\epsilon$ , and for specified level of precision  $\alpha$ , the required sample size  $n_a(\epsilon, \alpha)$  or  $n_r(\epsilon, \alpha)$  is then

$$n_a(\epsilon, \alpha) = \frac{4\sigma^2 z_{\frac{\alpha}{2}}^2}{\epsilon^2} \quad (24)$$

$$n_r(\epsilon, \alpha) = \frac{4\sigma^2 z_{\frac{\alpha}{2}}^2}{\mu^2 \epsilon^2} \quad (25)$$

For detailed discussion about statistic aspect refer to Whitt (2005). Generally to calculate how many replications are needed for the precision required at first a certain number of replication of each configuration are run, the confidence interval and initial level of precision are determined. Thus the number of replication is calculated through equation (24) by fixing the desired precision.

The goal of this study is to show that applying an EOQ model for perishable product (taking into account the deterioration rate of its) and moving from the FIFO policy to the LSFO policy it is possible to improving warehouse performance. To achieve this goal the optimal order quantity (EOQ) is determined depending on some uncertainty sources that are internal the supply chain but sometimes uncontrollable as demand rate and

deterioration rate. Thus a sensitivity analysis consisting in a three factorial experimental plan is presented in which each factor varies on two quantitative levels.

Experimental plan has been replicated to ensure a desired precision in the measures. To execute the experimental plan the simulation tool has been chosen as enabling the modeling of a real-life system. The simulation tool allows to model the system behavior when some stochastic input as RSL of products entering the warehouse are present, to determine RSL of products leaving the warehouse that represent the response of the experimental plan proposed.

## 5. PROPOSED METHODOLOGY

The study here presented focuses on the postharvest operations of perishable products that are transferred from the field to the refrigerated warehouse where they are before being shipped to the final retailer.

The first step carried out concerns the determination of the optimal warehouse management policy.

Initially the problem of determining the optimal batch size is considered, taking into account the deterioration cost of the product. It is in fact well known that traditional optimal policies such as the Economic Order Quantity (Wilson 1934) model which do not take into account deterioration costs, result in batch sizes which are generally unfeasible for perishable goods. According to the methodology here proposed the optimal order size is determined by equation (7) on the basis of the model reported in Section 3, and the  $TC_u$  is determined by equation (14).

Once the optimal batch size is determined, the proposed analysis focuses on the evaluation of the expected RSL of the products as they leave the warehouse after storage. This is carried out by considering an initial RSL value attributed to the products harvested as they enter the warehouse. Due to the inherent variability to the maturation level, the uncontrollable environmental conditions, and the variable duration of the harvesting operations, the RSL of the products entering the warehouse presents an intrinsic variability, it has therefore been modeled by means of a stochastic random variable. Such uncertainty has been taken into account in the analysis of the storage system by means of a simulation approach, considering the RSL of the products entering the warehouse, the demand rate, the deterioration rate as input parameters and evaluating the RSL of the products leaving the warehouse as the output parameter. Such output value is calculated in each run on the basis of the initial RSL and considering the deterioration of the products during the storage time, which is assumed to follow a Weibull function.

Thus the effect of the system parameters on the output variable has been explored by means of a simulation experiments plan specifically designed. The experimental plan has been replicated to ensure a desired precision of the measures obtained. Finally the effect of a shelf-life based picking rule has been

compared to the traditional FIFO rule generally employed for perishable goods and a sensitivity analysis has been performed.

## 6. EXPERIMENTAL APPLICATION

In this paragraph a numerical application is proposed, based on experimental data. In particular a warehouse is considered for the allocation of a perishable product characterized by a Weibull deterioration model having  $\alpha=0.2$ ,  $\beta=1.5$ . The optimal batch size is determined considering  $A=50\text{€}$ ,  $R=16$  SKUs/day and  $h=0.1\text{€/SKU/period}$ . The product cost ( $C$ ) has been considered equal to  $0.5\text{€/SKU}$  and the salvage value of the perished product is  $\gamma C=0.01\text{€/SKU}$ .

Products that remain unsold at the end of the generic warehouse cycle are considered perished and thus sold at their salvage value.

In such conditions the  $TC_u$  is of  $22.51\text{€}$  which corresponds to a cycle time of 3.7 days. The optimal order quantity has been determined by equation (7) and is equal to 93 SKUs. The EOQ and the related optimal  $TC_u$  for different values of  $\alpha$  is reported in Figure 1, and the corresponding Inventory level  $I(t)$  is reported in Figure 2. Such figures confirm that the optimal order size warehouse decreases when  $\alpha$  increases.

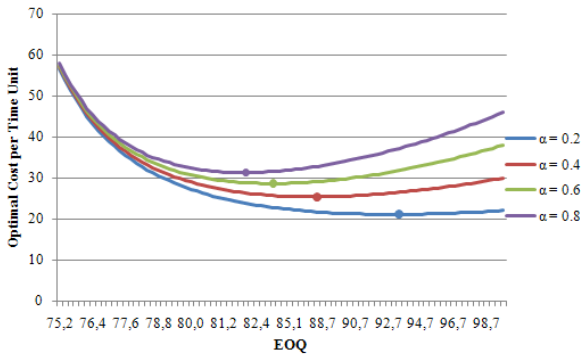


Figure 1: Optimal Order Quantity and Optimal Total Cost per Time Unit for several value of  $\alpha$

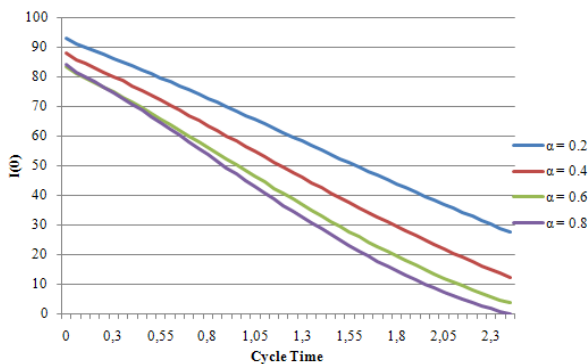


Figure 2: Trend of Inventory level for several value of  $\alpha$

For the same values of  $\alpha$  and  $\beta$  the  $TC_u$  is reported in Figure 3 which shows that the cycle time related to the minimum total cost per unit time decreases when the  $\alpha$  increases.

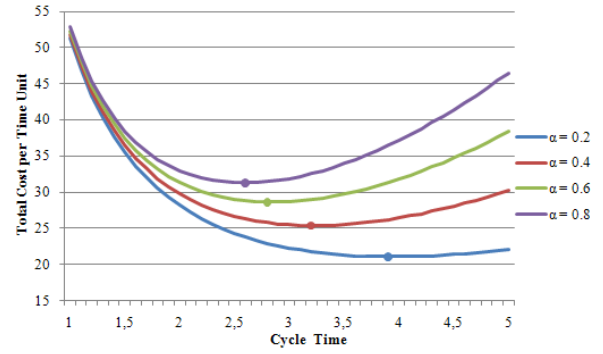


Figure 3: Cycle Time for different value of  $\alpha$

Once the optimal order size has been determined the effect of the inherent variability in the RSL of the products entering the warehouse has been addressed. In this study the RSL of the batches entering the warehouse has been assumed to be a stochastic random variable distributed according to a Normal density probability function with the following parameters:  $N(\mu_{SL}=24 \text{ days}, \sigma_{SL}=3.5)$ . A simulation model has hence been generated in order to determine the distribution function of the RSL values of the products leaving the warehouse after the storage period, taking into account the deterioration process.

The simulation model thus generated has been employed to determine the performance of the system and to analyze the effect of different storage policies and picking rules. In particular, in the warehouse system considered three fundamental factors have been selected which affect warehouse performance. Such parameters are RSL of the incoming products, the demand rate and the deterioration rate. Based on the assumptions made in section 4 a three factors experimental full plan has been generated in which each factor varies on two levels, thus resulting in  $2^3=8$  configurations, as shown in Table 1.

Table 1: Experimental Plan

Scenario	RSL (Days at 10°C)	$\theta$ (Unit/unit time)	R (Batch/Day)
1	20	0.3	16
2	24	0.3	16
3	20	0.6	16
4	24	0.6	16
5	20	0.3	24
6	24	0.3	24
7	20	0.6	24
8	24	0.6	24

This experimental plan reports the input parameters used in the simulation model to get the response in terms of the warehouse performance.

The influence of two different policies FIFO and LSFO has also been investigated. This aspect has been modeled by ranking the products according to their RSL, once they are stored in the warehouse, and by

using such rank in the picking list, when the LSFO policy is adopted. On the contrary, when the traditional FIFO policy is adopted, the products leave the warehouse based on their arrival order, products entering the warehouse the same day are therefore undistinguishable: they are thus inserted in the picking list randomly. In other words, when the LSFO policy is adopted, products inside the warehouse can be ranked even when they enter the warehouse in the same time. This ranking is clearly determined by the stochastic variation in the RSL of the incoming products.

The simulation model hence aims to determine the average RSL of products when they leave the warehouse, when the FIFO and LSFO policies are applied. To build the two simulation models the following assumptions and notation has been used: the interarrival time of the batches is deterministic and equal to  $(EOQ - \text{Total perished}) / \text{Demand rate}$ . For each configuration the EOQ quantity arriving at the warehouse is determined by equation (7) based on the demand rate and the deterioration rate reported in Table 1. The EOQ corresponds to one batch. For each SKU in the batch is assigned a RSL value within the normal distribution; the RSL of the products entering the warehouse decreases with time each day the stocks are held in the warehouse. The RSL at the shipping time is then determined based on equation (5), by fixing  $E_a$  equal to  $59.7 \text{ KJ mol}^{-1}$  (which is the activation energy of  $\text{CO}_2$  production for the avocado fruit, Fonseca 2001) and by considering that the storage temperature is equal to  $10^\circ\text{C}$ .

To ensure that the values determined by the simulation satisfy the principles of quality control, a required precision has been fixed equal to 95%. Thus 10 warehouse replenishments cycles corresponding to 10 replications of the simulation have been carried out in the two cases in which LSFO and FIFO are applied and the confidence interval for each measure of interest has been determined. Then the equation (24) has been employed to calculate the number of replications needed in the two cases. The number of replications necessary to ensure the desired precision for each of the measures of interest for the experimental plan is 23 equivalent to  $8 \times 23 = 184$  tests for each of the two scenarios (FIFO and LSFO). The performances of the warehouse system for the two policies under study are illustrated in Table 2 and Figure 4.

Table 2: Results of the Experimental Plan

Scenario	Average RSL (Days)		$\sigma_{\text{RSL}}$ (Days)		Average Number of products with a $\text{RSL}_{\text{exit}}$ equal to Average $\text{RSL}_{\text{exit}} \pm 1$	
	LSFO	FIFO	LSFO	FIFO	LSFO	FIFO
1	16.835	16.862	1.442	3.596	75	30
2	20.835	20.862	1.442	3.596	75	30
3	17.424	17.429	1.416	3.391	46	29

4	21.424	21.429	1.416	3.391	45	29
5	18.008	18.034	1.996	3.273	57	32
6	22.008	22.034	1.996	3.273	57	32
7	18.420	18.426	1.936	3.144	56	30
8	22.420	22.426	1.936	3.144	56	30
Average value	19.672	19.6875	1.697	3.35	58.375	30.25

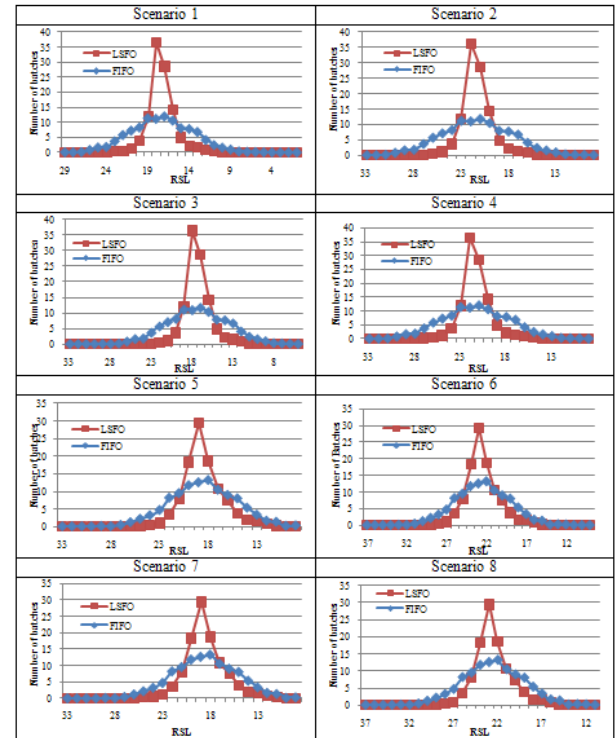


Figure 4: Average Number of products leaving the warehouse and their RSL for the two picking policies in each scenario of the Experimental Plan

For both the FIFO and the LSFO policy, in each test configuration and for each run, the number of products leaving the warehouse with the same  $\text{RSL}_{\text{exit}}$  has been calculated. The average of such values in the 23 runs has finally been determined and reported in Table 2. The results show that the average  $\text{RSL}_{\text{exit}}$  of products leaving the warehouse is the same for both LSFO and FIFO policies, while the average standard deviation differs substantially (in the FIFO policy is more than two times higher than LSFO). In any case, as expected, the best value of the response is obtained when the RSL of the incoming products is at its high level, the deterioration rate is at its low level and the demand rate is at its high level (Scenario 6 of the experimental plan). Table 2 also shows that when the LSFO policy is applied a greater number of products have a RSL about equal than the average RSL value  $\pm 1$  compared to the case in which the FIFO policy is applied allowing an improving in quality of 46.64%.

Based on the experimental plan realized an analysis of the factors has been conducted to determine

their impact on the response in the two cases in which LSFO and FIFO policies are applied. Results are shown in Table 3 and 4.

Table 3: Factorial Analysis for LSFO and FIFO policy

LSFO	Effect	Coefficient	T	P
Constant		19.6718	893.41	0.00
RSL	4.0	2.0	90.83	0.000
$\theta$	0.5008	0.2504	11.37	0.000
R	1.0847	0.5424	24.63	0.000
FIFO				
Constant		19.6875	895.64	0.00
RSL	4.0	2.0	90.99	0.000
$\theta$	0.4794	0.239	10.90	0.000
R	1.0846	0.5423	24.67	0.000

Table 4 ANOVA for LSFO and FIFO policy

ANOVA		
LSFO	DF	SS
Main Effects	3	34.8548
Residual Error	4	0.0155
Total	7	34.8703
FIFO		
Main Effects	3	34.8121
Residual Error	4	0.0155
Total	7	34.8275

Results show that in both cases the initial RSL, the demand rate and the deterioration rate has positively influence on the response. In both cases the effect which contributes mostly to improving the performance is initial RSL. The ANOVA analysis shows that the three main factors are responsible for the 99.9% of the total variance.

Based on coefficients shown in Tables 3 the response surface are determined for the experimental plan with equation (16) by considering only the first order terms. The two response surfaces referred respectively to LSFO and FIFO responses are:

$$y = 19.67 + 2RSL + 0.250\theta + 0.542R + \epsilon \quad (26)$$

and

$$y = 19.68 + 2RSL + 0.239\theta + 0.542R + \epsilon \quad (27)$$

The high value of R-Sq equal to 99.9% for LSFO and FIFO shows that the linear model obtained with the two response surfaces actually represents the relation between the predictors (initial RSL, deterioration rate and demand rate) and the response ( $RSL_{exit}$ ). This result can be employed to investigate the behavior of the response without the need of to run further simulations thus allowing to gain fast information about the warehouse system behavior when the input parameters vary.

## 7. CONCLUSIONS

The optimal management of supply chain for perishable products is a relevant research topic which has recently gained attention due to the poor sustainability of such systems. Agri-Food supply chains are in fact affected by several inefficiencies which typically result in a high industrial cost and in an ethical and social concern. Since deterioration phenomena are generally influenced by the temperature, the logistic operations are carried out in refrigerated conditions, in the so-called cold chain. Ensuring a specified temperature through the supply chain, however, does not ensure the avoidance of deteriorated products through the supply chain, since other important supply chain parameters such as the order quantity or the lead times have a strong effect on the performance of the chain. Finally, some uncertain parameters are involved in the process, which must be taken into account. Optimizing the performance of such systems is hence a complex task, which involves a good knowledge of the deterioration processes in order to properly assess the deterioration costs. In this paper a simulation model is proposed to evaluate the performance of a warehouse system for perishable products taking into account the effect of the uncertainties which typically affect the supply chain. In particular, a mathematical model is employed to determine the optimal stock level and the effect of shelf-life based picking policy on the performance of the supply chain. The results of the experimental plan proposed show that in the case considered the average  $RSL_{exit}$  of products leaving the warehouse is about the same for both LSFO and FIFO policies, while the average standard deviation differs substantially (in the FIFO policy is about two times higher than LSFO). When the LSFO policy is applied, thus, a greater number of products have a  $RSL_{exit}$  about equal than the average  $RSL_{exit}$  value  $\pm 1$  to compared to the case in which the FIFO policy is applied allowing an improvement in the quality of 46.64%. The effect of input parameters has been studied emphasizing the role of the deterioration rate as the factor which the most influences the response. Finally the response surface has been built representing a tool able to predict the warehouse behavior for any variation of the input factors considered allowing the optimization of perishable warehouse management policies.

## REFERENCES

- Andradottir, S., 1998. A review of simulation optimization techniques. *Proceedings of the 1998 Winter Simulation Conference*, 151–158.
- Banerjee, A., Burton, J., Banerjee, S., 2003. A simulation study of lateral shipments in single supplier, multiple buyers supply chain networks. *International Journal of Production Economics*, 81-82, 103-114.
- Begum R., Sahoo R. R., Sahu S. K., and Mishra M., 2010. An EOQ model for varying Items with Weibull distribution deterioration and price-

- dependent demand. *Journal of Scientific Research*, 2 (1), 24-36.
- Broekmeulen, R., Van Donselaar, K., 2007. A replenishment policy for a perishable inventory system based on estimated aging and retrieval behavior. *BETA working paper* nr. 218.
- Doyle, J.P. 1995. Seafood shelf life as a function of temperature. Alaska Sea Grant Marine Advisory Program, 30, 1-5.
- Fonseca S.C., Oliveira F., Brecht J., 2001. Modelling respiration rate of fresh fruits and vegetables for modified atmosphere packages: a review. *Journal of Food Engineering*, 52, 99-119.
- Fu, M., Hu, J. Q. 1997. Conditional Monte-Carlo: Gradient Estimation and Optimization Applications. *Kluwer Academic Publishers*.
- Giannakourou M.C., Taoukis P.S., 2003. Kinetic modelling of vitamin C loss in frozen green vegetables under variable storage conditions. *Food Chemistry*, 83, 33-41.
- Ghosh S.K., Chaudhuri K.S., 2005. An eoq model for a deteriorating item with trended demand, and variable backlogging with shortages in all cycles. *Advanced Modeling and Optimization*, 7 (1), 57-68.
- Gong, Y., 2009. *Stochastic Modelling and Analysis of Warehouse Operations*. Thesis (PhD). Erasmus Universiteit Rotterdam.
- Kelton, W. D., Barton, R., 2003. Experimental design for simulation. *Proceedings of the 35th Conference on Winter Simulation: Driving Innovation*, 2044-2050, December 7-10, 2003, New Orleans, Louisiana, USA.
- Ketzenberg, M.E., Bloemhof, J., 2009. The Value of RFID Technology Enabled Information to Manage Perishables. *Erim Report Series Research in Management*, ERS-2009-020-LIS.
- Law, A. M., Kelton W. D., 2000. *Simulation Modeling & Analysis*. Third Edition New York: McGraw-Hill.
- Manna S.K., Chaudhuri K.S., 2006. An eoq model with ramp type demand rate, time dependent deterioration rate, unit production cost and shortages. *European Journal of Operational Research*, 171 (2), 557-566.
- Nita H., Poonam P., 2008. Optimal ordering policy for the time dependent deterioration with associated salvage value when delay in payments is permissible. *Revista Investigación Operacional*, 2 (2), 117-129.
- Raissi, S., 2009. Developing New Processes and Optimizing Performance Using Response Surface Methodology. *World Academy of Science, Engineering and Technology*, 49, 1039-1042.
- Sezen, B., Kitapçı, H., 2007. Spreadsheet simulation for the supply chain inventory problem. *Production Planning & Control: The Management of Operations*, 18 (1), 9 - 15.
- Shibsankar Sana and K. S. Chaudhuri, 2004. A stock-review eoq model with stock-dependent demand, quadratic deterioration rate. *Advanced Modeling and Optimization*, 6 (2), 25-32.
- Singh, T. K., Cadwallader, K. R., 2004. Chapter 9. In: Steele, R., 1st edition. *Understanding and measuring the shelf-life of food*. Australia: Food Science Australia, p. 171.
- Taoukis P.S., Bili M., Giannakourou, M., 1998. Application of shelf life modeling of chilled salad products to a TT based distribution and stock rotation system. *International Society for Horticultural Science*, 476.
- Taoukis, P.S., Koutsoumanis, K., Nychas, G.J.E., 1999. Use of time-temperature integrators and predictive modelling for shelf life control of chilled fish under dynamic storage conditions. *International Journal of Food Microbiology*, 53, 21-31.
- Tarun Jeet Singh, Shiv Raj Singh, Rajul Dutt , 2009. An eoq model for perishable items with power demand and partial backlogging. *International Journal of Operations and Quantitative Management*, 15(1), 65-72.
- Van Donselaar, K., Van Woensel, T., Broekmeulen, R., Fransoo, J., 2006. Inventory control of perishables in supermarkets. *Journal of Production Economics*, 104 (2), 462-472.
- Whitt, W., 2005. Analysis for the design of simulation experiments. In: Henderson, S., Nelson, B., Simulation. Elsevier series of Handbooks, *Operations Research and Management Science*, Chapter 13.
- Wilson, R. H., 1934. A Scientific Routine for Stock Control. *Harvard Business Review*, 13, 116-128.
- Young, J., Blau, G., Pekny, J., Reklaitis, G., Eversdyk, D., 2004. A simulation based optimization approach to supply chain management under demand uncertainty. *Computers & chemical engineering*, 28 (10), 2087-2106.
- Zhao, X., Xie, J., Wei, J., 2002. The Impact of Forecast Errors on Early Order Commitment in a Supply Chain. *Decision Sciences*, 33 (2).

# COMMODO – COMPLEX MATERIAL MODELLING OPERATIONS

## A COMPREHENSIVE APPROACH TO THE MODELLING OF COMPLEX MATERIALS WITH MACHINE LEARNING MODELS WITHIN FINITE ELEMENT SIMULATIONS

Andreas Kuhn, Toni Palau, Gerolf Schlager<sup>(a)</sup>, Helmut J. Böhm, Sergio Nogales<sup>(b)</sup>,  
Victor Oancea, Ritwick Roy<sup>(c)</sup>, Andrea Rauh, Jürgen Lescheticky<sup>(d)</sup>

<sup>(a)</sup> ANDATA GmbH, Hallein, Austria

<sup>(b)</sup> Institute of Lightweight Design and Structural Biomechanics, Vienna University of Technology, Vienna, Austria

<sup>(c)</sup> ABAQUS SIMULIA, Providence, RI, USA

<sup>(d)</sup> BMW AG, Munich, Germany

<sup>(a)</sup> [andreas.kuhn, toni.palau, gerolf.schlager}@andata.at](mailto:{andreas.kuhn, toni.palau, gerolf.schlager}@andata.at),

<sup>(b)</sup> [hjb, snogales}@ilsb.tuwien.ac.at](mailto:{hjb, snogales}@ilsb.tuwien.ac.at),

<sup>(c)</sup> [victor.oancea, ritwick.roy}@3ds.com](mailto:{victor.oancea, ritwick.roy}@3ds.com),

<sup>(d)</sup> [andrea.rauh, juergen.lescheticky}@bmw.de](mailto:{andrea.rauh, juergen.lescheticky}@bmw.de)

### ABSTRACT

Due to the increasing usage of complex materials in lightweight design the development of proper material models for the prediction of damage and failure within Finite Element simulations has become an extensive task. Other fields of application already have shown that the introduction of Soft Computing and Machine Learning methods can be very beneficial for getting the complexity under control. The contribution aims at sketching a systematic approach to the application of machine learning methods in the field of material modelling. The focus is put not only on the definition of well performing mathematical models, but also on process aspects of generating and maintaining the mathematical models within reproducible, requirement-driven and controlled iterative environments for Computer Aided Engineering.

Keywords: material modelling, finite element simulation, soft computing, machine learning, artificial neural networks

### 1. INTRODUCTION

The Finite Element Method (FEM) is a well established technique in a wide range of disciplines, most successfully applied in mechanical engineering. In the automotive and aerospace industries the FEM is commonly used for structural design and development of new products. It enables modern lightweight designs and facilitates the development and usage of new and highly specialized materials (composites, advanced aluminium alloys, high strength steels, etc.) as well as joining technologies (spot welding, adhesive bonding, etc.).

The quality of FEM results depends strongly on the availability of appropriate material models that describe the nonlinear behaviour and failure of these materials.

The development and identification of new material models has become an increasingly complex and expensive process, partly as a consequence of the conventional approach of using purely analytical, physically motivated mathematical models for predicting material behaviour. Other fields of application have shown, however, that example based approaches, such as Machine Learning and Pattern Recognition, may outperform conventional methods and significantly reduce the development effort (Bishop 2006). Furthermore, their superiority even increases with the complexity of the problem.

At present the application of Machine Learning and Soft Computing methods to material modelling in computational mechanics, see, e.g., Lefik and Schrefler (2003), Hashash, Jung and Ghaboussi (2004), Aquino and Brigham (2006), or Kessler, El-Gizawy and Smith (2007), is rather marginal. The research project *CoMMoDO* (*Complex Material Modelling Operations*), which is funded by the Austrian Research Promotion Agency and the Austrian Federal Ministry of Transport, Innovation, and Technology under the initiative "*ModSim Computational Mathematics*" within the program "*Research, Innovation and Technology in Information Technology*", is intended to overcome this deficit. The proposed solution strategy is the smart combination

- of Machine Learning as well as Soft Computing methods (example based modelling)
- plus the interpretation of experimental data in the framework of time-series classification and forecasting problems
- with an object-oriented design approach in order to develop a requirements-driven process,



for the modelling of material behaviour in FEM simulations.

The present paper aims at giving an overview of several points of attack where Soft Computing and Machine Learning techniques can be used for achieving improvements in the modelling of complex material behaviour to be applied in Finite Element Methods. A number of approaches to tackling this problem are summarised, which can be understood as escalating steps in a controlled process for developing and building material models. Although here applications to the modelling of weld spots are presented as examples for most of these steps, the basic CoMMoD idea and methodology are applicable to other complex material types and joining technologies as well.

## 2. ESTIMATING THE KEY FEATURES OF MATERIAL BEHAVIOUR

When carrying out material tests one typically obtains a number of curves like the one shown in Figure 1, where the applied force is plotted against the displacements of a test specimen. In the first run we describe the curve via parameters that characterize its key features. In the case of Figure 1, these parameters could be the elastic stiffness,  $E$ , of the initial slope of the curve, the maximum force  $F_{max}$ , or some integral associated with energies, e.g.,  $W_{max}$ . Such parameters may have a physical meaning and are often used as key parameters and coefficients in conventional analytical material models. Nevertheless, their prediction and estimation based on material specifications and load conditions can be challenging.

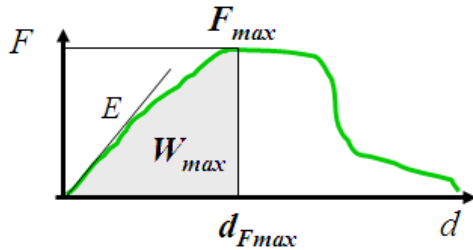


Figure 1: Generic force-displacement curve from a material test.

### 2.1. Description of Physical Parameters

In our first process step we use Machine Learning methods for the definition and construction of models that describe such physical parameters. Assuming that a suitable number of test results are available, the first step consists in evaluating the measurements extracted from them. This process step can also be seen as a first Data Mining session on all the material data of the given test set. This way one gains insight into the necessary parameters and material specifications that contribute to the problem, as to which material properties are important for the description of physical parameters of the material behaviour. This first step constitutes a method for describing the material

behaviour, which works at least for materials that are not excessively complex.

An example of this can be seen in Figure 2, where the response of a model for estimating the maximum axial force sustained by weld spots under different loading conditions and for different combinations of sheet materials and sheet thicknesses is plotted against the values measured in a series of some hundred experiments. Each dot represents a test case. Blue dots indicate that the sample was used for training the Machine Learning model. Red and green dots denote test and validation samples, which were not seen by the model during the training and tuning phase and therefore can be used for validation to assess the generalization properties of the model.

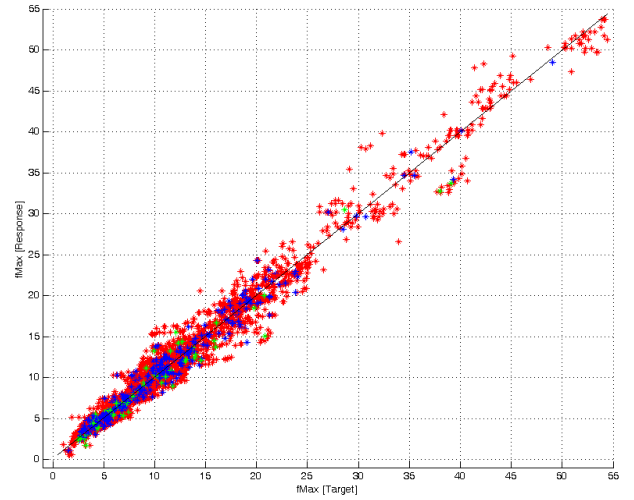


Figure 2: Estimates of maximum force sustained by a weld spot.

The resulting models can be employed for estimating the material parameters for various conditions and may be inserted accordingly into the Finite Element material description.

### 2.2. Process Scheme for Material Model Design

It is a very important point that the generation and maintenance of the prediction models are embedded into a proper process scheme.

Of course, the models cannot predict behaviour that has not been seen before and that is not represented by the test cases and the parameterisation in some way. However, the adaptation of the models to new data can be done very quickly if this behaviour is already described by the given or measured parameters. In case parameters are missing from the model one gets the confirmation that additional parameters are required. The process also automatically supports the identification of these parameters. In the case that some candidate parameters are available and measured in the data set, the identification of these parameters can be performed with the help of so called feature selection algorithms (Kittler 1986, Somol and Pudil 2000).

The flow chart shown in Figure 3 represents a standard procedure within Data Mining processes. The

important issue is that it should be implemented as a standard procedure in material modelling as well.

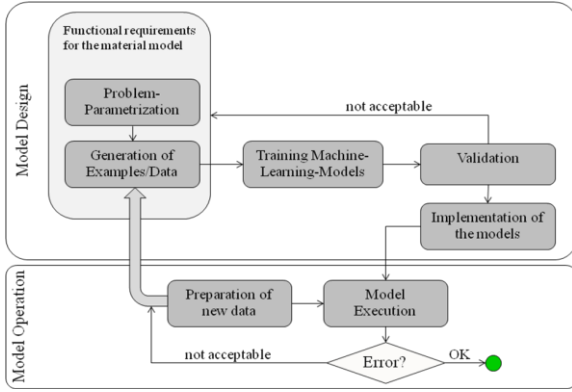


Figure 3: Machine Learning based process for material modelling.

When executing a series of material tests, one needs a good guess of the parameters required for the description of the desired behaviour to start with. These parameters then need to be measured and collected consistently during the material tests. It is important to support this procedure at a strategic level, so that the test data sets collected over time and different test runs and organizations can be put into one common context and database. Then Data Mining methods can be applied by generating Machine Learning models. The general Data Mining procedures (Witten and Frank 2005), with their accompanying validation processes, provide feedback about which parameters are relevant for describing and predicting the desired material behaviour. The methods give referable evidence whether the chosen parameters and test cases are sufficient or if new parameters are needed for a correct prediction of the material behaviour. Once the resulting models meet the performance goals, they can be implemented for daily use. When new material test data is available, the model predictions should be promptly evaluated and assessed. If the predictions do not match the experimental data, the models can be easily adapted by including the new data in further training of the prediction models. If the new data shows behaviour that cannot be described with the given parameterisation, the design process described above needs to be started from the beginning.

Among the major advantages of the Machine Learning based approach are its capabilities for preserving and reusing given knowledge. Also, it can be adapted to new experience much more quickly and efficiently than is generally possible with purely analytical approaches. If models are formulated analytically, it commonly happens that one has to begin from scratch if new data does not fit the existing model structure, so that the danger of running into extensive trial and error scenarios increases with the complexity of the problem.

### 2.3. Material Model Validation using Monte Carlo methods

The best prediction of the key parameters is of little use if the underlying material model is not sufficiently flexible for describing the behaviour of the material under study. If the estimated key material parameters are inserted into the material model, a validation of the model's performance still has to be carried out. If the performance goals are not met, further efforts are required.

In order to assess the flexibility of a given material model, a standard approach is to use Monte Carlo Simulation (Metropolis and Ulam 1949, Rubinstein and Kroese 2008). In such an analysis, material model parameters are scattered within suitable ranges to test if the material model is sufficiently flexible for covering the behaviour measured in a set of experiments. If this flexibility is found to be sufficient, the rest of the required material model parameters that were not estimated in the previous step can be identified by means of an optimisation process, for example an Evolutionary Algorithm (Bäck 1996) that seeks to minimise the deviation from the experimental measurements. The search space for this optimisation has been covered by the exploratory Monte Carlo Simulation carried out before, ensuring that this optimisation is able to achieve the required parameter identification.

An example of such an analysis is presented in Figure 4, which shows several performance parameters for the resulting Monte Carlo samples (black dots) and some results from material tests (red dots). The cloud of

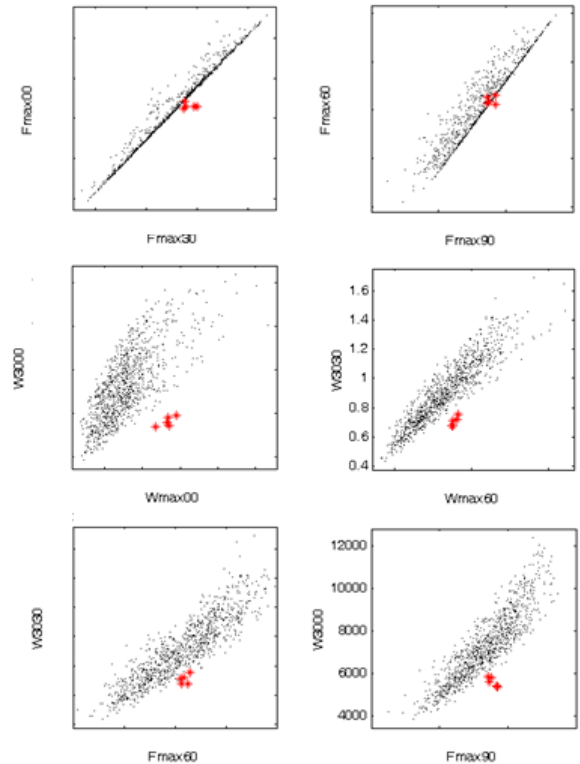


Figure 4: Validating a model by means of a Monte Carlo Simulation approach.



black dots identifies the region of parameter space that can be reached with a given analytical model. In the example shown, the analytical model lacks sufficient flexibility for concurrently reaching the material test results for several key parameters. It can be concluded that this analytical model is not capable of satisfactorily reproducing the targeted material behaviour.

### 3. EXAMPLE BASED DESCRIPTION

In the case of complex material behaviour, analytical approaches may reach their limitations in providing suitable descriptions or models, and example based approaches like Machine Learning may in principle be the better choice or at least a valuable complement. For example, the universal approximation properties shown by Artificial Neural Networks allow the reproduction of any given functional relationship (Hornik, Stinchcombe and White 1989), as long as a sufficient number of neurons are used in the hidden layer(s).

#### 3.1. Interpretation as a Time-Series Problem

In order to extract the maximum information out of experimental data, the problem is formulated as a time series prediction. From this point of view, each one of the samples constituting, for example, an experimental force-displacement curve can be seen as a training example for the Machine Learning models.

This different view of the problem helps in overcoming the main counterargument used against Machine Learning models as well, namely that such approaches require excessive amounts of data and numbers of tests for proper models to be trained. This line of reasoning comes from the practical use and mindset of Design of Experiments approaches, where the main features of the results are already determined by the choice of the model's structure in the beginning. The test data are then only set up and used for fixing some coefficients of the predetermined assumption. When using analytical models, most aspects of their behaviour are already predetermined by the choice of ansatz functions.

With the interpretation as a time series problem the rate effects and path dependent behaviour become conceptually uncomplicated, too. Historic values, accumulated signal features or special filter values can be introduced as input parameters in an easy way. A further alternative is using recurrent Artificial Neural Networks for the incorporation of history dependent behaviour.

#### 3.2. Example based description of weld spot behaviour

For the special case of modelling the behaviour of weld spots, the force-displacement curves obtained in some hundred experimental tests were used to train a neural network that models the responses of specimens that were obtained by spot welding sheets of different compositions and thicknesses and were subjected to different load cases. The quality of one of the resulting "global" models can be seen in Figure 5. There, the black lines are the measured results from the material

tests, whereas the red lines and red and green crosses depict validation and test curve samples that were not used in the training process. Note that this figure shows only 20 out of about hundred similar curves with comparable performance, the generalization properties of the neural network being excellent in this case.

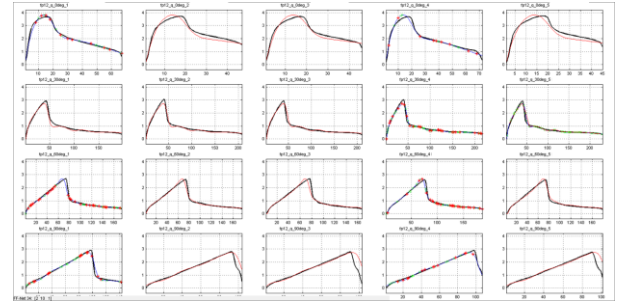


Figure 5: Example based training of a global model of a spot weld specimen.

Despite being very satisfactory, these results do not provide a solution to the problem of modelling the material behaviour of weld spots, because the load-displacement curves in Figure 5 pertain to whole specimens (and thus include the responses of the two sheets plus some influence of the testing apparatus) rather than to the weld spots.

#### 3.3. Classification and Hybrid Modelling

Another way of benefiting from the advantages of Machine Learning approaches in the modelling of material behaviour can take the form of using them only for some components of an existing analytical model. For example, the indication of the presence of damage or the classification of the damage mode may be done via a time series classification model. Such a model may take as inputs strain and stress states, with some history values, and a number of additional material and load specifying parameters, to return values 1 or 0, indicating whether or not a given integration point shows damage. It is then up to the analytical model, into which the classification model is embedded, to make appropriate use of this information.

The advantage of this approach is the fact that normally classification problems are technically much easier to solve than regression problems. However, in practice the onset of damage is very difficult to detect in material tests. In addition, care must be taken of the issue that the measured curves incorporate the effects of damage, whereas damage classifiers require data that is free of damage.

Before going more deeply into the issues of example based models for weld spot behaviour, however, a short discussion of some general aspects of weld spot models for FE analysis will be presented in Section 4.

### 4. WELD SPOT MODELS

Finite Element models of weld spots fall into two groups. On the one hand, detailed models have been

proposed, see, e.g., Lamouroux et al. (2007), which are outside the scope of the present discussion. On the other hand, “simplified” weld spot models have been developed explicitly for use in large scale analysis, such as crash simulations. Such models must combine computational efficiency, the capability of handling nonlinear behaviour with sufficient accuracy for general load paths, and features that facilitate automatic mesh generation. The latter issue leads to the demand of supporting the use of non-congruent meshes for the sheets to be joined by spot welding.

The present work is based on the Finite Element code ABAQUS/Explicit (SIMULIA, Providence, RI), which comes with a simplified weld spot model in the form of the FASTENER option. This combines connector elements, which allow modelling various types of nonlinear material behaviour including plasticity, damage and failure, with coupling constraints for consistently handling a mesh-independent connection to the sheets, which are modelled by shell elements. Material responses of weld spots are introduced in terms of forces and displacements, i.e., they do not describe a material behaviour in the strict sense, but rather a “structural behaviour” of the weld spot and the closely neighbouring regions of the sheets (for brevity, however, the expression “material behaviour” is used throughout the present work). The path dependences of plasticity and damaged behaviour are handled via status variables, damage indicators and damage evolution algorithms. The resulting, highly robust material descriptions are complex analytical models in the sense of Section 3, and their parameters must be identified from suitable experiments.

Such experiments involve testing to destruction standardized test specimens, such as the KS-II specimen (Hahn et al. 2000) shown in Figure 6, which consists of two U-shaped sheets that are joined, e.g., by a spot weld. When suitable fixtures are employed, a universal testing machine can be used to carry out tests for different load cases, such as normal, shear and mixed loading as well as peeling.

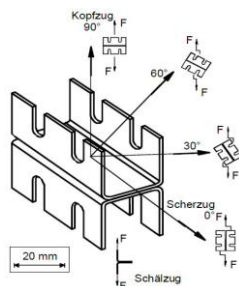


Figure 6: KS-II test specimen used for characterising weld spots.

The results of experiments of this type are global force-displacement curves that contain contributions from the weld spot, the test fixtures and the metal sheets, with the latter dominating the displacement responses in most cases. If the sheets are very stiff, such data can provide reasonable approximations to the

behaviour of the weld spot proper. In general, however, these global force-displacement curves characterise the behaviour of the whole test setup rather than the local responses of the weld spot.

Therefore, the global results cannot be used directly for describing the material behaviour of any type of weld spot model. This also holds true for FE-based simplified weld spot models of the type discussed above, considerably complicating their practical use. In analogy, when a neural network is trained with the global force-displacement curves, as is the case for the approach introduced in Section 3.2 and the data shown in Figure 5, the resulting model in general does not represent the local behaviour of the weld spot and cannot be used in Finite Element models of the latter.

Some additional complications have to be kept in mind. On the one hand, radial loading at the global level (i.e., keeping constant the direction of the force acting on the specimen) does not necessarily imply locally radial loading of the weld spot due to geometrical effects of the deformation of the sheets. On the other hand, the weld spot model not only must be capable of describing the failure of the weld spot proper (e.g., “nugget fracture” or “nugget pullout”) but also failure of the sheets in the immediate vicinity of the weld spot (“sheet tearing”).

To overcome the issue of global vs. local behaviour, a solution in the form of introducing an indirect training procedure is proposed, which accounts for the differences between behaviour at the local (element) and the global (component) levels.

## 5. INDIRECT TRAINING OF A NEURAL WELD SPOT MODEL FOR RADIAL TENSILE LOADS WITHIN A FINITE ELEMENT SIMULATION

In order to account for the complex and nonlinear behaviour of weld spots, the local material description in the connector part of an ABAQUS FASTENER is to be modelled by a neural network. This neural network is to be trained with the only data available, the global load-displacement curves obtained from a series of experiments. Once trained, the neural network is expected to reproduce the behaviour encoded in these data.

As noted above, extracting the local behaviour of weld spots from measurements of the global responses lies at the root of a major complication in the use of simplified weld spot models. The most common expedient in such situations is to make recourse to involved calibration methods for finding suitable parameters for the weld spot model. In contrast, a systematic indirect training process applicable to weld spot models is proposed in the following, in which the material behaviour is described by neural networks. In this process the neural network model for the weld spots’ material behaviour is embedded in a Finite Element simulation during training and thus contributes to predicting the global behaviour of the whole component of which the spot weld is a part. Comparing

the results of a given simulation with the experimental data obviously allows assessing the performance of the full model. Because the local behaviour of the weld spot, described by the neural network, is the only variable part of the global model of the sample, the latter's performance allows drawing conclusions on the performance of the local model. On this basis, in turn, the weights and biases of the neural network can be updated, i.e., training can be carried out.

In a first step, discussed in the following, the use of the neural network model is restricted to globally radial and monotonic load cases, so that issues of the path dependence of plasticity and damaged behaviour are of little relevance. The extension of the approach to non-radial and non-monotonous load paths is ongoing work.

Even though this problem may sound very specific to weld spots and their test configurations, the problem pattern is generic and occurs in many types of material tests, namely that the material behaviour cannot be measured directly at the same level of interface as, required for the Finite Element formulation. Once we find a method for training Machine Learning models not for their direct output but for the effect of their output at a higher level of abstraction (such as the behaviour of the whole test configuration), the problem is formulated in a way that any type of test can be used to train the models. For the special case of weld spots this may imply testing not only specimens with a single weld spot but also test structures containing several ones, so that a sufficient variety of load paths can be sampled

### 5.1. Integration into the Finite Element Simulation as User-Defined Element

As mentioned above, in ABAQUS/Explicit the standard tool for modelling weld spots is the FASTENER option, within which the proper place for integrating new material descriptions – including models based on neural networks – are the connector elements. At present, however, ABAQUS/Explicit does not provide a user subroutine interface for specifying the mechanical behaviour of connectors, so a workaround had to be used. This took the form of combining kinematics appropriate for weld spots, as defined by SIMULIA's BUSHING connector type, with a user defined element (VUEL) having two nodes with three translational and three rotational degrees of freedom each. The neural network was integrated as a code sequence within this VUEL. To complete the FASTENER surrogate, the weights of the coupling constraints to the shells are required, which can be extracted from the output of an analysis employing a standard FASTENER. Bushing kinematics was provided by SIMULIA as linked functions.

The task of the neural network implemented within the user defined element consists in returning a set of suitable generalized force increments for any set of increments of the generalized displacements passed in from the Finite Element code. Here, both force and displacement vectors are given in the connector's local coordinate system. Because an explicit time integration

scheme was chosen, there is no need for providing Jacobians.

Within this framework, the weights and biases of the neural network can either be specified as parameters within the user defined element or they can be provided via the ABAQUS input file. Both of these options fully support dynamic updating in the course of training of the neural network.

### 5.2. Assessment of Model Performance

For the task at hand the performance of a given neural network model is assessed by comparing the force-displacement curve measured in a suitable experiment with the one predicted by a model of the whole sample which incorporates the VUEL containing the neural network. The task of the performance assessment function is to formalise the computation of such a performance, i.e., to provide an algorithm for computing a performance value that correlates with how well the simulated material behaviour used for the weld spot allows matching the experimental measurements.

Performance is assessed by combining several criteria, each one focusing on a specific aspect of the general performance. The assessment criteria used are the following:

- Difference between experimental and simulated force-displacement curves using a *mean squared error (MSE)*, evaluated at discrete points chosen with consideration of the shape of the curve.
- Agreement in terms of maximum force: The simulation should predict a similar value for the maximum force as the one measured in the experiment.
- Agreement in terms of displacement at maximum force: The simulation should achieve the maximum force at a similar value of total displacement as measured in the experiment.
- Agreement in terms of displacement at failure: The simulation should predict a similar value for the displacement at total failure of the weld spot as the one measured in the experiment.
- Characteristics of the local curve can be taken into account as well, for example maximum force reached, convexity, etc.

For each one of these assessment criteria, an error value normalised to the interval [0, 1] is generated, with 0 denoting the most desirable outcome, i.e. corresponding to the best performing model. Then a single scalar error measure is obtained from these normalised errors by means of a weighted average. This allows for a dynamic adjustment of the relative importance of some of the assessment criteria with respect to others.

Once a performance assessment function has been defined, it can be used as objective function for an optimisation process that considers the weights and

biases of the neural network as optimisation variables and seeks to minimize the error value. This optimisation constitutes, indeed, the training of the neural network.

### 5.3. Indirect Training of the Neural Material Model

In general, the training of a neural network can be understood as the solution of an optimisation problem in which values of weights and biases are sought, which minimize the network error obtained by comparison of the network's response and the available targets. By minimizing the network error through this optimisation, the neural network is trained to reproduce the behaviour represented by the target data.

The same interpretation can be applied when the network's response cannot be directly compared to the available target values. In such cases, the definition of a suitable performance assessment function that is able to map the local (element) network responses into global (component) curves comparable to the targets turns out to be the key to make this kind of training feasible.

The performance function requires an additional step to map the local network responses to global curves that are comparable to the targets, hence the name "indirect training" for this kind of process. In the case of the neural connector model, the Finite Element simulation of the whole specimen is part of the performance assessment function, as shown in Figure 7.

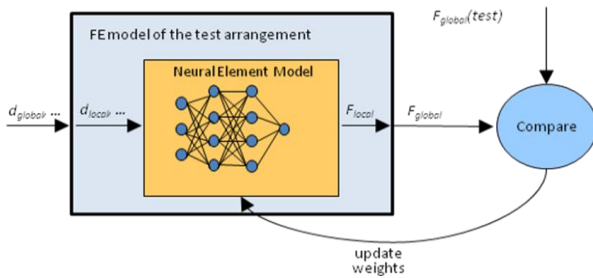


Figure 7: Schematic indirect training process.

In order to solve a generic optimisation problem, two steps must be performed:

1. *Initialisation*: Find one or more suitable initial points to start the optimisation from.
2. *Optimisation*: Apply an optimisation algorithm to find a point that minimises or maximises the objective function, starting from the given initial point(s).

In most applications, the initialisation step turns out to be at least as important as (and sometimes even more challenging than) the optimisation step itself in order to find an optimum of the objective function.

The heuristics used to initialise the neural network weights and biases comprised a pure random weight initialisation and the neural network specific *Nguyen-Widrow rule* (Nguyen and Widrow 1990), which takes into account network topology and ranges of inputs in providing initial values for the network weights and biases.

Depending on how many optima a given algorithm is designed to find, local and global optimisation methods can be distinguished. A *local optimisation* algorithm is able to find the local optimum closest to the current initial point. Other local optima, potentially better, cannot be found by a local optimisation algorithm. If a single optimum is known to exist, for example in the case of a convex objective function, a local optimisation algorithm may find the (unique) global optimum. For this reason, local optimisation is sometimes called *convex optimisation*. On the other hand, a *global optimisation* algorithm is designed to find several local optima and, from them, return the global optimum. These algorithms are often based on a population of quasi-independent agents that search for the global optimum, taking advantage of swarm intelligence based approaches.

As to the optimisation methods used for indirect training, the *Nelder-Mead simplex* algorithm (Lagarias, Reed, Wright and Wright 1998) was used for local optimisation, because it is a derivative-free method that does not require numerical estimates of the gradient of the objective function, which in the present indirect training setting includes the whole Finite Element simulation of the specimen. For global optimisation tasks *genetic algorithms* (Goldberg 1989) and *evolution strategies* (Bäck 1996) were used, all of which are derivative-free as well.

### 5.4. Software Framework for Indirect Training

For the sake of flexible testing of many different combinations of optimisation algorithms and initialisation heuristics, a generic optimisation software framework was devised that allows a very flexible replacement of the different modules thanks to its modular architecture.

In this software framework, *problems* and *solvers* are distinguished and kept independent from one another. A *problem* comprises a description of the search space (dimension, variable ranges, etc.) and an implementation of the objective function. A *solver* comprises an implementation of an initialisation strategy and an implementation of an optimisation algorithm. Applying a given solver to a given problem returns a solution, i.e., the optimum found by the solver starting from the specified initialisation of the objective function described in the problem.

For the indirect training of neural networks to be used as material models of a user-defined element within a Finite Element simulation of a spot weld specimen, the objective function was obtained by the following steps:

1. Set the current values of weights and biases in the neural network based user-defined element.
2. Perform a Finite Element simulation of the specimen using this parameterization of the user-defined element.
3. Compare the force-displacement curves obtained from this simulation with the



corresponding experimental measurements by means of the performance assessment function discussed above.

4. Return a value for the objective function based on the performance assessment function.

This software framework has been implemented using MATLAB (MathWorks Inc., Natick, MA), so that the existing and readily available MATLAB implementations of several optimisation algorithms (*Optimization Toolbox*, *Global Optimization Toolbox*) and initialisation heuristics for neural networks (*Neural Network Toolbox*) can be used.

### 5.5. Results from Indirect Training

In order to provide a proof of concept, an indirect training setup as presented above was tested against a synthetic problem for which both local and global load-displacement responses are known. The global behaviour for this test case was obtained by a Finite Element simulation, in which the local force-displacement behaviour was described by the standard ABAQUS FASTENER model for weld spots. This global force-displacement curve was then used as target for the indirect training of a neural material model.

Selected results of the indirect training process can be seen in Figure 8 and Figure 9, where two different optimisation algorithms were used together with the Nguyen-Widrow weight initialisation rule. For both cases a feed-forward neural network topology with 5 neurons in the hidden layer was employed. As the plots show, such an indirect training setup can give rise to valid models of the responses of both weld spot and specimen. As shown by the examples this approach is well able to cover the gap between local and global behaviour in the case of a weld spot model.

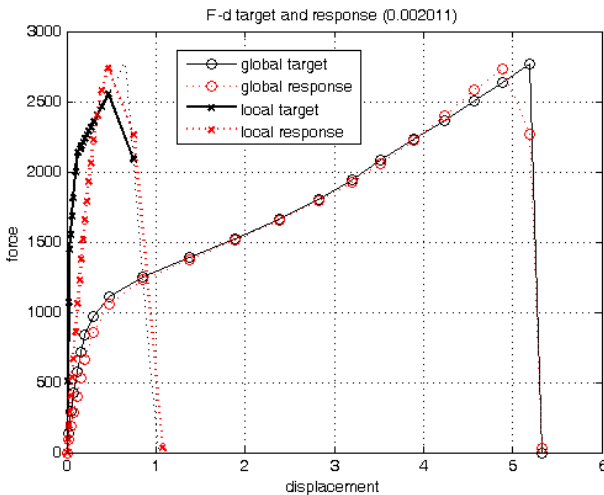


Figure 8: Global and local responses of a synthetic weld spot obtained by training with the Nelder-Mead simplex algorithm.

In general, local optimisation methods, such as the Nelder-Mead simplex algorithm, tend to reach slightly

better solutions as long as the initialisation already shows an acceptable performance. Global optimisation methods, such as Evolution Strategies, may perform marginally worse, but in general show a higher probability of finding an acceptable solution, being less dependent on the quality of the initialisation.

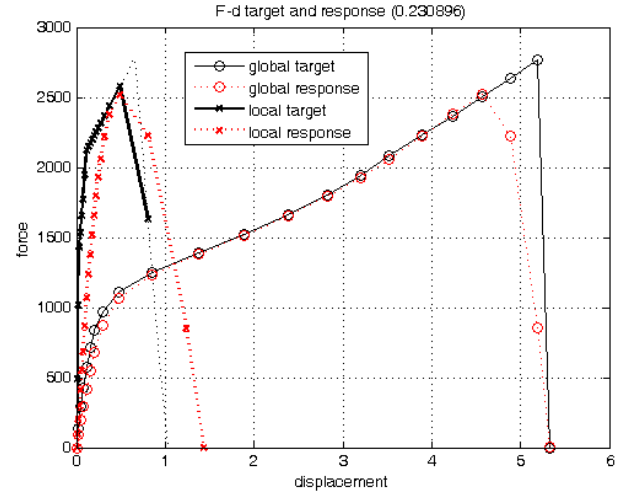


Figure 9: Global and local responses of a synthetic weld spot obtained by training with an Evolution Strategies algorithm.

When a model describing an actual weld spot, or any other kind of complex material or joint, is to be developed using this approach, the actual targets for the training process must, of course, be measurements obtained from appropriate experiments.

## 6. SUMMARY

The contribution discussed several approaches to using Machine Learning methods for material modelling within Finite Element simulation. A controlled process is sketched that ranges from the estimation and prediction of key feature parameters via descriptions of specimen behaviour at a global level to the use of Artificial Neural Networks for the direct description of material behaviour at the element level, escalating the use of Machine Learning models in dependence on the complexity of the material behaviour to be modelled.

For the example of weld spots a neural element model could be trained by introducing a scheme of indirect training. This approach allows evaluating the performance of the Machine Learning model at a higher abstraction level of the system, e.g., by using measurements involving the full test configuration. As a first step, the feasibility and prospective advantages of the approach were demonstrated for monotonous, globally radial loading. The generalization to arbitrary loads was under development at the time of writing the paper.

## 7. VISION AND OUTLOOK

The last major open issue in the presented approach is the extension to arbitrary loading conditions and

histories. Once this is achieved, the method can be extended and applied to any kind of material behaviour, e.g., to the failure of CFRP and other composite structures, adhesives, etc,

This new approach requires changing the way of developing and carrying out material tests, which will have to represent operationally relevant loading conditions rather than test cases defined explicitly for identifying some specific material parameters. This is expected to allow avoiding the complexity trap of an exponential increase of development effort with increasing complexity of the target behaviour, which tends to bedevil analytical models. The interim results reported here support the assessment that this goal can, in fact, be reached.

## ACKNOWLEDGMENTS

The paper presents interim results obtained within the research project *CoMModO (Complex Material Modelling Operations)*, which is funded and supported by the Austrian Research Promotion Agency and the Austrian Federal Ministry of Transport, Innovation, and Technology under the initiative "*ModSim Computational Mathematics*" within the program "*Research, Innovation and Technology in Information Technology*".

The research work has benefited greatly from the unconditional support of the crash simulation department of BMW, Munich, Germany. This paper is also dedicated to the memory of Dr. Heinrich Werner, whose strategic foresight contributed considerably to the initiation and mentoring of the project. Unfortunately he is no longer with us to see and discuss the results.

## REFERENCES

- Aquino, W., Brigham, J.C., 2006. Self-learning finite elements for inverse estimation of thermal constitutive models. *International Journal of Heat and Mass Transfer* 49, 2466–2478.
- Bäck, T., 1996. *Evolutionary Algorithms in Theory and Practice*. New York: Oxford University Press.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. New York: Springer.
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization & Machine Learning*. Reading, MA: Addison-Wesley.
- Hahn, O., Besserlich, G., Dölle, N., Jendrny, J., Koyro, M., Meschut, G., Thesing, T., 2000. Prüfung und Berechnung geklebter Blech-Profil-Verbindungen aus Aluminium. *Schweißen & Schneiden*, 52, 266–271.
- Hashash, Y.M.A., Jung, S., Ghaboussi, J., 2004. Numerical implementation of a neural network based material model in finite element analysis. *International Journal for Numerical Methods in Engineering*, 59, 989–1005.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, 2 (5), 359–366.

- Kessler, B.S., El-Gizawy, A.S., Smith, D.E., 2007. Incorporating neural network material models within finite element analysis for rheological behavior prediction. *Journal of Pressure Vessel Technology*, 129, 58–65.
- Kittler, J., 1986. Feature selection and extraction. In: Young, C.M., Fu, K.S., eds. *Handbook of pattern recognition and image processing*. New York: Academic Press, 60–81.
- Lagarias, J.C., Reeds, M.H., Wright, M.H. and Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9 (1), 112–147.
- Lamoureux, E.H.J., Coutellier, D., Dölle, N., Kümmerlen, D., 2007. Detailed model of spot welded joints to simulate the failure of car assemblies. *International Journal on Interactive Design and Manufacturing* 1, 33–40.
- Lefik, M., Schrefler, B.A., 2003. Artificial neural network as an incremental non-linear constitutive model for a finite element code. *Computer Methods in Applied Mechanics and Engineering*, 192, 3265–3283.
- Metropolis, M., Ulam, S., 1949. The Monte Carlo method. *Journal of the American Statistical Association*, 44 (247), 335–341.
- Nguyen, D., Widrow, B., 1990. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proceedings of the International Joint Conference on Neural Networks*, 3, pp. 21–26. June 17–21, San Diego, CA.
- Rubinstein, R.Y., Kroese, D.P., 2008. *Simulation and the Monte Carlo Method*. Hoboken, NJ: John Wiley & Sons.
- Somol, P., Pudil, P., 2000. Oscillating search algorithms for feature selection. *Proceedings of the 15<sup>th</sup> International Conference on Pattern Recognition*, pp. 406–409. September 3–8, Barcelona (Spain).
- Wegner, S., Hooputra, H., Zhou, F., 2006. Modeling of Self-Piercing Rivets Using Fasteners in Crash Analysis, *Proceedings of the 2006 ABAQUS Users' Conference*, pp. 511–526. May 23–25, Cambridge, MA, USA.
- Witten, I.H., Frank, E., 2005. *Data Mining*. San Francisco, CA: Morgan Kaufmann Publishers.

## AUTHORS' BIOGRAPHIES

**Andreas Kuhn** holds master's degrees in Technical Mathematics and Mechanical Engineering from Vienna University of Technology, where he also obtained his Ph.D. for work on the application of numerical simulation techniques to issues of satellite dynamics. After several years of experience in Computer Aided Engineering in the German automotive industry he founded the company ANDATA in 2004, where he currently holds the position of CTO. He is specialized in the application of Computational Intelligence and stochastic simulation techniques to the development and safeguarding of complex technical systems.

**Toni Palau** obtained his master's degree in Applied Mathematics at the Technical University of Catalonia (UPC-Barcelona Tech), Spain, and has occupied diverse technology related positions, covering the topics of Computer Vision, High Performance Computing and Parallelism, Optimisation, Monte Carlo Simulation, Machine Learning, Data Mining, Computational Intelligence and Finite Element Simulation. At present he works as a research engineer at ANDATA.

**Gerolf Schlager** studied Technical Physics at Vienna University of Technology. He was a doctoral student at the European Organization for Nuclear Research (CERN) in Geneva for more than four years and received his Ph.D. for his studies of the energy calibration of the ATLAS calorimeter system. At CERN he was introduced to stochastic simulation of particle collisions and the associated data mining techniques. Since 2007 he has been one of the principal research engineers at ANDATA on the topic of material modelling with Soft Computing approaches.

**Helmut J. Böhm** holds master's degrees in physics from Johannes-Kepler University in Linz, Austria, and Rutgers University, New Brunswick, NJ. He obtained his doctoral degree in Mechanical Engineering and his habilitation in Micromechanics of Materials at Vienna University of Technology, Austria, where he headed the Christian Doppler Laboratory for Functionally Oriented Materials Design from 1998 to 2004. At present he is associate professor at the Institute of Lightweight Design and Structural Biomechanics of Vienna University of Technology.

**Sergio Nogales** holds bachelor and master's degrees in Aeronautical and Materials Engineering from Technical University of Madrid (UPM), Spain. He obtained his doctoral degree at Vienna University of Technology, Austria. At present he is a post-doctoral researcher at the Institute of Lightweight Design and Structural Biomechanics of Vienna University of Technology.

**Victor Oancea** and **Ritwick Roy** both hold Ph.D. degrees and are Principal Development Engineers in the ABAQUS/Explicit development group at SIMULIA.

**Andrea Rauh** graduated in Mechanical Engineering from TU München, Munich, Germany. She joined BMW in 2009 and is currently working in the field of CAE method development for crash simulation with the focus on composite structures and joining technologies.

**Jürgen Lescheticky** studied Aerospace Engineering at the University of Stuttgart, Germany. After nine years of working as an engineering consultant, he joined BMW in 1995 as a project engineer for crash simulation. Since 1999, he has been head of the crash simulation department with the main responsibility for CAE method development.



# Multi-Objective Optimization in Urban Design

Michele Bruno<sup>(a)</sup>, Kerri Henderson<sup>(b)</sup>, Hong Min Kim<sup>(c)</sup>

<sup>(a)</sup> Columbia University 1172 Amsterdam Ave. New York, New York, USA, 10027

<sup>(b)</sup> Columbia University 1172 Amsterdam Ave. New York, New York, USA, 10027

<sup>(c)</sup> Columbia University 1172 Amsterdam Ave. New York, New York, USA, 10027

<sup>(a)</sup> [mb3408@columbia.edu](mailto:mb3408@columbia.edu), <sup>(b)</sup> [kh2388@columbia.edu](mailto:kh2388@columbia.edu), <sup>(c)</sup> [hk2601@columbia.edu](mailto:hk2601@columbia.edu)

## ABSTRACT

Urban Design is a multi-objective task. Traditionally, urban spaces are designed hierarchically; organizational inputs are idealized uniquely, and negotiated through sequential overlay. In our investigation, parametric modeling (with the software application Catia) and evolutionary optimization employing genetic algorithms (with the software application Mode Frontier) enable the exploration of a non-linear design space whereby multiple objectives may be optimized concurrently. This paper describes an experiment that builds from prior research in multi-objective optimization of architectural design and applies that workflow to multi-objective optimization in urban design. The experiment employs given constraints, custom procedural algorithms and genetic algorithms to examine a wide design space and identify designs that perform well in multiple arenas. Design, data and latent influences are exposed and negotiated quantitatively to render topological variation through optimization. By using multi-objective optimization we define and apply quantitative metrics in order to examine the potential for a new workflow in urban design.

Keywords: Optimization, Catia, Mode Frontier

## 1. INTRODUCTION

The context of our research elaborates on the recent work of David Benjamin and Ian Keogh in Multi-Objective Optimization in Architectural Design (Keough and Benjamin 2010). Benjamin and Keough created an automated workflow that linked parametric modeling (*Catia*), structural analysis through a custom-designed software (*Catbot*) and a multi-objective optimization engine (*Mode Frontier*). Their workflow was used to compute multiple architectural design permutations and aid as a tool to evaluate those designs. Building on their investigations and gained knowledge, we hope to broaden the potential influence of this workflow to include urban design. Our research does not integrate the structural analysis loop but uses parametric software and a multi-objective optimization at the scale of the city. Our experiment abstracts buildings to basic geometric primitives such as cylinders in order to study programmatic and spatial relationships. We outline the importance of the

definition of metrics for the success of the experiment and discuss how explicit metrics could influence current practices in urban design.

Urban design concerns the arrangement, design and functionality of cities. The discipline traverses many fields and interests such as architecture, urban planning, construction, politics, economics, real estate development, environmental systems and social theory. In some cases, urban design is influenced disproportionately by one or more of these fields, or by a particular stakeholder. In other cases, early decisions may have a much stronger influence than later decisions. In yet other cases, decisions may be made to satisfy each objective in sequence, which rules out some possible design results. As an alternative to these examples, we propose a workflow of multi-objective optimization in which many design criteria are evaluated simultaneously, with relatively equal influence.

*Optimization* software computes a parametric model through its range of possible permutations to find a set of high-performance designs. The application of this technique is novel in the context of urban design. In engineering, architecture and product design, optimization is often tied to simulation software such as finite element analysis (Kicinger et al. 2005). Inputs are identifiable and quantifiable; permitted tolerances are determined specific to the project, or are taken from known rules of thumb. Using primarily known materials, practices and tolerances for inputs, the workflow often produces designs that are both novel and high performance (Koza et al. 2003).

Often in urban design, projects are developed hierarchically; organizational inputs are idealized uniquely, and negotiated through sequential overlay. Complex problems in urban design may present multiple primary design factors to multiple invested parties (Galster et al. 2001). The strength of the computational process is the software's ability to evaluate multiple objectives concurrently and render a range of high-performance designs.

There are many quantitative factors to be considered in the urban design process. Zoning; program; density; solar gain; shadow projections; wind velocity, location to city service points for energy, water, and waste collection; traffic flow and projected

economic revenue are just a few of the factors involved in the process. Furthermore, there are often qualitative factors that are addressed in urban design; they can include quality of life, cultural distinction and aesthetics. These qualitative factors require metrics for design and critical evaluation. Urban design lays the foundation for the new buildings, public spaces and services that shape our lives.

New technologies enable new workflows to address the complexity of urban design projects. Automated genetic algorithms have been used to exploit parametric permutations by generating, evaluating and improving the performance of possible design options (Keller 2006). This workflow is not geared toward a specific task; it is a tool to aid reflective, responsible design practice.

## 2. WORKFLOW

Our workflow begins with a set of constraints, generates design permutations through custom procedural algorithms and evolves high-performing designs through genetic algorithms. Beginning with design constraints is a familiar launching point for architects and urban designers. Given constraints can include the site, existing infrastructures, services, budgets and legal parameters. Identifying and drawing the given constraints create the initial environment in which to operate.

Inputs are identified through conversation with involved parties. An input is any quantifiable factor, specified by an acceptable range that would support a desired state or objective. For example: Input: building height range 50'–75'. Ideal building height = 75'. Objective: maximize building height. Once given constraints and inputs have been identified, they may be drawn or modeled digitally using parametric software (*Catia*). The custom procedural algorithm is the *Catia* script. The architect or urban designer writes this script. In doing so, he or she sets up the relationships between the inputs and the parameters that can affect their values. The designer may also set up rules to further articulate relationships between design parameters (example: when x is 2, y is 0.5x). The role of the designer is to identify and create the morphological identity of the inputs. Using the custom procedural script, the designer builds the domain of influence of the genetic algorithm, setting the breadth of the potential design space (Figure 1).

The design of a good experiment establishes clear design metrics, bases input parameters upon valid data, is procedural in its modeling techniques and enables the genetic algorithm to explore a wide design space through the custom procedural script.

Connected to scripting is the notion of State Change. State Change is a function of an If/Then condition. That is to say that if x is true, proceed with State A, if y is true, proceed with State B, etc. The State can effect morphological or topological changes in the design. State Change widens design space in that it enables the algorithm to explore possible relationships

and design permutations that would not necessarily occur to the independent designer.

Given Information - Existing conditions  
Custom Procedural Algorithm -*Catia* scripting  
Genetic Algorithm - *Mode Frontier*

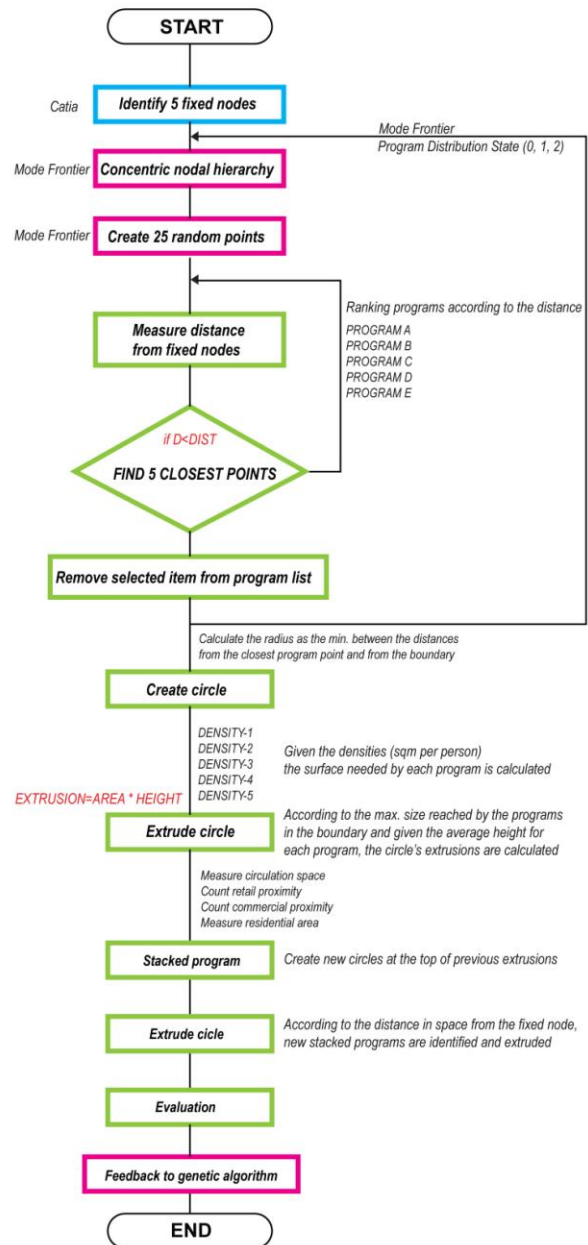


Figure 1. Workflow diagram specifying each agent's role in the process.

Once the model, parametric relationships and script are set, the *Catia* file is linked to the *optimization* software, *Mode Frontier*. *Mode Frontier* is an evolutionary computational software that employs a genetic algorithm as a search heuristic to generate multiple design permutations. It differs from stochastic search in that it learns from ratings of previous permutations. Specifically, our experiment employed the MOGA-II. MOGA-II is a scheduler based on a

multi-objective genetic algorithm (MOGA) designed for fast Pareto convergence. MOGA II supports directional crossover, implements elitism, enforces user-defined constraints and allows Steady State evolution. *Mode Frontier* starts initially with a random population of input parameters. Through generational growth, it evaluates the results of each cycle to ultimately reach a set of designs that offer the best possible outputs for the objectives. For each permutation, *Mode Frontier* generates data sets that the designer can then evaluate and compare with other permutations of the experiment.

Within *Mode Frontier*, objectives are set for the experiment. An objective is a value to which the outputs should be optimized. It can be set to “minimize” or “maximize” global parameters, or at specific target. The objectives articulate the purpose of the experiment. When defining metrics for the objectives, it is possible to create conditions that produce competing objectives. Competing objectives are ones where the conditions champion the maximum performance of one objective and diminish the performance of another. Establishing the objectives involves design decisions that are as crucial as the design of the inputs.

In setting up the experiment, it is the role of the designer to clearly construct the metrics by which the inputs are evaluated. In the case of urban design, objectives can be based on known rules of thumb. Objectives may also be developed as a way of quantifying less mathematical inputs such as quality of life or aesthetics. Often both types of metrics play a role in the experiment (DeLanda 2002). Defining metrics to evaluate design creates a new workflow and design culture in many ways:

1. Each design must begin with the question: what are the necessary inputs for urban design? What does it take to plan a great city?
2. It challenges the architect or designer to set a range of acceptable parameters for each possible case, identifying and expanding the definition of what makes a “good” design.
3. The explicit definition of metrics lessens the importance of subjective preconceptions in the design process. Once rules are established, design evaluation can be more critical and thorough. This novel approach could identify high-performance designs that reach beyond established practices.
4. By re-programming design methodology, this new workflow opens up conversation with representative stakeholders, designers, engineers, investors and community members early on.

To adequately address the numerous variables involved in urban design, the initial setup of experiment is paramount.

For our experiment, we decided to focus on five inputs that we believe to be influential for the urban design of Masdar, UAE: program, density, proximity and mixed-use quality.

In terms of computational resources, we have found that with a PC computer (Intel® Core™ Quad CPU, 4GB RAM) running for 24 hours, with 5 inputs, we can evolve 1,500 design permutations. Increased inputs, model complexity and wider parameter ranges could warrant longer computation times, networked processing or the organization of multiple experiments.

### 3. MASDAR 2.0 (BETA)

Masdar is a new city being constructed 17 km east-southeast of Abu Dhabi. Masdar aims to be a highly efficient, sustainable, zero-carbon, zero-waste ecology development, relying entirely on solar energy and renewable resources (Adrian Smith and Gordon Gill Architecture 2010). Beginning Tabula Rasa, the design of the city invites not only questions of efficient operational practices but also an optimistic interrogation into the factors involved in creating a 21st century city. We hope to develop programmatic distribution for Masdar that not only can be evaluated by measurable criteria but also to create a new workflow and design culture.

This case study involves the use of two existing software applications: *Catia* and *Mode Frontier*. The morphology of the experiment is abstracted: points, circles, cylinders and color are used diagrammatically to represent relationships of program, density, proximity and mixed-use quality (Figure 2).

The procedure starts with a set of fixed nodes as given constraints. For this test the fixed nodes are based on the existing transportation system. These nodes are transit stations that have already been planned and constructed in Masdar. The previous master plan also used these hubs as a primary factor of influence.

In *Catia*, as part of the custom procedural algorithm, we generated a field of potential program locations. We used 100 possible locations. The spacing and number of nodes in this field are influential for the overall design output. It is the responsibility of the designer to set these constraints during the experiment.

The genetic algorithm in *Mode Frontier* creates 25 random points at the possible locations. The distance from all the points to all the nodes is measured through the custom procedural algorithm, and points are ranked in accordance with their node proximity. A list is generated for each node of its sequentially adjacent five points.

For the identified five programs or inputs, a nodal hierarchy is established by the designer (Figure 3). This determines which program should be placed closest to its node and thereon. This hierarchy can determine the conceptual base for the city. For our case study we defined three nodal hierarchies that were of interest to us.

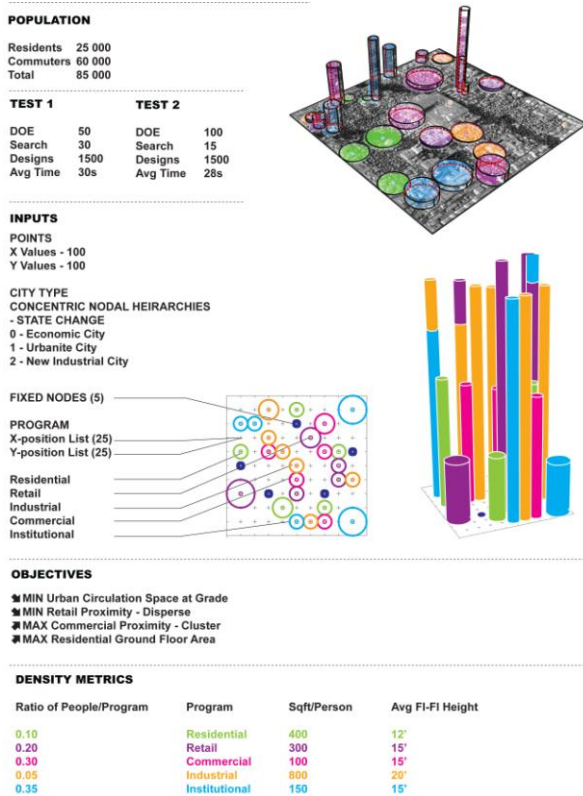


Figure 2. General setup of the experiment.

In *Mode Frontier*, State Change was employed to determine the nodal hierarchy of the given fixed point. This allowed the genetic algorithm to create a variety of urban spaces, widening the design space and enabling potentially unforeseen *optimizations* for our given inputs. Ratios of total program area of urban space to inhabitants were established. A set of rules was established for proximities between programs.

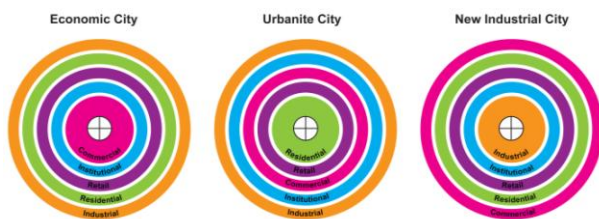


Figure 3. Nodal hierarchies determine the programmatic ordering sequence of geometry created during the experiment.

The proportion of programs within the same urban type is defined by stacking the new programs onto the previously extruded iteration. The hierarchy of programs is consistently valued according to the vertical distance from the transportation hubs (fixed nodes). This looping of programmatic distribution affects the degree of mixed-use program within a building and its neighborhood.

Open space is defined in the experiment as the space at grade that is not assigned to a building program. Open space includes public space, green

space, right of way and all circulation space for vehicles and pedestrians.

Program and circulation space are simultaneously and jointly optimized. The custom procedural algorithm also aims to embrace a planning model that overcomes traditional 20th-century zoning. Taking the perspective of developers, neighborhoods are classified according to adaptability of program combinations to height and potential for economic development.

Following the algorithmic computation, *Mode Frontier* returns data on each permutation's inputs, outputs and objectives. This data can be used in *Mode Frontier* to generate 4D bubble graphs and data charts. Using both visual (4D bubble graphs) and numerical (data charts) data, the designer can look for trends and high-performance results within an iterative process. An Utopia Point is the point on the graph where all objectives would be idealized. A Pareto curve is the set of all best designs. The experiment plays competing objectives against one another. It may not be possible to idealize every objective, but rather to establish a range of designs that achieve high threshold of performance for multiple objectives. Here, the designer re-enters the design process to evaluate influence of the urban design factors and weigh the best designs. Unlike multi-objective *optimization* in engineering fields, we aim to produce a range of high-performance designs that may be further evaluated post-computation by designers. This strives to champion the best possible design for the specific situation.

Our experiment returned a set of 62 best designs (Pareto designs). Each one achieved high performance for one or more objectives. Our initial objectives were to cluster commercial properties, disperse retail, maximize the ground floor area of residential properties and minimize the overall circulation space of Masdar at grade (Figure 4).

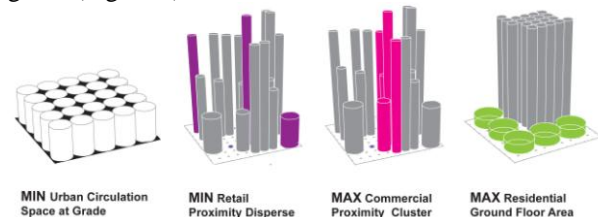


Figure 4. Predictions for each objective individually considered.

Upon examining the Pareto designs we found that three best-fit categories emerged: Best Clustering, Best Area Coverage and Best Overall. For Best Clustering, Design 1493 exhibited the closest proximity values of commercial properties and the best dispersement of retail properties specified in our objectives. The clumping of commercial and dispersement of retail created two identifiable business districts, though it performed less well in overall site coverage. For the Best Coverage category, Design 1244 covered more total area than any other design and showed the maximum ground coverage of residential program. Best



Clustering and Best Coverage are competing objectives; it would be impossible to optimize 100 percent for both. Our ambition was to explore designs that performed as best as possible for these categories.

The Best Overall permutation, Design 1177, performed well in clustering. In this case, two dense commercial areas are apparent, retail is reasonably dispersed, and there is a high presence of residential program at grade, and a high total coverage of land area, as prescribed by our initial input of factors. Design 1177 was chosen as the best-fit design of the experiment (Figures 5, 6 and 7).

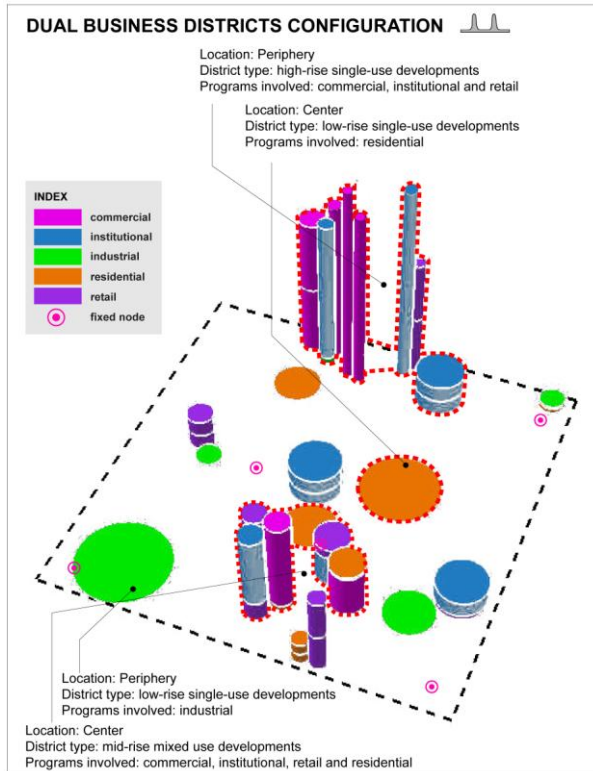


Figure 5. Design 1493: closer analysis.

#### 4. NEXT STEPS

Masdar 2.0 (beta) illustrates our initial set of experiments with programmatic *optimization* in urban design. We would like to extend this experiment to include different inputs within an *optimization* workflow. As part of our future work, we can consider additional types of inputs such as building, block and city morphology, circulation systems, city services, land value and potential economic *optimizations* as they relate to urban form and social inputs. Each type of input will require the definition of a metric by which they will be evaluated. Each type of input can be tested to compare how the input will perform in an environment of multi-objective *optimization*. This process is ideal if automated-testing returns results that are otherwise unattainable by traditional processes. It is possible that certain types of inputs are more suited to this process than others. We would like to test not only how different types of inputs can be optimized, but also how each type could perform in a multi-type, multi-objective experiment.

In our research, we discovered two potential limitations to *optimization*. The first is that the design of a good experiment is crucial. The design inputs must be valid. The parametric model and custom procedural script must be designed to enable the genetic algorithm to explore a wide design space. The experiment must also be designed to methodically and realistically return convincing results.

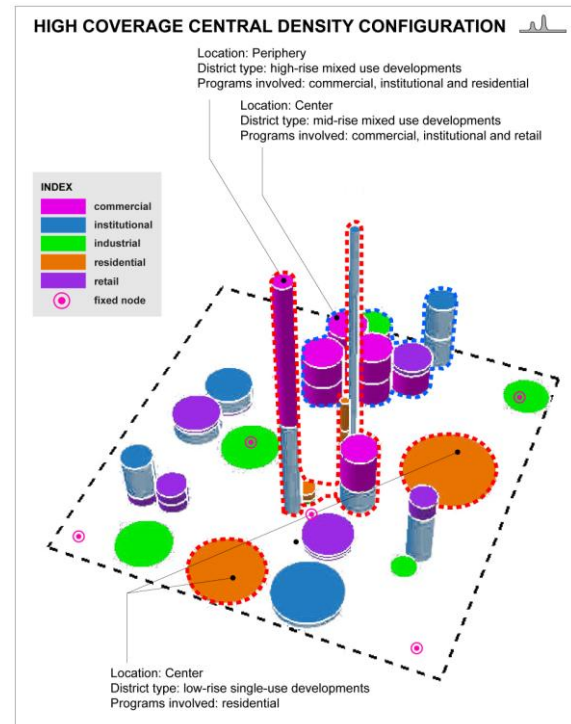


Figure 6. Design 1244: closer analysis.

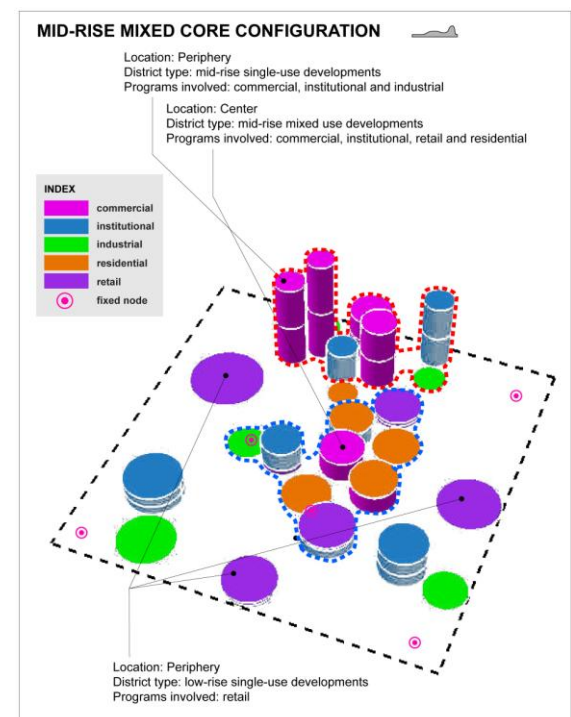


Figure 7. Design 1177: closer analysis.

The second concerns computational power. Succinctness in modeling and in defining parametric relationships can help control the computational needs of the experiment. Exponentially more computational power is needed as accuracy increases, approaching reality.

We recognize that urban design is viewed through many different lenses and must perform according to various criteria. The type of input or comparison of types of inputs can frame the scope of the experiment. It is important to be able to create the ability to test for as many different factors as are involved in urban design so that the conversation surrounding the potential design can be inclusive.

## 5. CONCLUSION

Building on previous research, we have adapted, applied and automated an existing workflow for *optimization* in architecture to urban design. Using the experiment's unique given constraints, we designed a specific parametric model, custom procedural algorithm and *optimization* objectives to create a new type of design experiment and broaden the potential influence a multi-objective workflow to other fields. Our workflow was executed on typical hardware (PC computers) using existing software without any previous advance knowledge of scripting, Engineering Knowledge Language or the function of genetic algorithms. The workflow requires the experiment itself to be well designed in order to be an advantageous tool. Depending on the number of inputs, input ranges and influences, complexity of the model, possible States, and overall scope, the experiment will require a specific design. The authorship of the design workflow is of great importance; it frames not only the parameters of the experiment but also the design project itself. The goal is to design a design space large enough to compute and return design possibilities that would otherwise be impossible to come to independently and to control the scope of the experiment so that it is possible to compute with typical hardware. The design of multiple experiments that build on one another is possible and interesting. The author must understand how factors relate to one another within a given experiment and throughout possibly multiple experiments. The experiment is only as good as the data and the design of the experiment itself.

The use of a functionalist algorithm may contribute to the development of a 21<sup>st</sup> century aesthetic. This new methodology does not rest uniquely on morphological output of data. The designer can define the metrics by which the data is evaluated, how that data is expressed and the possible ranges and potential interactions across the design space. The choice is to control more rigorously the factors that contribute to design practice and imagine new possibilities in design and workflow.

Automated *optimization* processes do not produce a single best design but a range of high-performance designs. The output data of the experiment must be

evaluated and judged by the parties involved. High-performance results that are surprising tend to expose latent assumptions embedded in design culture. In this exploration we seek to open a conversation about what makes a great urban design within the context of measurable factors. A more collaborative, accountable and quantifiable methodology in urban design will change the way developments, neighborhoods and cities are built. We believe that by bringing explicit factors of urban design to the table we can rigorously discuss why a particular design or approach may be favored over another. By openly discussing design priorities we can learn how to create more ideal places to live, work and play.

## Acknowledgments

This work was conducted under the guidance of David Benjamin with the help of Teaching Assistants Jesse Blankenship and Danil Nagy at Columbia University, School of Architecture, Planning and Preservation.

## References

- KEOUGH, I., AND BENJAMIN, D. 2010. *Multi-Objective Optimization in Architectural Design*. Symposium on Simulation for Architecture and Urban Design, 1–9.
- KICINGER, R., ARCISZEWSKI, T. AND DEJONG, K. 2005. *Evolutionary Design of Steel Structures in Tall Buildings*. Journal of Computing in Civil Engineering, vol. 19, no. 3 (July): 223–238.
- KOZA, J.R., KEANE, M.A. AND STREETER, M.J. 2003. *Evolving Inventions*. Scientific American Magazine, February: 52–58.
- GALSTER, G., ET AL. 2001. *Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept*. Housing Policy Debate, vol. 12, issue 4: 1–37.
- KELLER, S. 2006. *Fenland Tech: Architectural Science in Postwar Cambridge*. Grey Room, vol. 23, Spring: 40–65.
- DELANDA, M. 2002. *Deleuze and The Use of Genetic Algorithms in Architecture*. In Rahim, Ali (ed.). Contemporary Techniques in Architecture. London: Wiley-Academy, 9–13.
- ADRIAN SMITH AND GORDON GILL ARCHITECTURE. 2010. Unpublished presentation on Masdar. Columbia University, Graduate School of Architecture, Planning and Preservation, October 27.

## AUTHORS BIOGRAPHY

Michele Bruno (MSAAD), Kerri Henderson (MArch) and Hong Min Kim (MSAAD) have just graduated from the Graduate School of Architecture, Planning and Preservation of Columbia University. Michele is concurrently a PhD candidate '0at the University of Pisa where he graduated previously with a Bachelors degree in Architecture and Engineering. Kerri is a graduate of the University of Waterloo, Canada where she earned an Honours Bachelor of Architectural Studies in 2006. Hong Min Kim has a Bachelor of Science in Architectural Engineering from Yonsei University, Korea. Michele, Kerri and Hong Min, joined forces at GSAPP to develop their shared interests in computation and optimization in architectural and urban design; they have become close friends.

# ADDITIVE FAULT TOLERANT CONTROL FOR DELAYED SYSTEM

*Nouceyba Abdelkrim*<sup>(1,3)</sup>, *Adel Tellili*<sup>(1,2)</sup> and *Mohamed Naceur Abdelkrim*<sup>(1,3)</sup>

<sup>1</sup>Université de Gabès, Unité de recherche Modélisation, Analyse et Commande des Systèmes (MACS)

<sup>2</sup>Institut Supérieur des Systèmes Industriels de Gabès (ISSIG)

<sup>3</sup>Ecole Nationale d'Ingénieurs de Gabs (ENIG)

e-mail: nouceyba.naceur@laposte.net

e-mail: Adel\_Tellili@Lycos.com

e-mail: naceur.abdelkrim@enig.rnu.tn

## ABSTRACT

The additive fault tolerant control (FTC) for delayed system is proposed in this paper. To design the additive control, two steps are necessary, the first one is the estimation of the sensor fault amplitude which is realized by using the Luenberger observer and the second one is the addition of the additive fault tolerant control law to the delayed system nominal control. The nominal control law of delayed system is designed by using the Lambert W method.

**Index Terms**— Additive FTC, delayed system, Lambert W method, Luenberger observer, sensor faults.

## 1. INTRODUCTION

Time delay phenomena appear naturally in the modeling of many systems. Actually, we remark that the majority of industrial system controls are implemented via a numerical calculator, so if a system haven't an intrinsic time delay, often a time delay appear via the control loop[1], [2]. Time delay systems can be represented by delay differential equations (DDEs), which belong to the class of functional differential equations (FDEs), and have been extensively studied in [3], [4], [5] and [6]. During recent decades, the Lambert W function has been used to develop an approach for the solution of linear time-invariant (LTI) systems of DDEs with a single delay [7], [8]. The Lambert W approach is used to resolve many problem such that stability [9], [10], design by state feedback [11].

Moreover to the presence of time delay, many type of fault can affect the delayed system such as sensor fault, actuator fault and system fault. So it's necessary to design a control able to tolerating potential faults in these systems in order to improve the reliability and availability while providing a desirable performance [12]. This field control is known as fault-tolerant control systems ; it can maintain overall system stability and desired performance in the event of such failures [13-15].

This work deals with the design of an additive FTC developed in [13, 14] as extension to delayed system where its nominal control is designed by using the Lambert W approach.

## 2. DESIGN OF THE NOMINAL CONTROL OF DELAYED SYSTEM BY USING THE MLW METHOD

In this section we construct the nominal control of delayed system by using the Lambert W method.

### 2.1. Definition of Lambert W function

Let's  $x \in \mathbb{C}$  be a solution, to determinate, of the equation  $xe^x = y$  for  $y \in \mathbb{C}$ , This type of equation can be resolve by using the Lambert W function  $W_k$  such that [11,16,17]  $x = W_k(y)$ .

With:

$k$  branches of Lambert W function and  $k \in ]-\infty, +\infty[$ ,  $k = 0$  principal branch [11,16,17].

Consider the delayed system:

$$\dot{x}(t) = Ax(t) + A_d x(t-d) \quad (1)$$

where  $A$  and  $A_d$  are  $n \times n$  matrices,  $x(t)$  is an  $n \times 1$  state vector and  $d$  is a constant time delay. this can be represented by the delayed differential equation (DDE):

$$\dot{x}(t) - Ax(t) - A_d x(t-d) = 0 \quad (2)$$

Let's  $x(t) = e^{St}x_0$  be a solution of (2) with  $S$ : matrix, with appropriate dimension, to determinate. Replacing the expression of  $x(t)$  in (2) we find:

$$(S - A - A_d e^{S(-d)})e^{St}x_0 = 0 \quad (3)$$

by consequence we find:

$$(S - A - A_d e^{S(-d)}) = 0 \quad (4)$$

then:

$$S - A = A_d e^{-dS} \quad (5)$$

multiplying the equation (5) by  $(d e^{dS} e^{-dA})$  we find:

$$d(S - A)(e^{(S-A)d}) = A_d d e^{-A d} \quad (6)$$

In this step we can use Lambert W function with  $x = d(S - A)$  and  $y = A_d d e^{-A d}$ , by consequence the expression of  $S$  is:

$$S = \frac{1}{d} W(A_d d e^{-A d}) + A \quad (7)$$



equation (7) represent the characteristic equation for the general time-delay system [9]. The roots of equation (2) are the eigenvalues of the matrix  $S$  and can be used to describe the stability of the DDE (2), which represent the stability of general system.

## 2.2. Determination of the state feedback gain of delayed system

For a scalar DDE with state feedback as shown Eq.(8)

$$\begin{cases} \dot{x}(t) = ax(t) + a_d x(t-d) + bu(t) \\ u(t) = kx(t) \end{cases} \quad (8)$$

With  $k$  is the feedback gain determinate by using the Lambert W method The Lambert-W function can assign the real part of the rightmost pole exactly, by using this relation:

$$\text{Re}(S_0 = \frac{1}{d}W_0(a_d d e^{-(a+bk)d}) + (a+bk)) = \lambda_{desired} \quad (9)$$

For example in [11], for the system (8) with  $a = 1$ ,  $a_d = -1$ ,  $b = 1$  and  $d = 1$ ,

$$\text{Re}(S_0 = \frac{1}{d}W_0(a_d d e^{-(a+bk)d}) + a+bk) = -1 \quad (10)$$

Then, the resulting value of  $k$  is  $-3.5978$ .

After the design of the nominal control of delayed system we will determinate the additive control which can compensate the sensor fault.

## 3. DESIGN OF THE ADDITIVE CONTROL

The fault accommodation is based on the addition of an additive control  $u_{ad}$  to the nominal control law  $u$ . To design the additive control, two steps are necessary, the first one is the estimation of the sensor fault amplitude and the second one is the addition of the additive fault tolerant control law to the nominal control of delayed system.

### 3.1. Sensor fault amplitude estimation

When a sensor fault affects the closed loop system the tracking error between the reference input and the measurement will no longer be equal to zero. In this case, the nominal control law tries to bring the steady-state error back to zero. Hence, in the presence of sensor fault, the control law must be prevented from reacting. This can be achieved by cancelling the fault effect on the control input [13, 14].

A Luenberger observer [1], [18] will be used to estimate the sensor fault amplitude which affects the delayed system. Consider the linear time invariant LTI delayed system:

$$\begin{cases} \dot{x}(t) = Ax(t) + A_d x(t-d) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad (11)$$

If the number of outputs is greater than the number of control inputs, the designer of the control law selects the outputs that must be tracked and breaks down the output vector  $y(t)$  as follow:

$$y(t) = Cx(t) = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} x(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} \quad (12)$$

The feedback controller is required to cause the output vector  $y_1(t) \in \mathbb{R}^p (p \leq r)$  to track the reference input vector  $y_r$  such that in steady-state:

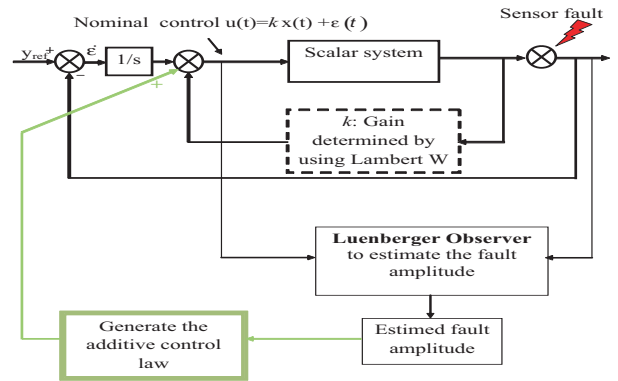
$$y_r(t) - y_1(t) = 0 \quad (13)$$

To achieve this objective, an integrator vector  $\varepsilon(t)$  is added which satisfy the following relation:

$$\dot{\varepsilon}(t) = y_r(t) - y_1(t) \quad (14)$$

With:  $\varepsilon(t) = \int (y_r(t) - y_1(t)) dt$

figure (1) represent the adding of integrator and the diagram of estimation and compensation of sensor fault.



**Figure 1.** Diagram of additive fault tolerant control.

So the nominal control represented by (8) becomes:

$$u(t) = kx(t) + \varepsilon(t) \quad (15)$$

The closed loop system becomes:

$$\begin{cases} \dot{X}(t) = A_f X(t) + A_{df} x(t-d) + G y_r(t) \\ y(t) = C_f X(t) \end{cases} \quad (16)$$

with:

$$\begin{cases} X(t) = \begin{bmatrix} x(t) \\ \varepsilon(t) \end{bmatrix}, A_f = \begin{bmatrix} (A+Bk) & B \\ -C & 0 \end{bmatrix}, \\ A_{df} = \begin{bmatrix} A_d \\ 0 \end{bmatrix} = \begin{bmatrix} A_{df1} \\ A_{df2} \end{bmatrix}, \\ C_f = [C \ 0], G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{cases}$$

For sensor faults, the output equation given in (11) is broken down according to (12), and can be written as:

$$y(t) = Cx(t) + Ff(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} \quad (17)$$

If the delayed system is observable the Luenberger observer can be written as follow:

$$\begin{aligned}\dot{\hat{X}}(t) &= A_f \hat{X}(t) + A_{df} \hat{x}(t-d) + G y_r(t) \\ &+ L(y(t) - \hat{y}(t)) \\ \hat{y}(t) &= C_f \hat{X}(t)\end{aligned}\quad (18)$$

Where:  $L$ : observer gain,  $\hat{y}(t)$ ,  $\hat{x}(t-d)$  and  $\hat{X}(t)$  are respectively estimated output, delayed estimated state vector and augmented estimated state vector. Then we define the observation error  $e(t) = X(t) - \hat{X}(t)$ , so the error dynamic of this relation is governed by this equation:

$$\begin{aligned}\dot{e}(t) &= \dot{X}(t) - \dot{\hat{X}}(t) \\ &= A_f X(t) + A_{df} x(t-d) + G y_r(t) \\ &- [A_f \hat{X}(t) + A_{df} \hat{x}(t-d) + G y_r(t) \\ &+ L(y(t) - \hat{y}(t))] \\ \dot{e}(t) &= M e(t) + A_{df} \tilde{x}(t-d) - L F f(t)\end{aligned}\quad (19)$$

with:

$$\begin{cases} M = A_f - L C_f = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \\ \tilde{x}(t-d) = x(t-d) - \hat{x}(t-d) \end{cases}$$

Knowing that in steady state where  $\dot{e}(t) = 0$ , then we can write this equation:

$$\dot{e}(t) = 0 = M e(t) + A_{df} \tilde{x}(t-d) - L F f(t) \quad (20)$$

And by consequence:

$$L F f(t) = M e(t) + A_{df} \tilde{x}(t-d) \quad (21)$$

In our case  $F$  is scalar and equal to 1, the gain of observer  $L = [L_1 \ L_2]^T$  is designed by using the augmented system, then  $LF$  is not invertible, for this reason we make the following decomposition:

$$L_1 F \hat{f}(t) = M_{11} \tilde{x}(t) + M_{12} \tilde{\varepsilon}(t) + A_{df1} \tilde{x}(t-d) \quad (22)$$

Then the expression of the sensor fault amplitude is given by this equation:

$$\hat{f}(t) = (L_1 F)^{-1} (M_{11} \tilde{x}(t) + M_{12} \tilde{\varepsilon}(t) + A_{df1} \tilde{x}(t-d)) \quad (23)$$

### 3.2. Compensation of sensor fault

The compensation for sensor fault effect on the closed-loop system can be achieved by adding a new control law to the nominal one [13, 14]:

$$u(t) = k x(t) + \varepsilon(t) + u_{ad}(t) \quad (24)$$

The output and the integrator are affected such that:

$$\begin{cases} y(t) = C x(t) + F f(t) \\ \varepsilon(t) = \varepsilon(t) + f_\varepsilon(t) \\ f_\varepsilon(t) = \int (-F f(t)) dt \end{cases} \quad (25)$$

If  $C = 1$ , by using (24) and (25) we can write the control law as follow:

$$u(t) = k x(t) + k F \hat{f}(t) + \varepsilon(t) + f_\varepsilon(t) + u_{ad}(t) \quad (26)$$

The effect of sensor fault, in control and by consequence in system, can be compensated by using the fault amplitude estimation and by calculating the additive control as follow:

$$u_{ad}(t) = -k F \hat{f}(t) - f_\varepsilon(t) \quad (27)$$

## 4. SIMULATION EXAMPLE

### 4.1. Estimation and compensation of sensor fault

Consider the linear time invariant delayed system:

$$\begin{cases} \dot{x}(t) = -x(t) - x(t-d) + u(t) \\ y(t) = x(t) \end{cases} \quad (28)$$

With:  $d = 1s$ .

The state feedback control determined by using the relation (9) leads to the numerical value of gain  $k = -21.86$ , this gain is obtained by placing -3 as desired eigenvalue of  $S$ .

The closed loop-system augmented by integrator becomes:

$$\begin{cases} \dot{X}(t) = \begin{bmatrix} -22.86 & 1 \\ -1 & 0 \end{bmatrix} X(t) + \begin{bmatrix} -1 \\ 0 \end{bmatrix} x(t-d) \\ \quad + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y_r(t) \\ y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} X(t) \end{cases} \quad (29)$$

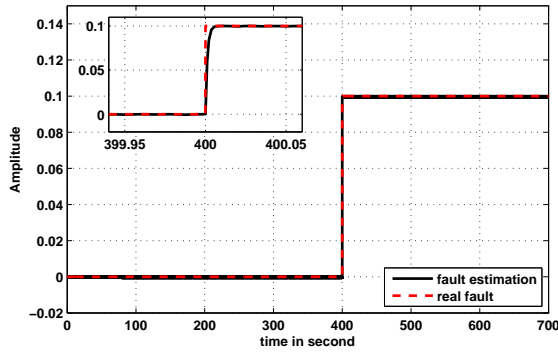
Then the nominal control law take this form:

$$u(t) = -21.86 x(t) + \varepsilon(t) \quad (30)$$

To estimate the sensor fault amplitude we use the Luenberger observer represented by (18). The observer gain  $L$  is determined such that the poles of  $(A_f - LC_f)$  equal 20 times of the ones of  $(A_f)$ . So the numerical value of  $L$  is:  $L = [838 \ 398]^T$ . It's clear that the matrix  $F$  is scalar and the gain  $L$  is a vector, so  $(LF)$  is not invertible, for this reason we use the decomposition represented by equation (22). Let us now examine the influence of sensor fault on the delayed system and the way to compensate for their effect. We will consider the type of sensor fault as a bias. The value of sensor fault is initialized to zero and it's switched to 0.1 at moment  $t = 400s$ , then the output equation of system (28) becomes:

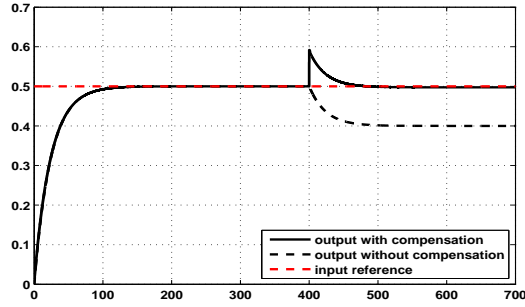
$$y_f(t) = y(t) + 0.1 \quad (31)$$

The input reference is equal to:  $y_r(t) = 0.5$ .



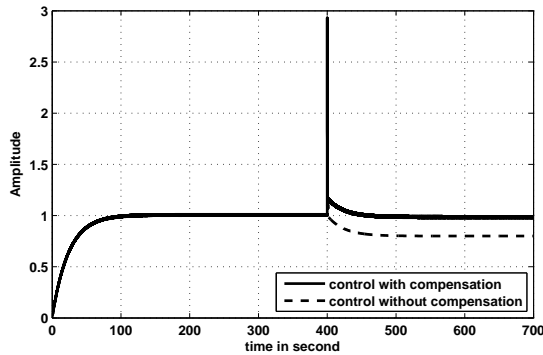
**Figure 2.** Sensor fault estimation.

Figure 2 shows that the observer can estimate the amplitude of sensor fault.



**Figure 3.** Time delay system output with and without fault compensation.

Figure 3 shows the time evolution of the output of delayed system in the occurrence of sensor fault. It's clear, in the curve of the output with compensation, that the additive control added to the nominal control can compensate the effect of sensor fault.



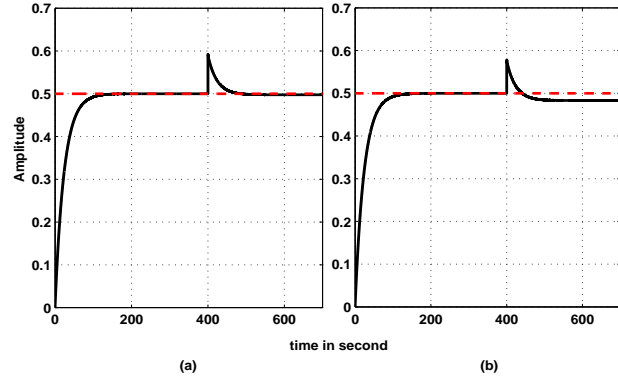
**Figure 4.** Time delay system control with and without fault compensation.

Figure 4 shows that the control law takes its steady value after compensation of sensor fault, but at a fault occurrence time the control law effect an important deviation. This deviation due to the important value of the ob-

server gain and it is relative to the dynamic of same delayed system. The deviation can be accepted if we haven't constraints in the control law.

#### 4.2. Effect of observer gain value to the compensation

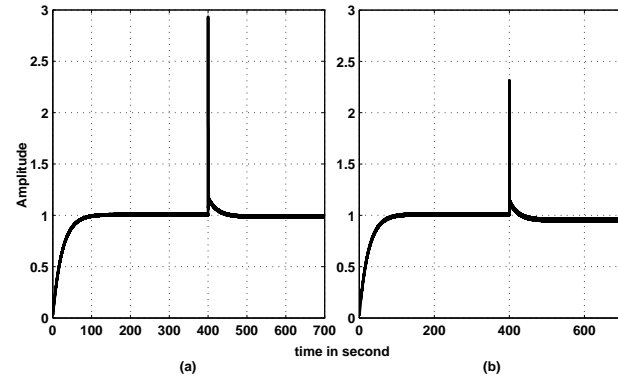
In this section we present the effect of the observer gain value of the fault compensation error.



**Figure 5.** Time delay system output with fault compensation for different value of observer gain.

- a: Case of  $L$  determinate such that the poles value of  $(A_f - LC_f)$  equal 20 times of the ones of  $(A_f)$ .
- b: Case of  $L$  determinate such that the poles value of  $(A_f - LC_f)$  equal 3 times of the ones of  $(A_f)$ .

It's clear that if the value of  $L$  is more important, the error of compensation tend to zero.



**Figure 6.** Time delay system control with fault compensation for different value of observer gain.

- a: Case of  $L$  determinate such that the poles value of  $(A_f - LC_f)$  equal 20 times of the ones of  $(A_f)$ .
- b: Case of  $L$  determinate such that the poles value of  $(A_f - LC_f)$  equal 3 times of the ones of  $(A_f)$ .

In other part, if the value of  $L$  is more important the deviation of control in the occurrence of fault is more important. The optimal case is choose according to the desired performance, for example, if we haven't constraints in the control law we can choose the case (a), but if the

control law had a saturation we must reduce the value of  $L$  by accepting an error of fault compensation, this case we encouraged to study the fault tolerant control with constraints.

## 5. CONCLUSION

In this paper we have treated the fault tolerant control of time-delayed system by using the Luenberger observer to estimate the fault amplitude. An additive control is added to the nominal control law determinate by using Lambert W approach to compensate the effect of fault.

We note that the use of observer provoke a deviation in the control evaluation at time of fault occurrence this we encourage to study the fault tolerant control with constraints.

## 6. REFERENCES

- [1] A. Seuret, Commande et observation des systèmes à retards variables : théorie et applications, thèse de doctorat de l'Ecole Centrale de Lille université des sciences et technologies de Lille, 4 Octobre 2006.
- [2] W. Kacem, Contribution la stabilité et la stabilisation des systèmes à retard. Thèse de doctorat de l'Ecole Nationale d'Ingénieurs de Sfax, 19 Décembre 2009.
- [3] S. Yi, P. W. Nelson, and A. Galip Ulsoy, "Controllability and Observability of Systems of Linear Delay Differential Equations Via the Matrix Lambert W Function, *IEEE Transactions on Automatic Control*, VOL. 53, NO. 3, April 2008.
- [4] H. Gorecki, S. Fuksa, P. Grabowski, and A. Korytowski, *Analysis and Synthesis of Time Delay Systems*, New York: Wiley, p. 369, 1989.
- [5] J. K. Hale and S. M. V. Lunel, *Introduction to Functional Differential Equations*, New York: Springer-Verlag, 1993.
- [6] J. P. Richard, Time-delay systems: An overview of some recent advances and open problems, *Automatica*, vol. 39, pp. 1667-1694, 2003.
- [7] F.M. Asl and A.G. Ulsoy, Analysis of a system of linear delay differential equations, *J. Dyn. Syst. Meas. Control*, vol. 125, pp. 215-223, 2003.
- [8] S. Yi and A. G. Ulsoy, Solution of a system of linear delay differential equations using the matrix Lambert function, in Proc. 25th Amer. Control Conf., Minneapolis, MN, Jun. 2006, pp. 2433-2438.
- [9] P. Kristel and F. Roger, Developing and automating time delay system stability analysis of dynamic systems using the matrix Lambert w (mlw) function method, May 2009.
- [10] S. Yi, P. W. Nelson, and A.G. Ulsoy, Survey on analysis of time delayed systems via the Lambert w function, *advances in dynamical systems*, 14(S2), pp. 296-301, 2007.
- [11] S. Yi, P. W. Nelson, and A.G. Ulsoy, Feedback control via eigenvalues assignment for time delayed systems using the Lambert w function, *Proceedings of the ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2007*, Las Vegas, Nevada, USA, pp:1-10 September 4-7, 2007.
- [12] Y. Zhang, J. Jiang, Bibliographical review on reconfigurable fault-tolerant control systems, *Annual Reviews in Control*, 32 (2008) 229-252.
- [13] H. Noura, Méthode d'accommodation aux défauts: théories et applications. Habilitation à diriger des recherches de l'université Henri Poincaré, Nancy 1, 26 mars 2002.
- [14] H. Noura, D. Theilliol, J.C. Ponsart and A. Chamseddine, Fault-Tolerant Control Systems design and practical application, Industrial Control Center, 2009.
- [15] N. Abdelkrim, A. Tellili and M.N. Abdelkrim, Additive Fault Tolerant Composite State feedback Control of Singularly Perturbed System, *JTEA: 6th, International Conference on Electrical Systems and Automatic Control*, 26-28 March, Hammamet, Tunisia, 2010.
- [16] S. Yi, A.G. Ulsoy and P. W. Nelson, Delay Differential Equations via the Matrix Lambert W Function and Bifurcation Analysis: Application to Machine Tool Chatter, *Mathematical Biosciences and Engineering*, vol. 4, no. 2, pp. 355-368, April. 2007.
- [17] S. Yi and A.G. Ulsoy, Solution of a System of Linear Delay Differential Equations Using the Matrix Lambert Function, *Proceedings of American Control Conference*, pp. 2433-2438, May. 2006.
- [18] D.G. Luenberger, An introduction to observers, *IEEE Trans. on Automatic Control*, 16, 596-602. 1971.

# CENTRALIZED AND DECENTRALIZED ADAPTIVE FAULT-TOLERANT CONTROL APPLIED TO INTERCONNECTED AND NETWORKED CONTROL SYSTEM

Amina CHALLOUF<sup>(a)</sup>, Adel TELLILI<sup>(b)</sup>, Christophe AUBRUN<sup>(c)</sup>, Mohamed Naceur ABDELKRIM<sup>(d)</sup>

<sup>(a,d)</sup> Unité de Recherche: Modélisation, Analyse et Commande des Systèmes (MACS),  
Ecole Nationale d'Ingénieurs de Gabès, Tunisie

<sup>(b)</sup> Institute Supérieure des Etudes Technologiques de Djerba, Tunisie

<sup>(c)</sup> Centre de Recherche en Automatique de Nancy (CRAN),  
Nancy-Université, CNRS UMR 7039

Facultés des Sciences et Techniques, France

<sup>(a)</sup> [amina\\_chalouf@yahoo.fr](mailto:amina_chalouf@yahoo.fr), <sup>(b)</sup> [Adel\\_Tellili@Lycos.com](mailto:Adel_Tellili@Lycos.com), <sup>(c)</sup> [christophe.Aubrun@cran.uhp-nancy.fr](mailto:christophe.Aubrun@cran.uhp-nancy.fr),  
<sup>(d)</sup> [naceur.abdelkrim@enig.rnu.tn](mailto:naceur.abdelkrim@enig.rnu.tn)

## ABSTRACT

In this work, we focus on monitoring and reconfiguration of an Adaptive Model Reference (MRA) Fault-Tolerant Control (FTC) for large-scale system. This particular class presents an interconnected and networked control system (INCS). Moreover, the system can be decomposed into N-interconnected subsystems communicating with network. Then the global output of INCS and one or more outputs of N-interconnected and networked control subsystems are attacked by sensor faults. Therefore, an active Fault-Tolerant (FT) approach, say the model reference adaptive control of linear systems, is used in order to guarantee not only the stability of an overall INCS globally, but also all local stabilities of N-networked control subsystems with strong interactions, delay and additive faults. Moreover, two architectures: centralized and decentralized adaptive controllers are designed to compensate the sensor faults for different internal structures of systems which are subject of this paper. The law adaptations which make the different faulty systems stable are given. A simulation example of an overall INCS consisting of three interconnected and networked control subsystems and involving stabilization of unstable steady-states is used to demonstrate the efficiency of the proposed approach.

Keywords: adaptive control, centralized control, decentralized control, fault-tolerant control (FTC), interconnected system, networked control system (NCS), reference model (RM), sensor failure

## 1. INTRODUCTION

The notion of Fault-Tolerant adaptive control has been an active area of research (Blanke and al 2003, Bodson and al 1997, Patton 1997, Patton and al 1997). The theory of large-scale systems is devoted to the problems that arise from some difficulties (Ioannou and al 1985, Huang and al 2010, Lina and al 2006, Mahmoud 1997, Patton and al 2007):

- Dimensionality;
- Information structure constraints;
- Uncertainty;
- Delays.

A system is considered large-scale if it is necessary to partition the given analysis or synthesis into manageable sub-problems. As a result, the overall system is no longer controlled by a centralized controller but by several independent controllers which all together represent a decentralized controller. This is the fundamental difference between feedback control of small and large systems. On one hand, the development in this direction has reached a level of important applications where adaptive controllers are used to enhance stability and improve operating conditions of defective systems (Bodson and al 1997, Ioannou and al 1985, Mahmoud 1997, Tsai and al 2009). On the other hand, her theory is developed to make a general practical use of adaptive controllers in both large-scale and networked control systems [Ioannou and al 1985, Mahmoud 1997, Patton and al 2007]. In interconnected and networked control systems, there will be more adaptive controllers located at different, possibly distant units, and in control centers. Besides, the dynamic of each local subsystem is not known exactly and the local outputs are corrupted with noise disturbances and faults via network control. In this paper, centralized and decentralized adaptive FTC based on reference model is included in presence of all interactions and over medium of communication between each subsystem. Two structures of adaptive controllers for interconnected and networked system is proposed (Bakule 2008, Chalouf and al 2009, Hasheng 2003, Hovakimyan and al 2006, Sundarapandian 2005). Each adaptive controller is designed to compensate the

additive faulty sensor. The stability of the overall adaptive control scheme is established.

The paper is organized as follows: Section 2 recalls the interconnected and networked control system modernization. Also, it represents the controller synthesis for LTI and continuous time systems. A simulation example of three interconnected and networked control subsystems subject to sensor faults is used in Section 3 to illustrate the effectiveness and performance of the two architecture (centralized and decentralized) active fault tolerant control system. Finally, a conclusion is included.

## 2. PROBLEM STATEMENT

### 2.1. Interconnected System

This class of the large-scale system can be composed of N-interconnected subsystems [Ioannou and al 1985, Mahmoud 1997].

Consider an overall interconnected system which is given by the following state spaces:

$$(S_p) \begin{cases} \dot{X}_p(t) = A_p X_p(t) + B_p U_p(t) + D_p w_p(t) \\ Y_p(t) = C_p X_p(t) + D_p U_p(t) + V_p v_p(t) \\ X_p(0) = \varphi(t) \end{cases} \quad (1)$$

The index p designs the processus parameters.

$A_p, B_p, C_p, D$  and  $V$  are the parameters system with appropriate dimensions.

On the other hand,

$$X_p(t) \in \mathbb{R}^n, U_p(t) \in \mathbb{R}^m, Y_p(t) \in \mathbb{R}^q$$

and  $w_p(t) \in \mathbb{R}^p$  and  $v_p(t) \in \mathbb{R}^p$  are the state vector, the control vector, the output vector and the external disturbance vectors respectively.

As black as this class of the large-scale system contains N-interconnected subsystems and based on equation (1), the partitioned matrix  $X_p, U_p$  and  $Y_p$  are:

$$X_p(t) = [X_{p1}^T, X_{p2}^T, \dots, X_{pN}^T]^T \quad (2)$$

$$U_p(t) = [U_{p1}^T, U_{p2}^T, \dots, U_{pN}^T]^T \quad (3)$$

$$Y_p(t) = [Y_{p1}^T, Y_{p2}^T, \dots, Y_{pN}^T]^T \quad (4)$$

$$w_p(t) = [w_{p1}^T, w_{p2}^T, \dots, w_{pN}^T]^T \quad (5)$$

In addition, the appropriate matrix  $A_p, B_p, C_p, D_p$  and  $V_p$  be partitioned as follows:

$$A_p = \begin{bmatrix} A_{p1} & A_{p12} & \dots & A_{p1N} \\ \vdots & \vdots & & \vdots \\ A_{pN1} & A_{pN2} & \dots & A_{pN} \end{bmatrix} \quad (6)$$

$$B_p = \begin{bmatrix} B_{p1} & B_{p12} & \dots & B_{p1N} \\ \vdots & \vdots & & \vdots \\ B_{pN1} & B_{pN2} & \dots & B_{pN} \end{bmatrix} \quad (7)$$

The matrix  $B$  will be full column rank.

$$C_p = [C_{p1} \quad C_{p2} \quad \dots, \quad C_{pN}] \quad (8)$$

and

$$v_p = [v_{p1} \quad v_{p2} \quad \dots, \quad v_{pN}] \quad (9)$$

Besides, the matrix  $A_p$  can be viewed as an interconnected matrix, too, consisting N-subsystems defined by them appropriate matrices  $(A_{pi}, B_{pi}, C_{pi})$ ,  $i=1, \dots, N$ .

So, we can write:

$$n = \sum_{i=1}^N n_i, m = \sum_{i=1}^N m_i, p = \sum_{i=1}^N p_i \text{ and } q = \sum_{i=1}^N q_i \quad (10)$$

Consequently, the parameters of an overall system (S) depend essentially on the parameters of N-interconnected subsystems ( $S_i$ ) which constituting (Ioannou and al 1985). In point of  $i^{\text{th}}$  subsystem is instable, the global system is also instable.

Like that a global system (S) can regroup N-interconnected subsystems ( $S_i$ ), the equation (1) can be rewritten into matrix form decomposing in all N-interconnected subsystems as follows:

$$(S_i) \begin{cases} \dot{X}_{pi}(s) = A_{pi} X_{pi}(s) + B_{pi} U_{pi}(s) \\ \quad + \sum_{j \neq i} (G_{pij} g_{pij}(s) + D_{pij} w_{pij}(s)) \\ Y_{pi}(s) = C_{pi} X_{pi}(s) + L_{pi} V_{pi}(s) \quad ; i, j = 1, \dots, N. \\ X_i(0) = \varphi_i(t) \end{cases} \quad (11)$$

Where:  $X_{pi} \in \mathbb{R}^{n_i}, U_{pi} \in \mathbb{R}^{m_i}$  and  $Y_{pi} \in \mathbb{R}^{q_i}$  are respectively the local states, the control and the output of the subsystems ( $S_i$ ).

$A_{pi} \in \mathbb{R}^{n_i \times n_i}, B_{pi} \in \mathbb{R}^{n_i \times m_i}$  and  $C_{pi} \in \mathbb{R}^{q_i \times n_i}$  are

respectively the parametric matrix of the local states, the control and the output of the  $i^{\text{th}}$  processus coupled with  $j^{\text{th}}$  subsystems ( $S_j$ ).

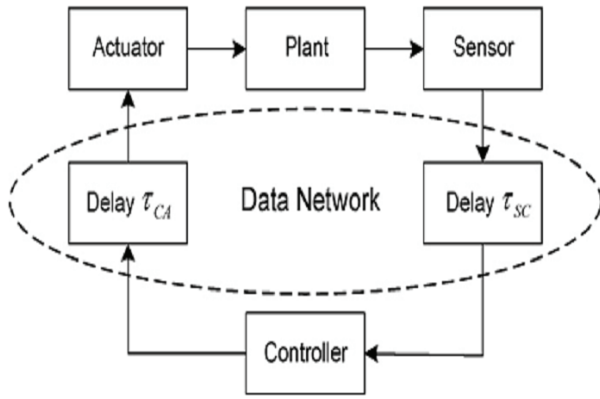


Figure1: Internal Structure of NCS

## 2.2. NCS model

The stage of modeling is rather significant at the time of the study of a process. The systems considered in this work belong to a particular class. They are the systems ordered by a medium of communication, they are known still by the systems ordered in network (NCS) [7,8,9]. This class of the system presents the phenomenon of delay in their dynamic of which its structure is illustrated by the fig.2:

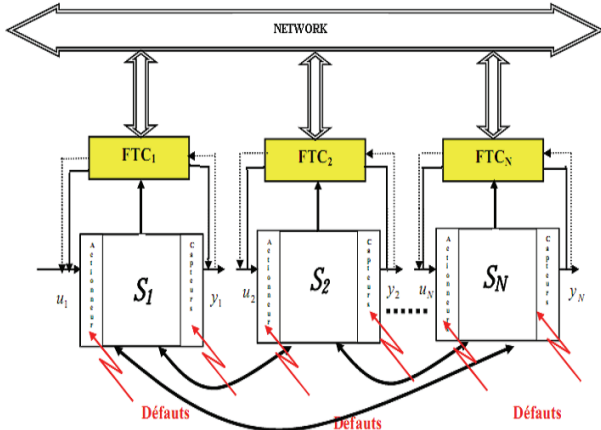


Figure 2: Structure of co-operative Interconnected NCSs and FTC

Contrary to the ordinary systems whose temporal evolutions can be given starting from the value of state  $X$  at the moment present  $\dot{X}(t) = f(t, X_p(t))$ . Those of the systems ordered in network depend primarily on of the last values of the state  $X_p(t)$ . Moreover, the latter present the phenomena of delay and the losses or duplications of the messages during the transmission of the data control/actuator:  $\tau_{CA}$  and sensor/control:  $\tau_{SC}$ .

Generally, three working frameworks are used to represent a late system: models on ring, infinite models of dimension on operators and models in the form of differential equations. In this present work, concerning the stability on the one hand and the modeling of communication TCP on the other hand, we are interested in the effects of the delay in dynamics of the NCS. Plus precisely, we thus consider systems (eq.1) rewritten in the form:

$$(S_p) \begin{cases} X_p(t) = A_p X_p(t) + B_p U_p(t - \tau) + D_p w_p(t) \\ Y_p(t) = C_p X_p(t) + D_p U_p(t - \tau) + V_p v_p(t) \\ X_p(0) = \varphi(t) \end{cases} \quad (12)$$

At the time of the phase of modeling, as well as the matrices defining a model, it is essential to determine the type of delay which affects the system (see fig 1).

There are three principal categories of delay:

- (a) delay in known;
- (b) raised delay;
- (c) limited delay.

The delay can be between control/actuator and/or sensor/control that as (see fig.1):

$$\tau = \tau_{sc} + \tau_{ca} \quad (13)$$

Where:

$$0 \leq \tau_{\min} \leq \tau \leq \tau_{\max} \quad (14)$$

$\tau_{\min}$  and  $\tau_{\max}$  represent respectively the minimum and maximum values of delay. All MIMO transfer matrix representations have appropriate dimensions and are proper real-rational matrices, stabilisable and detectable. A state space rational proper transfer - function is denoted by:

$$\begin{aligned} Sys_p &= \frac{U_p}{Y_p} = \begin{bmatrix} A_p & B_p \\ C_p & D_p \end{bmatrix} \\ &= C_p (sI - A_p)^{-1} B_p + D_p \end{aligned} \quad (15)$$

## 2.3. Model Reference Adaptive FTC design

Consider a LTI system defined by eq.12. To accommodate the last system, we were implementing adaptive FTC. His principal is explained by fig.2 offering the structure of co-operative INCSs and FTC. The last method is based on reference model. It proves the aptitude to reach the characteristics specified by the reference model in presence of additive faults on system [11, 12, 13, 14, 15]. The last idea is illustrated by fig.3. Based on fig.3 and in the time domain, the overall reference model is described by the equation:

$$\dot{Y}_m(t) = A_m Y_m(t) + B_m r(t) \quad (16)$$

The index m designs the known reference model parameters.

In this part, we apply the direct method of the fault tolerant adaptive control. The control input is given by:

$$U_p(t) = (C_0(t)r(t)) + (G_0(t)Y_p(t)). \quad (17)$$

With  $C_0(t)$  and  $G_0(t)$  are the parameters of the adjustable controller.

The derivation of time response plant is given by:

$$\dot{Y}_p(t) = (C_p A_p + C_p B_p G_0) X_p(t) + C_p B_p C_0 r(t) \quad (18)$$



When  $C_0(t) = C_0^*$  and  $G_0(t) = G_0^*$

The motivation being that there exist nominal parameter values of the adaptive controller:

$$\begin{cases} C_0^* = (C_p B_p)^{-1} B_m \\ G_0^* = (C_p B_p)^{-1} (A_m C_p - C_p A_p) \end{cases} \quad (19)$$

The update laws are given by:

$$\dot{\phi} = \begin{bmatrix} \dot{\phi}_r(t) \\ \dot{\phi}_Y(t) \end{bmatrix} = \begin{bmatrix} \dot{C}_0(t) = -g e_0(t) r(t) \\ \dot{G}_0(t) = -g e_0(t) Y(t) \end{bmatrix} \quad (20)$$

Thus, the expression of the adaptive law control becomes:

$$U(t) = ((\phi_r(t) - C_0^*)r(t)) + ((\phi_Y(t) - G_0^*)Y_p(t)) \quad (21)$$

Since, the error dynamic equation between the plant time responses  $Y_p(t)$  and reference model  $Y_m(t)$  with the adjustable parameters of the regulator is given by:

$$\dot{e}_0(t) = A_m e_0 + (C_p B_p) \begin{bmatrix} (C_0 - C_0^*)r(t) \\ +(G_0 - G_0^*)X_p(t) \end{bmatrix} \quad (22)$$

Now let us consider for the parameters error the following notations:

$$\begin{bmatrix} \phi_r(t) \\ \phi_Y(t) \end{bmatrix} = \begin{bmatrix} C_0(t) - C_0^* \\ G_0(t) - G_0^* \end{bmatrix} \quad (23)$$

In the following representation, the right-hand sides only contain states  $(e_0, \phi_r, \phi_Y)$ . Consider then the Lyapunov function candidate corresponding:

$$V(e_0, \phi_r, \phi_Y) = \frac{e_0^2}{2} + \frac{k_p}{2g} (\phi_r^2 + \phi_Y^2) \quad (24)$$

Taking the derivative of Lyapunov function:

$$\dot{V}(e_0, \phi_r, \phi_Y) = -A_m e_0^2 \quad (25)$$

where  $A_m > 0$  is arbitrarily chosen by the designer.

So that the system is stable, it is necessary that  $V \geq 0$  and  $\dot{V} \leq 0$  what is checked. This involves that the adaptive system is stable in the sense of Lyapunov (Bakule2008).

### 3. THREE INTERCONNECTED AND NETWORKED CONTROL SUBSYSTEMS EXAMPLE

In this simulation party, we interest at a numerical example where  $N=3$  illustrated by figure 3 in order to demonstrate the efficiency of the proposed control.

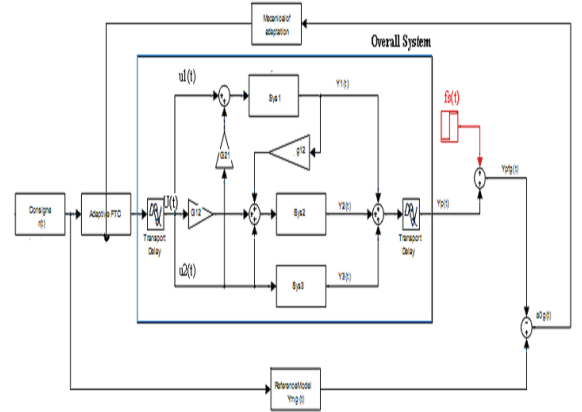


Figure 3: Functional diagram of Interconnected NCS in presence of sensor faulty

Consider the overall system (eq.1) consisting of three interconnected and networked control sub-systems ( $Sys_1$ ), ( $Sys_2$ ) and ( $Sys_3$ ) with unit reference input ( $r(t)=1$ ). On one hand based on fig.2, the transfer function of overall interconnected networked control reference model (eq.17) is giving as follows:

$$Sys_m = \frac{Y_m}{U_m} = e^{-0.348s} \frac{3s^3 + 11.26s^2 + 13.14s + 4.703}{s^4 + 5.5s^3 + 11.06s^2 + 9.688s + 3.125}$$

But, the transfer function of overall interconnected networked control system without external disturbances is giving as follows:

$$Sys_p = e^{-0.348s} \frac{3s^3 + 11.26s^2 + 13.14s + 4.703}{s^4 + 5.5s^3 + 11.06s^2 + 9.688s - 3.125}$$

So reference to the last transfer function, the overall interconnected and networked control system  $Sys_p$  is unstable.

On the other hand, the interconnection vector functions are respectively:  $G_{12}=-0.5$ ,  $G_{21}=0.5$ ,  $g_{12}=-0.579$  and the total delay is:  $\tau = 0.348$  where the delays into sensor/control and control/actuator are giving respectively:  $\tau_{sc} = \tau_{ca} = 0.174$ . The global Interconnected NCS ( $Sys_p$ ) is unstable.

Then as the fault  $f_s(t)$  is additive( fig.4) it can attack the total output  $Y_p(t)$  or one or more outputs  $Y_p(t)$  of the total system  $Sys_p$ . It is clearly, at the moment  $t = 20$ sec, we generate the offset fault with amplitude 1.

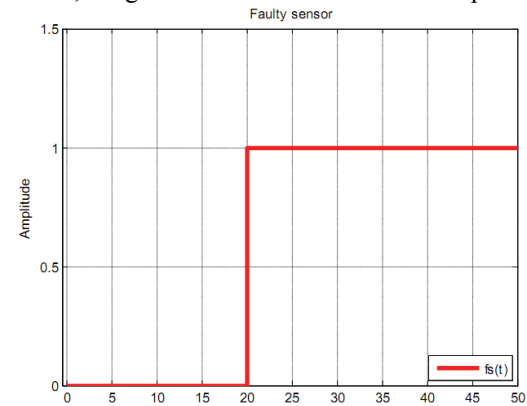


Figure 4: Offset fault

The time response  $Y_p(t)$  of the plant becomes:

$$Y_{pf}(t) = Y_p(t) + f_s(t), i = 1, \dots, N. \quad (26)$$

We use the control law (18) for various types of faults. The figures (4) and (5) illustrate the trajectories of the overall INCS time responses  $Y_{pf}(t)$ , the reference model  $Y_m(t)$  and the error  $e_0(t)$ .

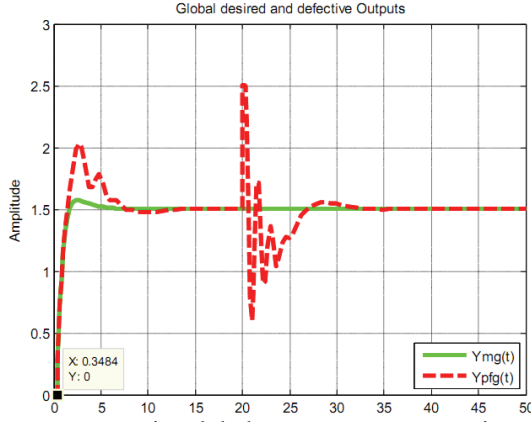


Figure 5: Dynamic global output responses via MR Adaptive FTC.

The figure 6 illustrates the dynamic global controls trajectories:  $U_p(t)$  were implemented on two different interconnected and networked control subsystems:  $Sys_{pi}$   $i=1,2,3$ , into to accommodate the defective overall Interconnected NCS:  $Sys_{pf}$ .

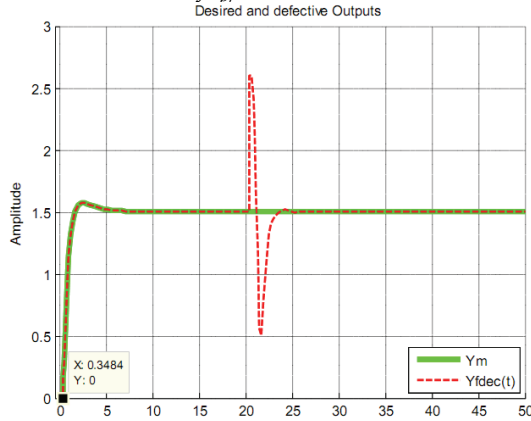


Figure 6: Dynamic distributed output responses via adaptive model reference FTC

#### 4. MAIN RESULTS AND INTERPRETATIONS

Here, to study the influence and the differences between centralized and decentralized architecturally adaptive controllers based on reference model, the figure5 and figure 6 illustrate the various response trajectories. It is clearly that the accommodation of Interconnected NCS<sub>f</sub> is realized by the proposed control laws such that are find simulation results. In addition, it is clearly that the adaptive FTC is implemented to three subsystems. They are LTI, continuous times, interconnected and networked control subsystems  $Sys_i, i=1,2,3$ . All constitute an overall interconnected NCS system  $Sys_p$  which his parameters depend essentially on all subsystems. Besides, the responses of defective plant  $Y_{pf}(t)$  are controlled by centralized control:  $U(t)$  and decentralized controls:  $u_1(t)$  and  $u_2(t)$

in the presence of faulty sensor (see respectively figure5 and 6). The global and distributed outputs respectively  $Y_{pfg}(t)$  and  $Y_{pf}(t)$  are reconfigured where a change then it recovers the desired output  $Y_m(t)$  of reference model. In figure7, the errors between the outputs defective processus and his reference model towards zeros for centralized and decentralized adaptive controllers.

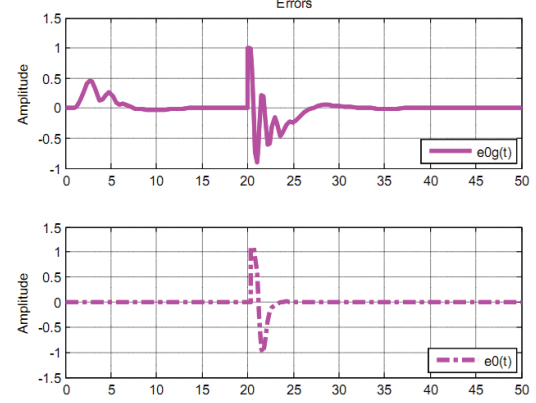


Figure 7: State errors

In the presence of additive defects, the accommodation of the N-interconnected and networked control subsystems is assured by the fault tolerant adaptive control based on reference model. We can note that the decentralized architecture of adaptive FTC implanted of an overall interconnected networked control system permits to build high performances and marvelous accommodation. The actual class of system may be robust to change all delay parameters. If the sum of delay  $\tau$  overstepping the limited values, the processus cannot be stable. The adaptive fault tolerant controllers are based on updated controller parameterizations  $C_0^*$  and  $G_0^*$ . They are updated by some adaptive laws and they are designed to minimize the deviation of output caused by the fault  $f_s(t)$ . Moreover, the figure8 illustrates simulation results of various trajectories (centralized and decentralized) controls.

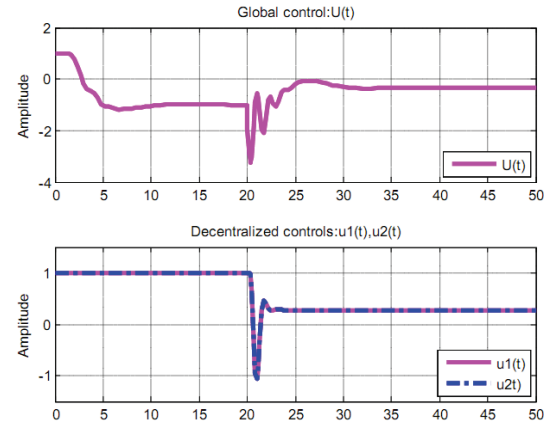


Figure 8: Trajectories of global and distributed controls

So, many strong points of the present fault tolerant adaptive control schemes in comparison with the existing techniques for two representations (centralized/decentralized) of INCSs are the following:

(i) they are two structures of large-scale system which differ from one another in the model used to represent the coupling between N-subsystems on which the controls are based;

(ii) sensor faulty dynamics are included in the adaptation loop and hence any on-line estimation parameters of overall system and N-interconnected subsystems will not cause significant of presence of sensor faults;

(iii) the present approach can be regarded as a generic one leading to different specific schemes that result from the choice of controller models of different degrees of complexity. There exists a tradeoff between models complexities and system performances for some tracking tasks. This can be used for the choice of appropriate fault tolerant adaptive control structures for desirable tasks;

(iv) the appropriate selection of decentralized techniques to be employed on N-interconnected subsystems distributed;

(v) a major point in the present development from the two architectures (centralized/decentralized) of controls that are high performances and marvelously reconfigurations of the distributed fault tolerant adaptive controls.

## 5. CONCLUSION

Model reference adaptive FTC implemented on interconnected and networked control system subject to compensate faulty sensor was studied in this paper. The main contributions include essentially:

1) Modeling NCSs subject to both access interactions and delay in presence of faulty sensors;

2) Determining different architectures (centralized and decentralize) of adaptive model reference controllers for N-interconnected and networked control subsystems.

A simulation example of three interconnected and networked control subsystems was studied to illustrate effectiveness of the proposed method.

## REFERENCES

- Blanke, M., Kinneert, M., LUNZE, J., Staroswiecki, M., 2003. Diagnosis and fault tolerant control. Springer, Berlin, Heidelberg, New York.
- Bodson, M., Greszkiewicz, E.J., 1997. Multivariable Adaptive Algorithms for Reconfigurable Flight Control. *IEEE Transaction on Control systems Technology*, 5( 2).
- Patton, R.J., 1997. Fault-tolerant control: The 1997 situation. pp.1033–1055. In *IFAC Fault Detection, Supervision and Safety for Technical Processes*, Kingston Upon Hull, UK.
- Patton R. J., Kilinkhieo, 2009. An Adaptive Approach to Active Fault-Tolerant Control. ). pp.54-61. *The Open Automation and Control Systems Journal*, 2(3).
- Ioannou, P., Koktovic, P., 1985. Decentralized Adaptive Control of Interconnected Systems with Reduced-order Models. pp. 401-412. *Automatica*, 21( 4) .
- Mahmoud, M.S., 1997. Adaptive Stabilization of a class of Interconnected Systems. pp. 225-238. *Computers Elect. Engng*, 23( 4).
- Huang, D., Nguang, S.K., 2010. Robust fault estimator design for uncertain networked control systems with random time delays: An ILMI approach. pp. 465-480. *Information Sciences*.180(3).
- Lina, L., Chen, H.C, 2006. Robust control of interconnected network schemes subject to random time varying communication delays. pp.647 - 658, *Int. J. Electron. Commun. (AEÜ)* 60(2).
- Patton, R.J., Kambhampati, C., Casavola, A., Zhang, P., Ding, S. Sauter, D., 2007. A Generic Strategy for Fault-Tolerance in Control Systems Distributed Over a Network. pp.1-17. *European Journal of Control*.12(4).
- Tsai, N.C.Sue, C.Yang, 2009. Integrated model reference adaptive control and time-varying angular rate estimation for micro-machined gyroscope. *International Journal of Control*.
- Bakule, L., 2008. Decentralized control: An overview. pp. 87-98. *Annual Reviews in Control*. 32(3).
- Challouf, A., Telli, A., Abdelkrim M. N., 2009. Fault-Tolerant Adaptive control of linear Systems applied to A Robot H4. *Le 3ème Colloque de Recherche Appliqué et de Transfert de Technologie " 2ème conférence internationale*. 11-12 November ( Tunisie).
- Hansheng, W., 2003. Decentralized adaptive robust control for a class of large scale systems with uncertainties in the interconnections. pp.253-265. *International Journal of Control*, 76(3).
- Hovakimyan and Ramachandra A., 2006. Decentralized adaptive output feedback control via input/output inversion. pp.1538-1551. *International Journal of Control*, 79( 12).
- Sundarapandian, V., 2005. Distributed Control Schemes for Large-Scale Interconnected Discrete-Time Linear Systems. pp. 313-319. *Mathematical and Computer Modelling*, 41(2).

# MODELLING NONI (*MORINDA CITRIFOLIA* L.) PULP SPRAY DRYING USING A CROSSED MIXTURE-PROCESS DESIGN AND FEED OPTIMIZATION

Eliosbel Márquez<sup>(a)</sup>, Débora Castro<sup>(b)</sup>, Margarita Nuñez<sup>(c)</sup>, José Luis Rodríguez<sup>(d)</sup>

<sup>(a,c,d)</sup>Food Industry Research Institute, Carretera Guatao km 3½,  
Havana 19 200, Cuba.

<sup>(b)</sup>Finlay Institute, Av. 27 No. 19805 esq. 198. Lisa, Havana 11600, Cuba.

<sup>(a)</sup>[eliosbel@iiaa.edu.cu](mailto:eliosbel@iiaa.edu.cu), <sup>(b)</sup>[dcastro@finlay.edu.cu](mailto:dcastro@finlay.edu.cu), <sup>(c)</sup>[margarita@iiaa.edu.cu](mailto:margarita@iiaa.edu.cu), <sup>(d)</sup>[joseluis@iiaa.edu.cu](mailto:joseluis@iiaa.edu.cu)

## ABSTRACT

The effects of noni pulp:maltodextrin (MD) ratio and dry matter (DM) of the feed to the spray drier on different bioactive substances and drying quality attributes of the dried product were evaluated by means of a crossed mixture-process design and the response surface methodology was applied. The mixture components in the feed, noni pulp and MD, were in the 80:20 to 90:10 range. The process factor was the feed's DM, at three levels, 10, 15 and 20 g/100g. The experiments took place in a Niro pilot-scale spray dryer. The lowest powder moisture attained by the model, 3.97 %, corresponded to a feed with 15 % DM and 80:20 pulp:MD ratio. When this ratio is increased, reconstituted product browning also increases with respect to the feed. Maximum ascorbic acid and total phenol were 467 mg/100g and 866 mg gallic acid /100g respectively, corresponding to the 90:10 pulp:MD ratio. The feed's DM only had a significant effect on powder moisture.

**Key words:** Noni pulp, antioxidants, spray drying, crossed mixture-process design.

## 1. INTRODUCTION

The noni tree (*Morinda citrifolia* L.) belonging to the Rubiaceae family, is native of Polynesia, India and Southeast Asia, but it has spread to other regions, such as Australia, Latin America and the Caribbean (Cribb and Cribb 1975; Elkins 1998; Chan et al. 2006). The noni fruit has been used as food and, for its therapeutic properties, in traditional medicine for over 2000 years (Earle 2001; Chan et al. 2006). Clinical and laboratory research has been carried out lately to determine its possible effectiveness (Dixon et al. 1999; McClatchey 2002). In the last few years the consumption of products of this fruit increased notably in the United States, Japan and Europe. The FAO/WHO combined report (2008) on the food standards programme estimated a noni market of 400 million USD in 2001, while the estimate in 2006 had increased to 2000 million USD, which made it the fastest growing market in the global health products sector. Noni juice has been accepted in the European Union as a novel food and the fruit was evaluated as to safety and approved for human

consumption (European Commission 2003; West et al. 2006).

Accordingly, the design and characterization of spray dried noni products, to be used by the food industry as functional ingredients in a range of health products and as novel nutritional supplements is considered interesting. As the high antioxidant capacity of this fruit is given mainly by the presence of phenols and ascorbic acid (Márquez et al. 2008), it is important to evaluate the influence of the characteristics of the spray dryer feed on the content and retention of these bioactive substances in the end product, a dry powder. On the other hand, it is well known that adhesiveness is the main characteristic limiting the usefulness of spray drying of sugar-rich products, such as fruit juices and pulps, because of low yields and operational problems in drying. The problem of adhesiveness in such dehydrated products has been related to its low glass transition temperature (T<sub>g</sub>), a specific property of amorphous materials, which not only present difficulties during drying but are also susceptible to collapse during storage. Therefore, a common approach to drying these products has been to add high molecular weight substances, such as starches and maltodextrin (MD), to increase T<sub>g</sub>, up to 75% of these solids being sometimes admitted in the dry matter (Bhandari et al. 1993; Roos et al. 1996; Bhandari, et al. 1997a; Bhandari, et al. 1997b; Bhandari & Howes 1999).

It is claimed (Cornell 1990), that the typical strategy for the experimental design of industrial processes has been to optimize the variables separately, using a mixture design for the formulas and a factorial design and response surface methods for the process. However, the interactions among compositional and process variables cannot be revealed by this simplistic approach; hence the importance of using a crossed design combining mixture components with process factors.

The objective of this study was to evaluate the influence of the feed's noni pulp: MD ratio and dry matter on the content of bioactive substances and drying quality attributes of spray dried noni pulp, by means of a D-optimal crossed mixture-process design as well as to optimize the feed's composition.

## 2. MATERIALS AND METHODS

### 2.1. Raw material

Ripe noni fruits (their ripeness degree characterized as “soft white”) were reduced to a pulp in a pilot scale Shuttevaerweg-60 pulpiers-trainer. This pulping procedure (with 80 % yield) eliminates only the seed, so integral use to the fruit is achieved. Some physical-chemical characteristics of the noni pulp thus obtained are: dry matter,  $10.03 \pm 0.44$  %; soluble solids,  $9.25 \pm 0.55$  °Brix; total dietetic fibre 1.9 %, pectin 0.96 %, pH,  $4.06 \pm 0.01$  and acidity,  $0.55 \pm 0.01$  g/100g malic acid, the major acid component (Márquez et al. 2008; Pino et al. 2009). Maltodextrin used (IMSA Mexico) had a dextrose equivalent (DE) of 9-12% and 5% moisture.

### 2.2. Modelling

The experiments were carried out using a D-optimal crossed mixture-process design of experiments (Cornell 1990; Montgomery 1991; Anderson and Whitcomb 1996, 1998; Núñez 2000), which combines mixture components with process factors. In this type of design the mixture components are dependent of each other but independent of the process factors. The mixture components were the ratio of noni pulp (P) to maltodextrin (MD) in the formula fed to the dryer, in the 80:20 to 90:10 ranges. For this range the solid pulp:MD ratio varies between 30:70 and 50:50. The process factor studied was the feed's dry matter (DM), at three levels, 10, 15 and 20 g/100g of liquid formula. The design matrix encompassed a total of 11 experiments and 14 experimental runs, with replicas in the experiment with feed consisting of 85% pulp:15% MD and DM of 15 g/100g liquid formula.

It is important to point out that the addition of water to the different P:MD rate of the experimental matrix, constitutes the only way to achieve an essential condition of this kind of design, that the mixture and process variables are independent of each other, it means that feed DM does not depend on or is not determined by P:MD rate, in such a way that for a same P:MD proportion is possible, by adding water, to obtained different feed DM to the dryer.

### 2.3. Spray drying

The noni pulp was manually mixed with maltodextrin and water according to the formula for each experiment, homogenized in a colloidal mill and spray dried. The experimental runs were made in a pilot-scale Niro Atomizer Minor (Denmark) spray drier, fitted with a centrifugal disk ( $2 \times 10^4 \text{ min}^{-1}$ ) under the same operating conditions: feed temperature, 30°C; inlet air temperature, 150 °C and outlet air temperature, 75-80 °C. Feed flow provided by a positive displacement pump (maximum capacity, 40 L/h) was regulated in order to assure the desired outlet temperature. In each experience 20 kg of liquid formula was dried.

### 2.4. Chemical – physical and sensory analysis

The following analyses were carried out in noni pulp and in the experimental powder samples: humidity (AOAC 1995); ascorbic acid (NC ISO 6557: 01); non

enzymatic browning (Hendel 1950) and total phenols (Kang and Saltveit 2002). Ascorbic acid and total phenol retention was calculated from the content (mg/100 g) of these compounds in the feed's dry matter and in the powdered product. The sensory evaluation of dehydrated samples colour of each experiment was carried out from the browning point of view with trained judges, by means of a categorical test of 5 points, where punctuation 5 corresponded to the pearl typical colour of the noni pulp and punctuation 1 corresponded to a much browned colour (NC-ISO 4121, 2005).

### 2.5. Statistical processing

The response variables of the design were humidity, browning of the reconstituted powder; increase in the reconstituted powder browning in regard to the feed; content and retention of ascorbic acid and total phenols. The response surface methodology was used to analyze the results (Montgomery 1991), and the noni pulp:maltodextrin ratios and the DM of the drier feed were optimized by the numeric optimization method, setting restrictions to the response variables by means of the DX version 6 programme.

## 3. RESULTS AND DISCUSSION

For the analysis of results it was kept in mind that the interpretation of the models is carried out by way of the significant terms in the coded models. The models where the process variable studied (the feed's dry matter) do not appear or has no significant influence can be interpreted properly as mixture models. The quadratic and cubic terms in a mixture model are considered relationships of non linear mixtures or combinations, not interactions, because their components cannot vary independently of each other (Núñez, 2000). The mixture models do not have an intercept. Estimates of the response variables by means of the models are performed by the DX programme with all its terms, both significant and not significant.

In the tables 1-4 appear the results of the coded models obtained for the assessed response variables. The quality indicators considered for each model were: significance of the analysis of variance of the regression ( $p < 0.05$ ); non significance of the lack of fit test ( $p > 0.05$ ); values of  $R^2 > 0.8$ ; significance of the coefficients of each term of the models ( $p < 0.05$ ) and standardized residues following a normal distribution with average zero and  $s = 1$  without atypical observations. All the indicators behaved satisfactorily and therefore the models had the quality required for their analysis.

### 3.1. Analysis for moisture model of dehydrated noni pulp

From the coded crossed model for the moisture of the spray dry noni pulp (Table1), was obtained that it decreases when MD increases with respect to the pulp in the feed and that the feed's DM has no significant influence independently, but is manifest in an interaction with the proportion of MD in the mixture, so that the smallest value of powder moisture estimated by

means of the model (3.97 %) is obtained with a feed's DM of 15% and an pulp : MD ratio of 80:20 (Fig. 1 and 2).

Although in general, in the spray dried products a larger feed's DM causes smaller moisture values in the final dry product, these results prove, in first place, that not all the pulps have the same behaviour and a high feed viscosity, due to the increase of the DM, from certain limits, may unfavourably influences on the drying. In second place, this is an example of the importance of the designs applied to spray drying, which take into account the interaction between process variables and mixture components to obtain minimal values of moisture in the dehydrated products.

Table 1: Results for the moisture (g/100 g of powder) model of dehydrated noni pulp

Term of model	Estimate coefficient	Significance
Intercept	8.86	
B	-4.89**	0.001
X	-1.02	0.065
X <sup>2</sup>	-0.88	0.281
BX	0.85	0.276
BX <sup>2</sup>	4.60**	0.006
Kind of model	Crossed lineal mixture x quadratic process	
Significance of variance analysis	0.0043**	
R <sup>2</sup>	0.845	

B: Proportion of maltodextrine respect to noni pulp in the feed

X: DM of the feed

$0 \leq A \leq 1$   $0 \leq B \leq 1$   $A+B=1$   $-1 \leq X \leq 1$

\*\*\*, \*\*, \* Coefficient and model significance to  $p < 0.001$ ;  $p < 0.01$ ;  $p < 0.05$ , respectively

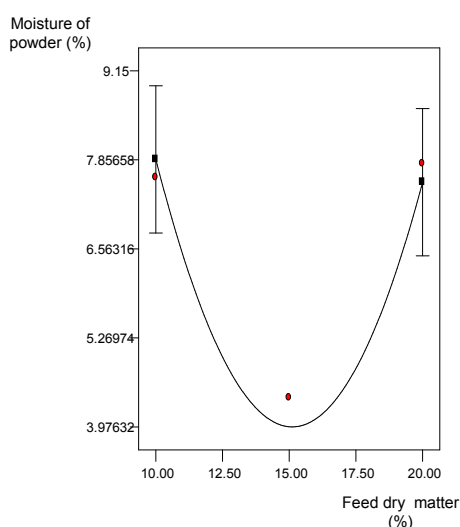


Fig. 1: Influence of feed dry matter on moisture of dehydrated noni pulp to P:MD rate 80:20 in the feed.

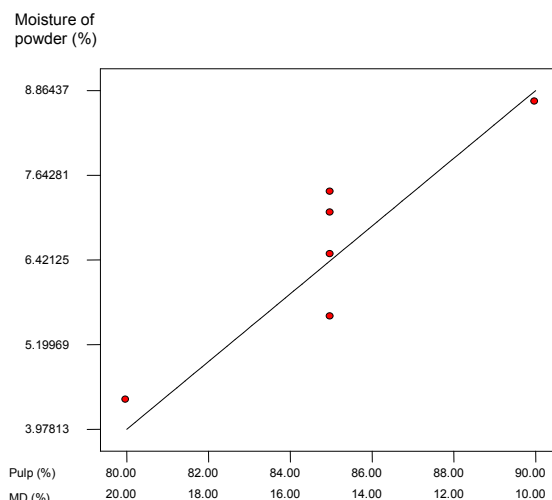


Fig. 2 .Influence of P:MD rate of the feed on the moisture of dehydrated noni pulp to a feed DM of 15 %

### 3.2. Analysis for browning models of dehydrated noni pulp

No browning model was found for the dehydrated noni pulp, but models were found for browning of the reconstituted powder and for the increase in browning of the reconstituted powder with regard to the feed. Browning of the reconstituted powder constitutes a quality indicator for possible applications of the powder, i.e., in reconstituted drinks, while the increase in browning (Table 2) allows the evaluation of the influence of the feed's composition on dried noni pulp browning. In both response variables the feed's DM showed no influence, the MD:pulp ratio did; hence, its interpretation is properly that of cubic mixture models.

Table 2: Results for the model of increase in browning of the reconstituted noni pulp powder respect to the feed.

Term of model	Estimate coefficient	Significance
A	0.064**	0.0038
B	0.031**	0.0038
AB	-0.13*	0.015
AB(A-B)	0.92***	0.0003
Kind of model	Cubic mixture	
Significance of variance analysis	0.0002***	
R <sup>2</sup>	0.856	

A: Proportion of noni pulp respect to maltodextrine in the feed

B: Proportion of maltodextrine respect to noni pulp in the feed

$0 \leq A \leq 1$   $0 \leq B \leq 1$   $A+B=1$

\*\*\*, \*\*, \* Coefficient and model significance to  $p < 0.001$ ;  $p < 0.01$ ;  $p < 0.05$ , respectively

As the maltodextrin of feed mixture increased, browning of the reconstituted powder decreased as well as too its browning with regard to the feed, probably



due to a bleaching effect of the MD, that attenuates pulp darkening produced by reducing sugars during drying. In fact, in some experiments negative lightly values were obtained in the increase of browning of powdered samples as compared with the fed formulae, which points out that instead of darkening a little clearing of the pulp takes place during drying.

The results of the sensory evaluation of powder colour were consistent with the analytic results of reconstituted pulp browning, as it was smaller and the punctuation higher in samples with smaller pulp to MD ratio. Nevertheless, none of the powdered samples from the experiments was rejected due to sensorial evaluation of colour.

### 3.3. Analysis of models for ascorbic acid and total phenols of dehydrated noni pulp

From the nutritional and healthy point of view it is important to evaluate the influence of spray drying conditions on the ascorbic acid content of the noni pulp, given its high concentration in this nutrients ( $150.94 \pm 11.74$  mg/100 g of powder, according to Márquez et al. 2008), similar to that of guava. From the table 3 can be observed for its interpretation that, as expected, a smaller MD:pulp ratio in the feed implies a larger proportion of pulp in the mixture and, therefore, of ascorbic acid in the dried product, so the maximum value estimated by the model was 467 mg/100 g for 90 % pulp and 10% MD.

Table 3: Results of ascorbic acid model (mg/100 g of dehydrated noni pulp)

Term of model	Estimate coefficient	Significance
Intercept	466.94	
B	1930.38*	0.022
X	0.28	0.995
BX	1056.91	0.4043
B <sup>2</sup>	-6111.17*	0.0128
B <sup>2</sup> X	-1928.61	0.5839
B <sup>3</sup>	4035.30*	0.0123
B <sup>3</sup> X	743.51	0.7448
Kind of model	Crossed Cubic mixture x lineal process	
Significance of variance analysis	0.0213*	
R <sup>2</sup>	0.876	

B: Proportion of maltodextrine respect to noni pulp in the feed

X: DM of the feed

$0 \leq A \leq 1$   $0 \leq B \leq 1$   $A+B=1$   $-1 \leq X \leq 1$

\*\*\*, \*\*, \* Coefficient and model significance to  $p < 0.001$ ;  $p < 0.01$ ;  $p < 0.05$ , respectively

Although model obtained was crossed: cubic mixture - linear process, as the feed's DM did not have a significant influence on the powder's ascorbic acid content, its interpretation is made like a cubic mixture model.

The noni pulp is also rich in other bioactive substances, as total phenols ( $181.30 \pm 15.84$  mg of gallic acid/100 g of powder), (Márquez et al. 2008) and its content in the dried product was favoured with the increase of the pulp:MD ratio in the feed (Table 4), the same performance observed for ascorbic acid, so the highest phenol content of the powder according to the model (866 mg gallic acid/100 g) was obtained with a 90:10 pulp:MD ratio in the feed, and the lowest (552 mg gallic acid/100 g) with a 80:20 ratio (Fig. 3), regardless of the feed's DM, which had no influence on this response variable and a quadratic mixture model was obtained.

Table 4: Results of total phenols model (mg gallic acid/100 g of dehydrated noni pulp)

Term of model	Estimate coefficient	Significance
A	865.98***	0.0001
B	552.58***	0.0001
AB	-319.23*	0.0272
Kind of model	Quadratic mixture	
Significance of variance analysis	0,0001***	
R <sup>2</sup>	0,845	

A: Proportion of noni pulp respect to maltodextrine in the feed

B: Proportion of maltodextrine respect to noni pulp in the feed

$0 \leq A \leq 1$   $0 \leq B \leq 1$   $A+B=1$

\*\*\*, \*\*, \* Coefficient and model significance to  $p < 0.001$ ;  $p < 0.01$ ;  $p < 0.05$ , respectively

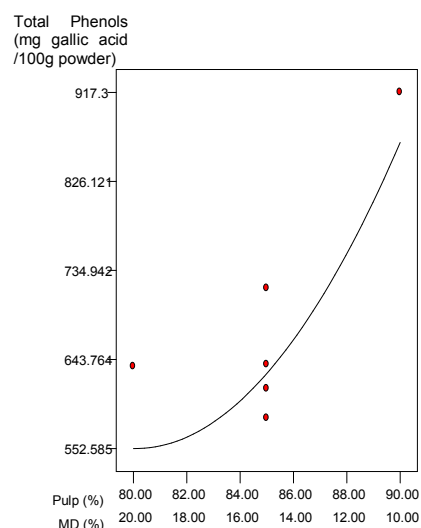


Fig. 3. Influence of P:MD ratio of feed on dehydrated noni pulp total phenols.

The maximum values of both ascorbic acid and total phenols corresponded to the higher purity powdered samples of the experimental design (45.3 – 46.5 g noni solids/100 g of the powdered sample) and the minimum values of these antioxidants corresponded well with the



lowest purity samples (26.3-28.7 g noni solids/100 g of the powdered sample). Powder purity mainly depends on noni solids in the DM of the powder (determined by the pulp:MD ratio of the feed) and, in a smaller measure, on the dry powder's DM.

Feed variables did not influence on ascorbic acid and total phenol retention during drying, so that models were not found for them, at least in the range studied, though the average retention values of the experiments were good,  $92.26 \pm 7.78$  % and  $98.08 \pm 2.70$  %, respectively. For both antioxidants the range of experimental retention values varied in the 73 to 100 % range.

### 3.4. Optimization of the feed

The restrictions imposed on the response variables for dryer feed optimization were: maximization of the ascorbic acid and total phenol contents and minimization of moisture, as well as of the increase in browning of the reconstituted powder respect to the feed. Table 5 shows the two optimum conditions of the feed and the table 6 shows the values of the response variables estimated by the models for these conditions. The first condition is considered the most convenient overly (Núñez, 2000) that is, the one that better satisfies the restrictions of all the response variables.

A new industry, the so-called “well-being or health industry” manufactures products that have a great demand in the market at the moment. The spray dried noni pulp can have multiple uses as functional ingredient for this industry, in the design of new products and nutritional and/or therapeutic supplements, e.g. capsules or pills, and in powdered or liquid food formulas. The dehydrated pulp, by itself, could also constitute a product for preparing reconstituted drinks; in this case, considering the unpleasant noni taste, the addition of some flavouring substances should be contemplated.

Table 5. Optimal feed conditions for spray drying of noni pulp.

Noni pulp (%)	Maltodextrine (%)	Dry matter (%)
84.9	15.1	17.9
80.2	19.8	13.1

Table 6: Estimates of response variables by the models for the optimal conditions of feed

Response variables	1 <sup>st</sup> optimal condition	2 <sup>nd</sup> optimal condition
Powder moisture (%)	6.52	4.67
Increase in reconstituted powder browning respect to feed	0.009	0.011
Ascorbic acid (mg/100g powder)	474.14	330.12
Total phenols (mg gallic acid/100g powder)	625.94	552.60

## 4. CONCLUSIONS

- It is possible to design powdered, functional and novel noni products with this methodology, using a crossed mixture-process experimental design to optimize the proportion of components of the mixture fed to the drier and the spray drying conditions simultaneously. This can be accomplished by means of the response surface methodology and the application of restrictions to different quality indicators of the dehydrated product.
- When the MD to noni pulp ratio in the spray drier feed is increased from 10 to 20 %, powder moisture decreases. The minimum humidity value in the powder (3.97 %) obtained by the model corresponded to a feed mixture with 15% DM and 80: 20 pulp:MD ratio.
- The maximum values of ascorbic acid and total phenols were 467 mg/100g and 866 mg gallic acid/100g, respectively, and they corresponded to the highest pulp:MD ratio in the design, 90:10, regardless of the feed's DM, which did not significantly influence any of these variables. Models were not obtained for retention of these antioxidants, but they were high (73 to 100%)
- Minimizing or maximizing the response variables, according to the case, two optimum feed conditions were obtained: 84.9:15.1 pulp:MD ratio with 17.9% DM and a 80.2:19.8 pulp: MD ratio with 13.1% DM.

## REFERENCES

- Anderson, M.J., Whitcomb, P.J., 1996. Optimize Your Process-Optimization Efforts. *Chemical Engineering Progress*, December.
- Anderson, M.J., Whitcomb, P.J., 1998. Find the Most Favorable Formulations. *Chemical Engineering Progress*, April.
- AOAC., 1995. *Official Methods of Analysis of the Association of Official Analytical Chemists*. 6th ed., Helrich K. (Ed.), AOAC International, Arlington, VA.
- Bhandari, B. R., Datta, N., & Howes, T., 1997b. Problems associated with spray drying of sugar-rich foods. *Drying Technology*, 15(2): 671-684.
- Bhandari, B. R., Datta, N., Crooks, R., Howes, T. & Rigby, S., 1997a. A semi-empirical approach to optimise the quantity of drying aids required to spray dry sugar-rich foods. *Drying Technology*, 15(10): 2509-2525.
- Bhandari, B. R., Senoussi, A., Dumoulin, E. D. and Lebert, A., 1993. Spray drying of concentrated fruit juices. *Drying Technology*, 11(5): 1081-1092.
- Bhandari, B.R. & Howes, T., 1999. Implication of glass transition for the drying and stability of dried foods. *Journal of Food Engineering*, 40: 71-79.
- Chan-Blanco, Y., Vaillant, F., Pérez, A. M., Reynes, M., Brillouet, J.M. & Brat, P., 2006. The noni fruit (*Morinda citrifolia* L.): A review of agricultural research, nutritional and therapeutic properties. *Journal of Food Composition and Analysis*, 19: 645-654.
- Cornell, J., 1990. *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture data*, 2nd ed., Example 7-4, John Wiley & Sons, Inc, New York.
- Cribb, A.B., Cribb, J.W., 1975. *Will foods in Australia*. William Collins, Pub. Pty. Ltd. Sidney, Australia.
- Dixon, A.R., H. McMillen & Etkin, N.L., 1999. Ferment This: The transformation of Noni, a traditional Polynesian medicine (*Morinda citrifolia*, Rubiaceae). *Ecological Botany* 53: 51-68.
- Earle, J. E., 2001. *Plantas medicinales en el Trópico Húmedo*. Ed. Guayacán, San José de Costa Rica.
- Elkins, R., 1998. Hawaiian noni (*Morinda citrifolia*) prize herb of Hawaii and the South Pacific. *Woodland Publishing*, 22-23.
- European Commission, 2003. 2003/426/EC: Commission decision of 5 June 2003 authorising the placing on the market of "noni juice" (juice of the fruit of *Morinda citrifolia* L.) as a novel food ingredient under regulation (EC)No 258/97 of the European Parliament and of the council. *Official J of the European Union L* 144, 12/06/2003, p 0012-0012.
- Hendel, E.C., Bailey, G.F. and Tailor, D.H., 1950. Measurement of non enzymatic browning of dehydrated vegetables during storage. *Food Technol.*, 4: 344.
- Kang, H., M. y Saltveit, M., E., 2002. Antioxydant capacity of lettuce leaf tissue increases after wounding. *J. Agric. Food Chem.* 50:7536- 7541.
- Márquez, E.; Castro, D.; Ventosa, M., Rodríguez, J. L.; García, T. y Zequeira, O., 2008. Caracterización de tres estados de madurez del noni (*Morinda citrifolia*, L.) y de productos derivados. En *Memorias CICTA 11*, 13-17 oct. La Habana, Cuba.
- McClatchey, W.C., 2002. From Polynesian Healers to Health Food Stores: Changing Ethnopharmacology of *Morinda citrifolia*. *Journal of Integrative Cancer Therapy*, 1(2): 110-120.
- Montgomery, D., 1991. *Diseño y Análisis de experimentos*, 3ra edición, Grupo editorial Iberoamérica s.a., 591p.
- NC- ISO 6557, 2001. Productos de frutas y vegetales. Determinación del contenido de ácido ascórbico.
- NC-ISO 4121, 2005. Análisis sensorial. Guía para el uso de escalas con respuestas cuantitativas.
- Núñez de Villavicencio, M., 2000. *Diseños de mezclas: Aspectos teórico-prácticos de su aplicación*. Tesis de maestría, Universidad de La Habana, Cuba.
- Pino, J.; Márquez, E.; Castro, D., 2009. Volatile and non volatile acids of noni (*Morinda citrifolia* L.) fruit. *J. Sci Food Agric.*, 89: 1247-1249.
- Roos, Y. H., Karel, M., & Kokini, J. L., 1996. Glass transitions in low moisture and frozen foods: Effect on shelf life and quality. *Food Technology*, November: 95-108.
- West, B.J., Jensen, C.J., Westendorf, J. & White, L.D., 2006. A Safety Review of Noni Fruit Juice. *Journal of Food Science*, 71: 100-106

# FACTORY LAYOUT PLANNER: HOW TO SPEED UP THE FACTORY DESIGN PROCESS IN A NATURAL AND COMFORTABLE WAY

Antonio Avai<sup>(a)</sup>, Giovanni Dal Maso<sup>(b)</sup>  
Paolo Pedrazzoli<sup>(c)</sup>, Diego Rovere<sup>(d)</sup>

<sup>(a) (b)</sup>Technology Transfer System s.r.l., Milano, Italy  
<sup>(c) (d)</sup>SUPSI-ICIMSI, Manno, Switzerland

<sup>(a)</sup>[avai@ttsnetwork.com](mailto:avai@ttsnetwork.com), <sup>(b)</sup>[dalmaso@ttsnetwork.com](mailto:dalmaso@ttsnetwork.com)  
<sup>(c)</sup>[paolo.pedrazzoli@icimsi.ch](mailto:paolo.pedrazzoli@icimsi.ch), <sup>(d)</sup>[diego.rovere@icimsi.ch](mailto:diego.rovere@icimsi.ch)

## ABSTRACT

This paper describes a new tool, named *Factory Layout Planner (FLP)*. The FLP is meant to speed up the factory design process and to facilitate the cooperation of heterogeneous actors usually involved in the design, simulation and analysis of future or existing layouts. The idea is to provide an innovative *environment* where people can perform the aforementioned activities in a natural, comfortable and creative way, without any constraints. In the vision of the authors, this requirement points to the massive use of 3D technology supported by a new human-computer interaction solution based on the multi-touch paradigm. The FLP provides both 3D and discrete event simulation capabilities, that are key enabling technologies in speeding up the layout planning process and in finding out the desired solution. A simple test case is reported to demonstrate the industrial applicability of the FLP.

**Keywords:** layout planning, multi-touch system, 3D and discrete event simulation.

## 1. INTRODUCTION

Manufacturing facilities are one of the most important strategic elements of a business enterprise. They are expensive and their lifespan is some decades. A properly designed plant layout is an important source of competitive advantage (Quartermann et al, 1996).

Planning a factory layout involves deciding where to put all the facilities, machines, equipment and staff in the manufacturing operation and the way in which materials and other inputs (like people and information) flow through the operations.

Traditionally, factory design processes are prone to error (Ding et al, 2010). Relatively, small changes in the position of a machine in a factory layout can affect considerably the flow of materials. This, in turn, impacts on the costs and effectiveness of the overall manufacturing operation. Therefore, any mistake can lead to inefficiency, inflexibility, large volumes of inventory and work in progress, high costs and unhappy customers. Changing a layout can be expensive and

difficult, so it is best to get it right the first time (Meyers and Stephens, 2000).

Thus a need for a new factory planning tool, easy to use and understand for all the involved actors, is becoming more and more pressing. 3D technology is mandatory for such a tool because 3D models look realistic, provide a better understanding of space requirements and can be animated to show the physical flow of materials. Users can have a 360 degrees view of the new factory layout and they can easily see how factory objects interact, to detect the weakness of the layout and to reduce the interpretation errors.

Another key feature is represented by the simulation technology, both 3D and discrete event, that promotes more informed decisions, before any equipment is installed (Boër et al., 1993).

Last but not less important, the way to interact with this tool should be very intuitive, quick and natural because several actors, with different skills, are involved in this strategic process.

Keeping in mind this context, the Factory Layout Planner (FLP) was developed with the main aims:

1. To accelerate and simplify the factory layout process by creating accurate factory 3D models.
2. To foster the effective collaboration between teams.
3. To facilitate the human computer interaction.
4. To reduce the time to production of new layouts.

This paper describes the main modules and technologies used in developing the FLP. It is organized as follows: section 2 presents a state of the art about factory layout planning tools; section 3 describes, in details, the architecture and the applications composing the FLP; section 4 provides a real test case and conclusions follow.

## 2. STATE OF THE ART

The state of the art in 3D factory layout planning and simulation for discrete manufacturing is represented by some big software solutions, generally, classified as PLM (*Product Lifecycle Management*) that provide a

support from the product to the process design and management.

The major references are: *Dassault Systemes – Delmia* and *Siemens - Tecnomatix*, which support both layout design and simulation and integrate modules for task programming and production process management. These suites require technical capabilities, training courses, changes in the mindset of the involved people, changes in the processes and, sometimes, in the company organization that can discourage small or medium enterprises. Another popular suite is *Autodesk – Factory design suite*, that allows integrating 2D layout data with 3D models, but here the simulation feature is still missing and it requires people with some CAD skills. Finally, *Visual Components* supports 3D components programming and assembly for an easy layout design, while simulation requires some specific programming competences.

On the market, there are other small solutions more oriented to discrete event simulation but with some features meant to support 3D animation. A first example is *Flexsim simulation*, that allows to drag and drop 3D objects, but requires consistent programming capabilities. Another example is *Arena*, where a “look-like 3D” layout creation is supported by an additional module. Last we mention *Simio*, that allows to generate 2D and again “look-like 3D” layouts. For those tools, simulation is focused only on the 2D objects. And again all these tools require skill and know-how in simulation modelling and programming.

Several research works have been proposed over the last decades addressing the benefits of using simulation during the design of new layouts for discrete manufacturing (Avai et al., 2011, Kyle et al., 2008, Voorhorst et al., 2008). Some of them describe the development of specific applications to facilitate the use of such models or the results' interpretations (Fagent et al., 2005). While Mert (Mert, 2005) investigated how simulation helps in finding out the most performing new layout when different techniques and principles are applied.

Additionally, the factory layout planning is considered in several European projects. “*EuroShoe*” project studied “ad hoc” solution for mass customized production plants; “*Difac*” (Digital Factory for Human Oriented Production) was focused on the planning of generic manufacturing factories and “*Eupass*” (Evolvable Ultra-Precision Assembly Systems) dealt with the layout configuration of a work cell.

FLP should inspire the creativity of the team as well, because, according to Harron (Harron et al., 2008) the design of a new layout requires both art and technology. It's not only an application of technology, but it's also a creative process.

The main features of the FLP can be described as follows:

- To allow planning all the components in a comfortable, intuitive and creative way.

- To provide 3D and discrete event simulation in a smart way.
- To guarantee a natural interaction.
- To allow the collaboration between different actors also remotely placed.
- To interface with other CAD systems.
- To provide some basic drawing functions.

The next paragraph will describe in details the innovative aspects of the FLP.

### 3. FLP OVERVIEW

The FLP is not a single program but a suite of applications built on top of a set of common libraries. These applications are targeted at different users and contexts and can speak with each other, share documents and data thanks to the client/server architecture. The main common libraries are:

1. The collaborative engine, that allows multiple users to act at the same time on the same layout in a local or distributed environment.
2. The 3D visual engine, that provides the user interface to edit the layout.
3. The simulation engine, that enables to perform discrete event simulation (DES) on the layout that the user is composing.

The FLP suite and its main technologies are described in the following subsections.

#### 3.1. FLP architecture

The FLP is based on a two-level architecture with a fat client: the server is mainly a synchronization manager and a repository. The documents are required data (e.g. catalogue of available components, results of simulations) and are stored on the server. The client applications can connect to the server to get all the needed data and to edit the layout. The used protocol was optimized to minimize network traffic and to be tolerant of network delay, allowing a seamless remote editing experience. The FLP server is targeted at IT staff and should run on server computers.

Thanks to client/server architecture, the collaboration on the layout planning can be both remote and local. While the first allows users, distributed all over the world, to cooperate in the layout creation, the latter allows users to act on the same device at the same time on a common model.

#### 3.2. FLP Desktop

This is the main application targeted at the users who design, refine layouts or perform 3D and discrete event simulation. It is mostly composed by a catalogue browser and a 3D editing window. Clearly, the 3D components' models can be imported from a CAD system and saved in the catalogue. Each company can create its own catalogue. With very simple operations, such as drag-and-drop or copy-and-paste, users can place layout elements, features and equipment in the 3D

editing window on a user-defined grid, dragging them from the catalogue. They can also add columns, doors, windows, fences, etc., in order to recreate the actual available space and to organize it in the best way. The “snap to 3D grid” allows users to place equipments easily. All the components in the catalogue are parametric, enabling re-shaping at the time of placement. The 3D editing window provides advanced capabilities, such as: direct manipulations, move, alignment, mirroring, offset from object and snap in order to speed up the process of creating a first layout and rapidly change the space organization when the factory layout evolves.

Moreover, users can define the material flows and dependencies between the resources.

### 3.3. FLP Multi-touch

The multi-touch application is a customized interface of the Desktop application tailored to work with touch devices.

According to Wikipedia, “*multi-touch refers to the ability to simultaneously detect and fully resolve 3 or more distinct positions of input touches*”. In fact, this application can manage up to 32 simultaneously touches. This means that, in a very natural way, multiple users can interact on the same layout at the same time to organize better the space, to move components, to plan a new factory layout. Furthermore, the multi-touch application allows technicians, managers, salesmen and stakeholders to interact with the FLP with a very intuitive touch interface without using any traditional input device such as keyboard or mouse.

This touch interface exploits new paradigms of human-computer interaction: while single or double touch devices are becoming common nowadays, multi-user interaction managed by a multi-touch interaction is genuinely innovative and can dramatically improve the touch experience (Ramanahally et al., 2009). The main challenges to build such a user interface are:

1. Multi-user aware widgets (i.e. buttons, input boxes, check boxes, menus), because traditional GUI has a single focused component, while in the FLP, many users can play with different widgets at the same time.
2. Adjustable widgets, all the people around a table can interact from their position, so the widgets have to be represented with the right orientation. In the traditional GUI, they have only one orientation.

### 3.4. FLP Layout editing

The layout editing was designed to address the following main needs:

1. The generation of executive 2D layout drawings.
2. The smart creation of simulation models.
3. The preparation of commercial presentations.

The layout editing in the traditional CAD systems is generally targeted towards the needs of technicians to prepare the executive drawings. The FLP is meant to be used by a wider range of end users with different aims. For instance, when a user places a component on the grid, it should connect itself with the nearest components if needed, and it should communicate with the others. In this way, the material flow is generated in an automatic way: the components are animated in order to visualize their paths as well as their machining. This gives a visual cue of how the layout will work, both for technical analysis and commercial presentations.

These requirements were addressed combining several resources describing each component:

- A geometrical 3D representation of the component composed of coloured meshes and a kinematic description.
- A DXF drawing is used for the executive 2D layout, and it also provides the clutter of the foundation.
- A logical description of the component in terms of ports (inputs/outputs) and parameters.
- A program (written in JavaScript or in custom DSL under development) implementing the behaviour.

Each application can access only the required resources of a component. For example, the visualization engine can display the 3D meshes; the simulation engine can execute the behaviour program. Nevertheless, changes to one aspect of a component are reflected to the other aspects. For instance, if a component is linked to another in the 3D window, this affects DES, creating a connection between the respective input and output ports.

### 3.5. FLP Simulation

The simulation is another core application of the FLP: the basic idea is to combine 3D simulation with DES. These technologies are fundamental in planning a layout, because, at macro level, manufacturing can be represented as “material in – process – material out”. Understanding the processes and the production sequences helps to ensure efficient manufacturing and control costs. DES allows to analyze the material flow and the resources' allocation at plant level (Voorhost et al., 2008), while 3D simulation allows to optimize the layout and the relative process at workcell level (Schenk et al., 2005).

From a technological point of view, each component has a set of input and output ports that are connected to each other to define the material flow and resource dependencies. Along these ports, the components can send and receive signals that can be primitive types such as integer or double values but also complex structures. Every time a new signal is published to the output port, it is propagated to all the connected input ports. In this way, the components can

react to changes in the status. Moreover, the components' logic can schedule a task to be executed after a predetermined time delay or when a certain condition is met.

The logic that defines the behaviour of a component is handled by a program executed inside an interpreter. It gives virtual access to the complete layout and its associated resources. It is possible, for example, to change the colour of the geometry of a downtime machine, to move a beam on the roller conveyor, etc..

The program controlling the behaviour of the component can be written in JavaScript, using an API to access and define the required interfaces, or a DSL (Domain Specific Language) that is currently under development. The DSL will be similar to a state chart diagram and it will allow creating and editing the program visually. A set of predefined behaviour will be provided for common components such as sources, sinks, buffers, conveyor transports and so on.

#### 4. REAL TEST CASE

This paragraph describes a simple real test case. It is meant to test some features of the FLP prototype and to demonstrate its industrial applicability.

This test case deals with an existing woodworking plant making panel doors for the furniture industry. It is composed by two identical and parallel processing lines as shown in Figure 1.

Each line is equipped with:

- A CNC drilling and routing work centre (see Figure 2).
- A brushing machine.
- A reversing device.
- A CNC drilling and routing work centre of panel side.
- A control station.

Looking at Figure 1, the material flows in the bottom-up direction.

At the beginning and at the end of these two lines there are two Cartesian robots in charge of loading and unloading the wooden boards. Identical panel doors, requiring the same processes, are stacked up on the infeed roller conveyors. A single panel is picked by a robot and put on the processing line. Each panel is machined at the routing work centre and then it's brushed by the brushing machine. When needed, it is turned upside down and comes back to the first centre, otherwise, it goes on through the drilling and routing work centre, where the panel sides are worked. Eventually, each panel is checked at the control station and then it is unloaded by the robot and put on the outfeed roller conveyor.

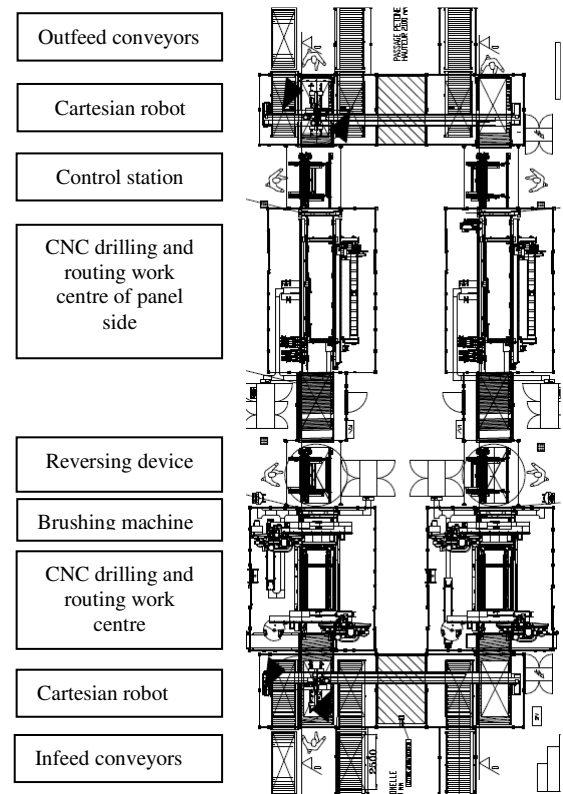


Figure 1: The layout of the plant

All the 3D models of the layout components were imported from a CAD system but their level of detail was even too high for this scope. Thus there was the need to simplify the 3D models trimming minor details such as screws, bolts, electric wires, and, sometimes, reducing the number of triangles of holes, cylinders, etc. Figure 2 shows the result of this simplification applied to the routing work centre. Unfortunately, this step cannot be completely automatized (some automated procedures are in place, but human intervention is still required) and it is very time consuming. However, it has to be done once, and then the model can be used as many times as needed in several different layouts. Eventually, the kinematics has been added to the machines and work centres whose movements have to be represented. These 3D models were saved in a proprietary XML format along with the other required information, and then imported in the catalogue of the FLP.

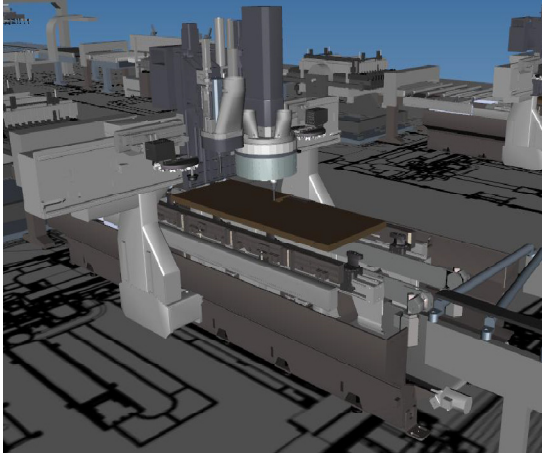


Figure 2: CNC drilling and routing work centre

The FLP, with a filled catalogue, is now ready to be tested by the plant designer along with the process manager working for the same company. Their goal is to re-create the existing woodworking plant, starting from scratch using the FLP provided with the multi touch feature. A very short training was made before starting the test case. The touch interface makes the layout creation very easy and intuitive and the team is more concentrated on the component placement than on understanding how to use the new tool.

Figure 3 shows the result of the planning of the woodworking plant.

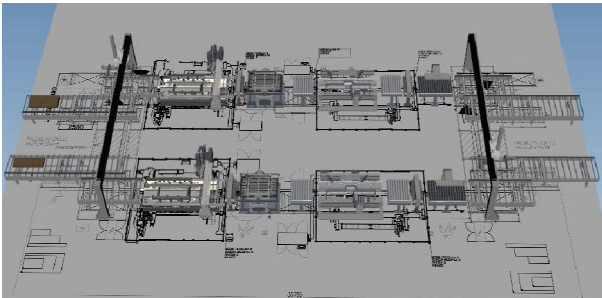


Figure 3: The final layout

The feedbacks gathered were positive in relation to the interface and the simplicity in creating a new layout and they appreciated the FLP speed during the walk-through. This test ran on a standard desktop PC, even though the total number of triangles of this plant was about 2,5 millions. Traditionally, CAD models are challenging to be visualized at interactive frame rates (~30 fps) on a normal hardware. Mesh simplification is a common used technique to reduce the number of triangles but to keep an acceptable level of constructive details only topology-preserving simplification algorithms can be applied (Luebke et al., 2002). Research specific for CAD models have highlighted that at least 2K triangles are needed to display a single mechanical component (Tang et al., 2010). Table 1 shows the number of triangles needed to visualize each component of the layout with an acceptable level of detail.

Table 1: Number of triangles for all the components of the final layout

Component	Triangles	#	Total triangles
Cartesian robot	174.624	2	349.248
Drilling and routing work centre	616.481	2	1.232.962
Brushing machine	4.732	4	18.928
Reversing device	313.744	2	627.488
Drilling and routing work centre of panel side	23.624	2	47.248
Roller conveyor_1500	6.576	4	26.304
Roller conveyor _4000	8.644	10	86.440
Roller conveyor _4000	9.272	2	18.544
Control station	9.552	2	19.104
<b>Total</b>			<b>2.426.266</b>

As far as it concerns the multi-touch interface, it was perceived as really immediate and comfortable in creating the layout. Only the view orientation lacked intuitiveness. Sometimes, it's not instantaneous to understand how to move the fingers on the monitor surface in order to get the desired 3D view. This functionality has to be further investigated in order to improve its usability.

Other concerns, arisen during this testing phase, regarded the accuracy in the component positioning and sizing.

Even though the kinematic simulation application is currently under development and thus not fully accessible for inexperienced users (as it still requires a great deal of programming), an IT expert worked it out in order to reproduce the panels' movement through the plant as well as their machining at the work centres.

The simulation of material flow and machines provides an added value, as expected, because it allows checking for bottleneck, collisions, etc., providing also some overall performances of the plant. The animation, along with the simulation results, eases the discussion about the changes to apply to the layout in order to increase its performances. An alternative solution was analyzed, as shown in Figure 4, where a new line was added just with a simple copy and paste operation.

The first impression about 3D animation was positive as well, and the company was surprised about the realistic level of the animation.



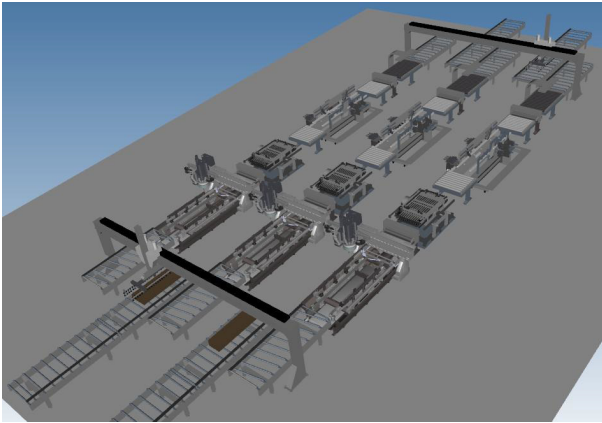


Figure 4: The revised layout

## 5. CONCLUSIONS

This paper describes the innovative Factory Layout Planner, whose main scope is to speed up and facilitate the layout planning process by combining and integrating different technologies such as 3D modelling, simulation and multi-touch interaction. Special attention was given to the interface and the processes to create and to change layouts in order to make them as easy and quick as possible. In fact, the FLP's target users are not only technicians, but also stakeholders, designer and salesmen.

The results of the simple test case were positive confirming the efficacy of the tool and its applicability in real manufacturing contexts.

Further activities will be concentrated on the development of:

- The simulation application
- A tool to drive the user in creating new components to add to the catalogue.
- Some supporting features such as: the measurement, photo realistic screenshots, etc..
- Generation of 2D plan and sections views with quotes.

## ACKNOWLEDGEMENT

This work has been partly funded by the European Commission through NMP Project DOROTHY: Design Of customeR dRiven shOes and multi-siTe factorY (No. FP7-NMP-2007-SMALL-1). The authors wish to acknowledge the Commission for their support. We also wish to acknowledge our gratitude and appreciation to all the DOROTHY project partners for their contribution during the development of various ideas and concepts presented in this paper.

## REFERENCES

Avai, A., Pietrarola, D., Pirovano, A., Aspesi, M., 2011. Design validation of an automatic and flexible steel fabrication facility using discrete event simulation. *Proceedings of Industrial Simulation Conference'2011*, pp. 153-169. Venice (Italy).

Boër, C.R., Avai A., El-Char J., Imperio E., 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 System*. Beijing (PRC).

- Ceruti, I.F., Dal Maso, G., Ghielmini, G., Pedrazzoli, P., Rovere D., 2010. Factory Layout Planner. *Proceedings of 16th International Conference on Concurrent Enterprising*. Lugano (Switzerland).
- Ding, J., Wang, Y., Chen, K., 2010. An interactive layout simulation system of virtual factory. *Applied Mechanics and Materials*, Vols 20-23, pp. 421-426.
- Fagent, P., Eriksson, U., Hermann, F., 2005. Applying discrete event simulation and an automated bottleneck analysis as an aid to detect running production constraints. *Proceedings of 2005 Winter Simulation Conference*, pp. 1401-1407. Arlington, (VA, USA).
- Fuller, R.B., 1982. Synergetics. Explorations in the geometry of thinking. Macmillan Pub Co (Ed.).
- Harrow, D. and Rosas, J., 2008. Process plant design & layout – Art or Technology ?. *Proceeding of Procemin 2008*, pp 41-54. Santiago (Chile).
- Kyle, R.G., Ludka, C.R., 2008. Simulating the furniture industry. *Proceedings of 2008 Winter Simulation Conference*, pp 1347-1350. Miami, (FL, USA).
- Luebke D.P., Reddy M., Cohen J.D., Varshney A., Watson B., Huebner R., 2002. Level of detail for 3D graphics. Morgan Kaufmann
- Mert, A., 2004. Simulation based layout planning of a production plant. *Proceedings of 2004 Winter Simulation Conference*, pp. 1079-184. Washington D.C. (USA).
- Meyers, F.E., Stephens, M.P., 2000. Manufacturing facilities design and material handling. Prentice-Hall, Inc.
- Quarterman, L., Amundsen, A., Nelson, W. and Tuttle H., 1996. Facilities and workspace design. *Engineering & Management press, Institute of Industrial Engineers*. Georgia (USA).
- Ramanahally P., Gilbert, S., Niedzielski, T., Velazquez, D. and Anagnost, C., 2009. Sparsh UI: A Multi-Touch Framework for Collaboration and Modular Gesture Recognition. *Proceedings of the World Conference on Innovative VR 2009*. Chalon sur Saone (France).
- Schenk, M., Strassburger, S., Kissner, H., 2005. Combining virtual reality and assembly simulation for production planning and worker qualification. *Proceeding of CARV 2005*, pp. 411-414. Munich (Germany)
- Tang Y.; Gu H.; 2010. CAD Model's Simplification and Conversion for Virtual Reality. *Information and Computing (ICIC)*, 2010 *Third International Conference on*, vol.4, pp.265-268, Wuxi (P.R.C.)
- Voorhorst, F.A., Avai, A. and Boër, C.R., 2008. Optimizing a highly flexible shoe production plant using simulation. *Proceedings of 7<sup>th</sup> international workshop on modelling & applied simulation*, pp. 233-239. Amantea (Italy).

## **AUTHORS BIOGRAPHY**

**Antonio Avai** is a partner and technical director of Technology Transfer System, 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 15 years of experience in discrete event simulation and its integration with other software tools and technologies, and has authored several papers on these topics.

**Giovanni Dal Maso** is a partner and technology consultant of Technology Transfer System. His interests include simulation, computer graphics, multi-touch devices, mobile and cross-platform development. He has been responsible of software development in several EU funded projects and managed the exploitation of the results into commercial products.

**Paolo Pedrazzoli** works for SUPSI, as responsible for the Bachelor of Science in Industrial Engineering and for the Master of Science in Precision manufacturing. There, he also acts as coordinator of two EU funded project. Additionally he participates in several National funded and EU funded projects. His field of interest lays in Lean Manufacturing, Assembly Systems, 3D simulation and CAx.

**Diego Rovere** works as a researcher for SUPSI. He is responsible of the development group of applied Simulation and professor in the Master of Science in Precision manufacturing. He participated to several EU founded projects managing software development. His interests are concentrated on 3D simulation engines, CN integration and real time physics simulation applied to manufacturing environments.

# RELEASE KINETICS OF CARDAMOM OIL FROM MICROCAPSULES PREPARED BY SPRAY AND FREEZE DRYING

Masoud Najaf Najafi<sup>a</sup>, Rassoul Kadkhodaei<sup>b</sup>

<sup>(a)</sup>Institute of Scientific - Applied Higher Education Jihad-e- Agriculture, Iran.

<sup>(b)</sup>Department of Food Technology, Khorasan Research Institute for Food Science and Technology, Mashhad, Iran.

<sup>(a)</sup>[masoudnajafi@yahoo.com](mailto:masoudnajafi@yahoo.com), <sup>(b)</sup>[rkadkhodaei@yahoo.com](mailto:rkadkhodaei@yahoo.com)

## ABSTRACT

The stability of encapsulated cardamom oil, which was prepared by spray and freeze drying, was studied in view of the release characteristics. HiCap 100 was used as the wall material. The powders were stored under the conditions of 45, 60 and 75% relative humidity at 25 and 50 °C. The rate of release of cardamom oil was analyzed using Avrami's equation. The results showed that spray drying was a better method to encapsulate 1,8-cineole as an indicator for the release of cardamom oil, than freeze drying. In constant temperature, the release of 1,8-cineole increased with increasing relative humidity. However, when comparing the different temperature, the release of 1,8-cineole at 50 °C was higher than that observed at 25 °C.

Keywords: cardamom oil, HiCap100, release, Relative humidity.

## 1. INTRODUCTION

Flavor plays an important role in consumer satisfaction and influences further consumption of foods. Flavor stability in different foods has been of increasing interest because of its relationship with the quality and acceptability of foods, but it is difficult to control. Manufacturing and storage processes, packaging materials and ingredients in foods often cause modifications in overall flavor by reducing aroma compound intensity or producing off-flavor components (Lubbers, Landy and Voilley 1998). To limit aroma degradation or loss during processing and storage, it is beneficial to encapsulate volatile ingredients prior to use in foods. The release characteristics of encapsulated flavors from the powder are quite important for estimating the storage period, as well as the controlled release applications in food (Reineccius 1995; Whorton and Reineccius 1995). The advantages of controlled release are: the active ingredients are released at controlled rates over prolonged periods of time; loss of ingredients during processing and cooking can be avoided or reduced; reactive or incompatible components can be separated (Dziezak 1988; Brannon-Peppas 1993). Soottitawat, Bigeard, Yoshii, Furuta, Ohkawara and Linko (2004) investigated the release characteristics and oxidation stability of encapsulated d-limonene, which gum arabic, soybean water-soluble

polysaccharide, or modified starch, blended with maltodextrin were used as the wall materials. The powders were stored under the conditions of 23-96% relative humidity at 50 °C. They have reported the release rate and the oxidation rate increased with increasing water activity, but around the glass transition temperature, the rates decreased sharply to increase again at a further increase of water activity. However, in the same manner for the shelf life of encapsulated flavor. Pena et al. (2009) have recently reported that Vanillin release from capsules was encountered to increase when increasing the stirring rate or when increasing the temperature of the release medium. Cardamom oil has a highly chemically reactive volatile compounds and unstable in the presence of air, moisture and high temperature, which has been used as a flavor. This major compounds are 1,8-cineole,  $\alpha$ -terpinyl acetate and limonene comprising two-third of the total volatiles (Prabhakaran Nair 2006). Krishnan, Bhosale, and Singhal (2005) used binary and ternary blends of gum arabic, maltodextrin, and modified starch as wall materials. They showed that the presence of gum Arabic in blend increased the protective effect of wall on core material and proposed a simple mathematical model for estimating the flavor retention. Controlled release of flavors from capsule matrices seemed to be a useful application.

In this study the encapsulation of emulsified cardamom oil by spray and freeze drying was carried out with modified starch such as HiCap 100 as wall material. The release kinetics of encapsulated powders was investigated using different method of drying and storage conditions (relative humidity and temperature).

## 2. MATERIALS AND METHODS

### 2.1. Materials

Cardamom oil and Tween 80 (GA) were purchased from Sigma-Aldrich Co., Ltd. (Germany). Modified starch (HiCap 100) were obtained from National starch (UK). The HiCap 100 was derived from waxy maize base, were modified with *n*-octenyl succinic anhydride (OSA) for use in the flavor encapsulation process. Other organic chemicals used were of analytical grade. Deionized water was used for the preparation of all solutions.

## 2.2. Preparation of spray and freeze dried powder

Wall material of HiCap 100 was added into deionized water to obtain 30% w/w mixture and allowed to hydrate overnight. After that, cardamom oil was added to solution at a mass ratio of 0.05 to wall material. The mixture was homogenized by using a rotor/stator type homogenizer (Ultra-Turrax T25, IKA, Germany) at rotational speed of 20,000 rpm for 1 min, followed by use of a 20 kHz ultrasonic processor (Vibracel 750 W, Sonics, USA) programmed to continuously insonify samples for 5 minutes at its maximum output power. The cardamom oil emulsion was fed through Büchi 190 Mini Spray Dryer (Büchi Labortechnik AG, Flawil, Switzerland), equipped with a centrifugal atomizer. The operating conditions of the spray dryer were air inlet temperature of  $180 \pm 10$  °C, air outlet temperature of  $90 \pm 10$  °C, feed rate of 10 ml/min, and air flow rate of 110 kg/h. For freeze drying of emulsions, they were introduced into large Petri dishes and frozen at -25 °C for 24 h followed by drying in a freeze dryer (Chrisa, Germany) under reduced pressure. The freeze dried mass was pulverized into a fine powder by using a pestle and mortar. Finished powder was stored in a hermetically sealed bottle at a desiccator of constant temperature and humidity to evaluate the release rate of the flavor encapsulated in powder.

## 2.3. The stability of encapsulated cardamom oil

Approximately one tenth of a gram of the encapsulated powder was weighed and spread in a thin layer glass plate, and placed in a desiccator. The relative humidity (RH) inside the desiccator was constant of 45, 60 and 75 RH at 25 and 50 °C. Constant RH was created using saturated salt solutions of magnesium nitrate for 45% RH, strontium chloride for 60% RH, and sodium chloride for 75% (Rockland 1960). The desiccator was sealed and held in an air bath to keep the temperature as above to reach equilibrium. Sample plates were placed in a desiccator to study the release and oxidation kinetics for 25 days. At fixed time intervals, the bottles were removed from the desiccator to residual amount of 1, 8-cineole in the powder was measured by the solvent extraction method described below. The retention of 1, 8-cineole in the releasing experiment was expressed as the ratio to the initial one. Avrami's equation, also called Weibull distribution function, which was successfully applied to the release time-courses of the encapsulated flavors (Sootittantawat et al., 2004; Yoshii et al., 2001) was also employed in this work.

## 2.4. Extraction of cardamom oil from the Powder

0.1 gram of powder was dispersed in 4 ml of water in a glass bottle, and then 2 ml of hexane was added, followed by forceful mixing with a vortex mixer for 1 min. To extract encapsulated cardamom oil into the organic solvent, the mixture was heated in a heating block with intermittent shaking at 45 °C for 20 min. Then, after cooling down, the extracted mixture was then centrifuged at 4000 rev/min for 20 min to separate

the organic phase from water. The organic phase solution left from above was dispersed in hexane; volume made to 10 ml in a standard volumetric flask and used to estimate 1,8-cineole by taking absorbance at 270 (Krishnan et al. 2005). This wavelength was found to correspond to the maximum absorbance of 1,8-cineole over the spectrum of wavelengths from 200 to 600 nm. A standard curve was obtained by measuring absorbance of 1, 8-cineole mixed in hexane at various concentrations (w/w). A blank sample of hexane was used in the spectrophotometer to prevent the effect of absorbance by hexane. All of the samples were analyzed in duplicate and the data were presented as an average.

## 2.5. Statistical analysis

All experiments were carried out based on fully factorial design and the results represent the mean of at least two replicates. General Linear Model of MINITAB (Version 14, 2004) was used for performing analysis of variance (ANOVA) to determine the difference between treatments at the significance level of  $P < 0.05$ . Duncan's multiple range test was employed to investigate the significant difference between treatments means at the probability levels of  $P < 0.05$  and  $P < 0.01$ . The graphs were drawn by Microsoft Excel 2003.

# 3. RESULTS AND DISCUSSION

## 3.1. Retention of flavours during spray and freeze drying

The retention of 1,8-cineole was affected by the method of the drying. The results showed that higher final retention of 1,8-cineole could be achieved by spray drying, indicating that this spray drying was a better method to encapsulate 1,8-cineole than freeze drying. Retention of 1,8-cineole during spray and freeze drying were 0.95 and 0.86, respectively. Rosenberg, Kopelman and Talmon (1990) observed similar results at a higher load of ethyl butyrate. They observed at spray-dried powder, an early crust formation on the surface of the droplet resulted in entrapping much of the flavor inside the droplet.

## 3.2. Influence of relative humidity on the release of 1,8-cineole from the powder

The release time-courses of 1,8-cineole from the spray and freeze-dried powder were measured at 25 and 50 °C and at 45, 60, and 75% relative humidity (RH). The effects of RH on the release of 1,8-cineole during storage are shown in Figures 1 and 2. In these Figures, the relative humidity greatly affected the release rate of 1,8-cineole. The dependence, however, was not simple. In constant temperature, the release of 1,8-cineole increased with increasing RH. However, when comparing the different temperature, the release of 1,8-cineole at 50 °C was higher than that observed at 25 °C. These results suggested that the release of 1,8-cineole was closely related at least to the water activity of the powder, which is in agreement with many earlier

observations. The loss of 1,8-cineole during storage may be caused by two mechanisms: diffusion of 1,8-cineole through the matrixes of the wall material and the oxidation of 1,8-cineole. However, as will be mentioned in the following section, the loss by oxidation was at most 5-6% of the initial 1,8-cineole content. Therefore, the release of 1,8-cineole may result mainly from diffusion as suggested by Whorton (1995). However, when comparing the two methods of drying, the release rate was observed to be slower for spray drying in this study.

### 3.3. Analysis of the release rate by avrami's equation

To evaluate the release rate constant of 1,8-cineole in the powder, Avrami's equation (eq 1) was applied to the release time-courses of the encapsulated 1,8-cineole as reported by yoshii et. al (2001) and soottitantawat et. al (2003),

$$R = \exp[-(kt)^n] \quad (1)$$

where  $R$  is the retention of 1,8-cineole,  $t$  is the storage time,  $k$  is the release rate constant, and  $n$  is a parameter representing the release mechanism. The release time-courses could be correlated well with Avrami's equation, as shown by the solid line in Figures 3 and 4. Eq (1) is also called the Weibull distribution function, which is successfully applied to describe the shelflife failure (Gacalar and Kubala 1975). As can be seen from Figures 3 and 4 the relative humidity had a pronounced effect on the release of 1,8-cineole, which was greatly accelerated by an increase in relative humidity. This suggested that the release of ethyl butyrate was closely related to the presence and concentration of water molecules surrounding the powder. Whorton and Reineccius 1995 have reported similar results for several esters. Also, for encapsulated powder, the release of 1,8-cineole markedly influenced by drying method. There was appreciable increase in the release rate when the freeze drying was used. On the other hand, freeze-dried powder was more lost at relative humidity in storage time than spray-dried powder. Spray drying was a better-controlled release agent than freeze drying.

Sun and Davidson (1998) successfully applied Avrami's Eq. (1) to analyze the time-dependent protein inactivation in amorphous sucrose and trehalose matrices. As is easily recognized,  $n=1$  represents the first order reaction. In case of solid-gas reaction, if the molecular diffusion of the reactant and product compounds through the ash layer is rate limiting,  $n=0.54$ . Taking double logarithm of both sides, Eq. (1) yields Eq. (2):

$$\ln(-\ln R) = n \ln k + n \ln t \quad (2)$$

From Eq. (2). one can get the parameter  $n$  as a slope by plotting  $\ln(-\ln R)$  vs.  $\ln t$ , and the release rate constant  $k$  from the interception at  $\ln t=0$ .

The release time-course of 1,8-cineole at different conditions were analyzed by plotting  $\ln(-\ln R)$  against  $\ln t$ . Figures 4 and 5 shows an example of the analysis by Eq. (2). As can be seen, the release of encapsulated powder prepared by spray and freeze drying fitted well to Eq. (2). As compared with the release rate at RH of 45%, the higher rate was observed at around RH of 75%.

This implies that the slower release in the low water activity region is most likely due to the lower mobility of 1,8-cineole molecules in the glassy state of the capsule matrixes (Whorton 1995). When RH increased up to a value of around 75%, the powders began to be rehydrated. At this stage, it may be assumed that the effective surface area decrease resulted in a decrease of 1,8-cineole evaporation from the surface of the powder particles. Most particles were detected to be clumped and adhered together into a paste like mass, which explains the rubbery form of the capsule matrixes. A similar speculation was also proposed by Whorton and Reineccius 1995.

## 4. CONCLUSION

The retention of 1,8-cineole was influenced by the type of drying method. spray drying was found to be a superior method over freeze drying to retain 1,8-cineole during storage. The release time-course of 1,8-cineole fitted well to the Avrami's equation. Inspecting the results obtained for the effect of temperature and moisture content on the release of 1,8 cineol indicated that at water activities and all temperatures under question, the spray-dried microcapsules showed slower release rates than the freeze-dried ones.

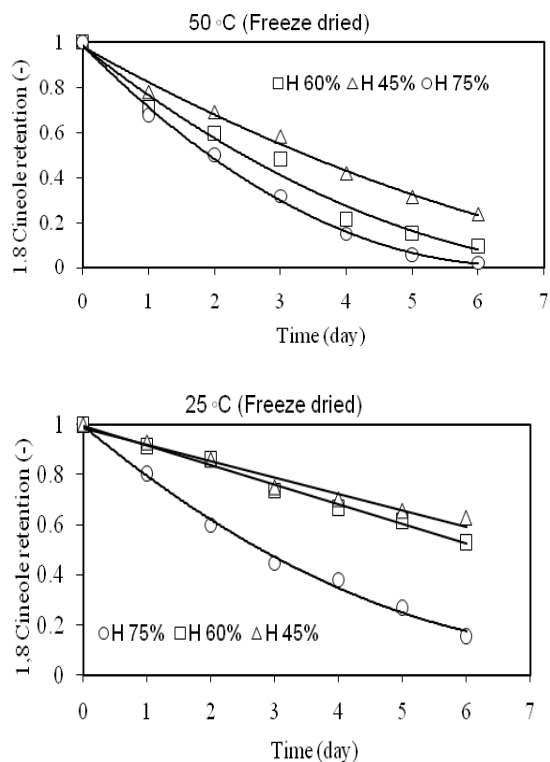


Figure 2: Effects of relative humidity and storage temperature on the release of 1,8-cineole from freeze-dried powder.

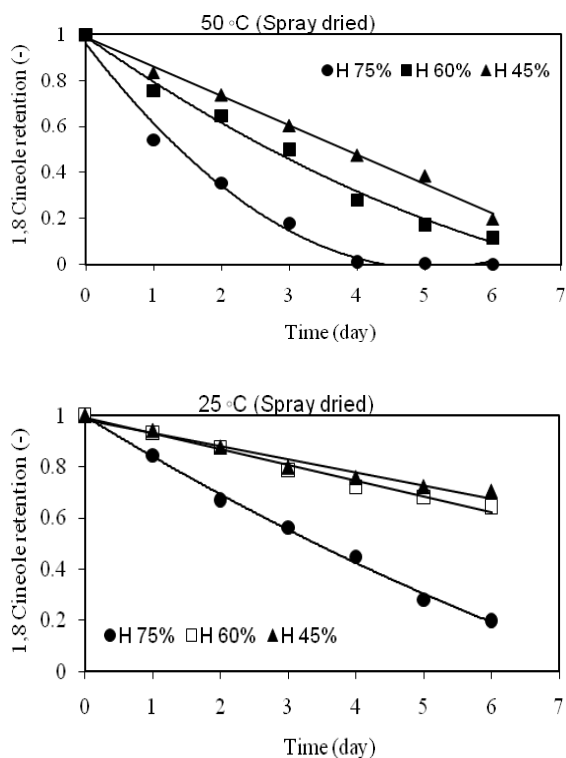


Figure 2: Effects of relative humidity and storage temperature on the release of 1,8-cineole from spray-dried powder.

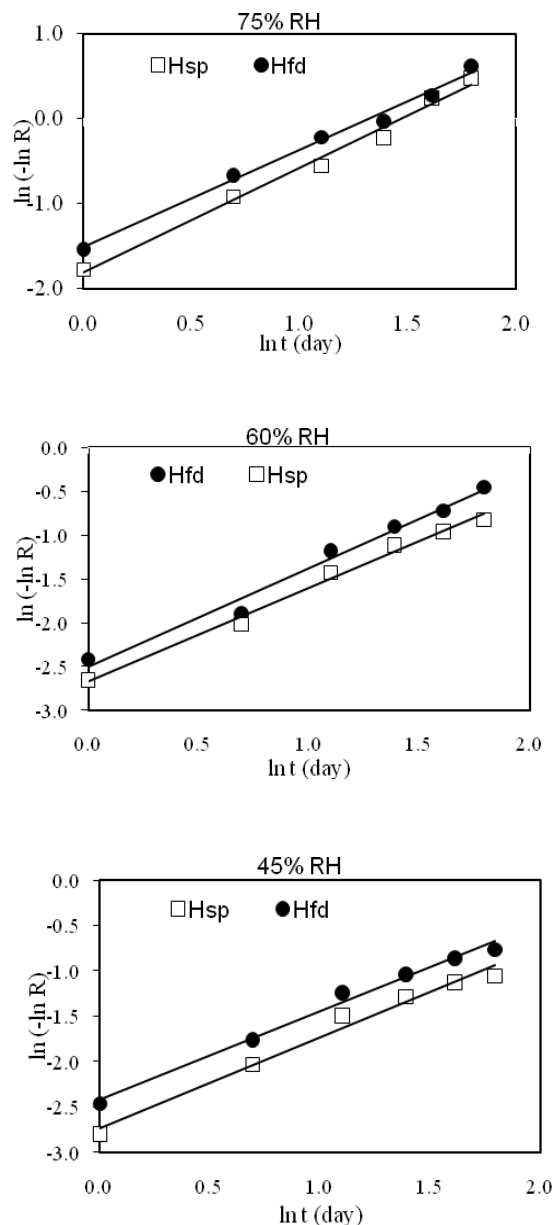


Figure 3: Correlation of release time-course of 1,8-cineole by Avrami's equation at 25 °C for spray (sp) and freeze-dried (fd) powders.

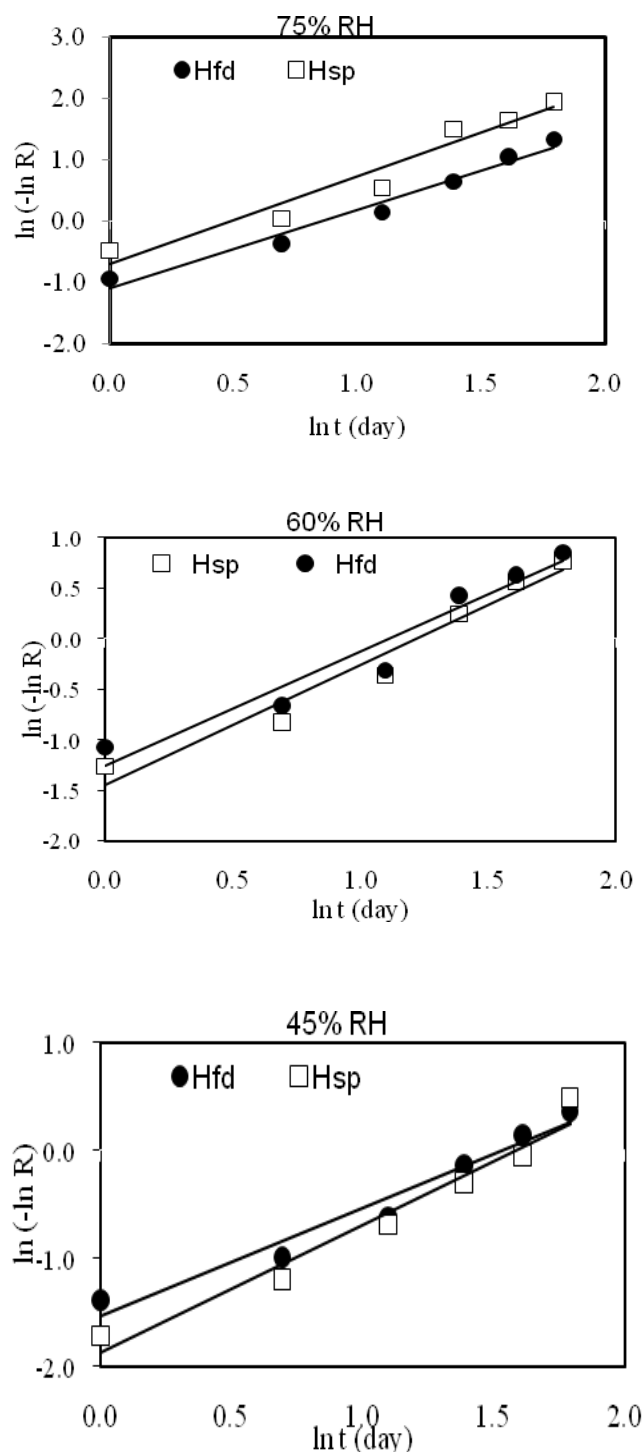


Figure 4: Correlation of release time-course of 1,8-cineole by Avrami's equation at 50°C for spray (sp) and freeze-dried (fd) powders.

## REFERENCES

Brannon-Peppas, L., 1993. Properties and applications. In: Polymeric Delivery Systems (edited by M.A. El-Nokaly, D.M. Piatt & B.A. Charpentier), Pp. 52. ACS Symposium Series 520. Washington, DC: American

Chemical Society. crystallization. *Biochimica et Biophysica Acta* 1425, 235-244.

Dziezak, J.D., 1988. Microencapsulation and encapsulation ingredients. *Food Technology*, 42, 136-151.

Gacalar, M. C. Jr.; Kubala, J. J., 1975. Statistical models for shelf life failures. *Journal of Food Science*. 40, 404-409.

Krishnan, S., Bhosale, R., and Singhal, R. S., 2005. Microencapsulation of cardamom oleoresin: Evaluation of blends of gum arabic, maltodextrin and a modified starch as wall materials. *Carbohydrate Polymers*, 61, 95-102.

Lubbers, S., Landy, P and Voilley, A., 1998. Retention and release of aroma compounds in food containing proteins. *Journal of Food Technology*, 52, 68-74.

Pena, B., Casals, M., Torras, C., Gumi, T., and Garcia-Valls, R., 2009. Vanillin Release from Polysulfone Macrocapsules. *Industrial Engineering Chemistry Resereach*, 48, 1562-1565.

Prabhakaran Nair, K. P., 2006. The agronomy and economy of Cardamom (*Elettaria Cardamomum* M.): The "Queen of spices". *Advanced agronomy*, 91, 179-410.

Reineccius, G.A., 1995. Liposomes for controlled release in the food industry. In: Encapsulation and Controlled Release of Food Ingredients (edited by S.J. Risch & G.A. Reineccius). Pp. 113-131. ACS Symposium Series 590. Washington, DC: American Chemical Society.

Rockland, L. B., 1960. Saturated salt solution for static control of relative humidity between 5 and 50 °C. *Analytical Chemistry*. 32, 1375-1376.

Rosenberg, M., Kopelman, I.J. and Talmon, Y., 1990. Factors affecting retention in spray-drying microencapsulation of volatile material. *Journal of Agriculture and Food Chemistry*, 38, 1288-1294.

Soottitantawat, A., Bigeard, F., Yoshii, H., Furuta, T., Ohkawara, M., Linko, P., 2004. Effect of Water Activity on the Release Characteristics and Oxidative Stability of D-Limonene Encapsulated by Spray Drying. *Journal of Agriculture and Food Chemistry*, 52, 1269-1276.

Sun, W. D., and Davidson, P., 1998. Protein inactivation in amorphous sucrose and trehalose matrices: effect of phase separation and crystallization. *Biochimica et Biophysica Acta* 1425, 235-244.

Whorton, C. and Reineccius, G.A., 1995. Evaluation of the mechanisms associated with the release of encapsulated flavor from maltodextrin matrices. In: Encapsulation and Controlled Release of Food Ingredients (edited by S.J. Rish & G.A. Reineccius). Pp. 143-160. Washington, DC: American Chemical Society.

Whorton, C., 1995. Whorton, C.Ž1995.. Factors influencing volatile release from encapsulation matrices. S. J. Risch & G. A. Reineccius, Encapsulation and Controlled Release of Food Ingredients. ACS Symp. Ser. No.590, 134-142.



Yoshii, H., Sootitawat, A., Liu, X.D. et al., 2001. Flavor release from spray-dried maltodextrin/gum Arabic or soy matrices as a function of storage relative humidity. *Innovative Food Science and Emerging Technologies*, 2, 55–61.

# GENETIC ALGORITHM APPROACH TO MODELLING FRACTAL MANUFACTURING LAYOUT

Julian Aririguzo C<sup>(a)</sup> and Sameh M Saad <sup>(b)</sup>

<sup>(a)(b)</sup>Faculty of Arts, Computing, Engineering and Sciences  
Sheffield Hallam University  
Sheffield, UK

<sup>(a)</sup>J.Aririguzo@shu.ac.uk <sup>(b)</sup>S.Saad@shu.ac.uk

## ABSTRACT

Fractal Manufacturing System (FrMS) basically structurally builds up from units called 'fractals' or fractal objects which are independent entities and contain essential features and congenital attributes of the entire manufacturing configuration. They can self-adapt quickly to dynamic changes in an unpredictable manufacturing environment. They are also self-regulating and fall under organizational work groups, each within its own area of competence. An optimal shop floor design and implementation is key and an integral part of achieving a successful FrMS. and is concerned with issues of shop floor planning, arrangement and function layout. The fractal shop floor layout develops from individual cells and is conceptually capable of producing a variety of products with minimal reconfiguration. Keen attention is paid to determination of capacity level, cell composition and flow distances of products. In this paper, Genetic Algorithm (GA) is adopted to efficiently and effectively design flexible FrMS shop floor layout, needed in agile manufacturing system to cope with new and dynamic manufacturing environments that need to adapt to changing products and technologies. Its stochastic search algorithm is used in simulating natural evolutionary process techniques, which in turn solves the many FrMS combinatorial optimization problems. The design implementation is done using MATLAB. The end result interestingly is a fault tolerant structure that is better suited to survive and stand the pressure for lead time reduction and inventories, product customization and challenges of a dynamic and unpredictable operational environment.

Keywords: Fractal manufacturing system, Manufacturing layout design, Genetic algorithm.

## 1. INTRODUCTION

The conceptual fractal shop floor builds up from individual cells called fractals and is capable of producing a variety of products with minimal reconfiguration (Venkatadri *et al.* 1997; Montreuil *et al.*, 1999). This is due mainly to their ability to self-adapt quickly to dynamic changes in an unpredictable manufacturing environment. The fractal layout is an extension of the cellular layout (Askin *et al.* (1999) and in fact, each fractal cell is a multifunctional mini shop (Venkatadri *et al.*, 1997) since it could produce most of the product types routed to it and have layout

specification that produce varied products. This decentralized production layout allows for flexible mass customization. However, there are many challenges posed by the design and implementation of this strategy. A design and simulation of the model of shop floor layout for FrMS is presented in this paper paying attention to determination of capacity level and cell composition using genetic algorithm approach. The procedure is based on an iterative algorithm, implemented using MATLAB and used to calculate material travelling distances for each fractal cell and this continuously optimizes the layout, flow assignment and improves the overall performance of these parameters to create maximum space utilization. The rest of the paper is organized as follows; section two briefly looked at the general fractal manufacturing layout, section three made an overview of the GA and MATLAB and sets the scene for the application of the GA to the layout design and experimentation. Section four discusses the essentials of the proposed fractal manufacturing layout. Section five implements the GA approach, while section six discusses output results and the paper is finally concluded in section seven.

## 2. FRACTAL MANUFACTURING LAYOUT

The fractal workstation layout is created to minimize the capacity requirements and material travelling distances (Saad and Lassila 2004). The layout design concerns the arrangement of physical production resources within the production facility (Chase and Aquilano 1992) and the planning of which involves the determination and allocation of the available space to a given number of resources (Azadivar and Wang 2000) and emphasizes minimization of flow distances in order to improve product flow and general layout performance (Montreuil *et al.*, 1999). This involves various aggregate steps; capacity planning, fractal cell creation, flow assignment and cell/ global layout (Venkatadri *et al.* 1997). These processes can significantly affect the efficiency of the planned manufacturing system in terms of shop floor control, equipment utilisation, materials handling, materials management, and worker productivity (Co and Araar 1988). The manner and nature of the flow of materials through the facility is of crucial importance and this includes the flow rate, throughput time of products and the routes taken by the products (Wild 1993).

## 2.1. Capacity Planning

The consideration of the resource requirements, demand, capacity, work methods, handling and movement, departmental area requirements, and shape and location restrictions are all issues of capacity planning (Wild 1993, Gau and Meller 1999). The general and perhaps most important objectives are how to minimize the physical movement and handling of materials, maximize the capacity utilisation (Wild 1993) and ensure a smooth work flow (Chase and Aquilano 1992) in accordance with the system plan. Other very important issues include; systems design, machine reliability, parts scheduling, etc. These are all issues involved in the capacity planning process. It is worth mentioning that the capacity planning task requires optimal value of input data to satisfy product demand, minimize investment and operations cost and go into production within the pre-specified production time, though cost of material transfer could be traded-off against initial investment cost (Montreuil *et al.*, 1999). This study first determines the required capacity levels for each machine type and the number and composition of fractal cells. Then an iterative algorithm continuously optimizes the layout and flow assignment according to the performance of the system.

## 2.2. Fractal Cell Creation

The fractal layout system uses cells to group machines together and to control and limit product routings. The number of fractal cells and workstation composition of each cell is significant in the overall manufacturing system. Its layout can be seen as an extension of a cellular layout (Askin *et al.* 1999) due to the structure of the shop floor, and fractal cells are multifunctional and are able to process most of the product types routed into the system. Each cell needs to contain exactly one replicate of workstation type. They also share workstations, but each cell has equal compositions. These identical cells are flexible and standardized. They can respond well to short term changes, uncertainties and unpredictable incidents or events such as machine breakdown, product mix, and transfer devices going offline (Montreuil *et al.*, 1999).

## 2.3. Flow Assignment

The fractal cell, a set of neighboring workstations is the basic unit in the fractal layout system (Venkatadri *et al.* 1997 & Montreuil *et al.* 1999). The system flexibility is believed to increase because all fractal cells have roughly the same composition of machines and are capable of processing most of the products routed to them. It helps to alleviate flow congestion of products and improve flow efficiency. The flow score is measured and analyzed in order to estimate the frequency and distance travelled. The satisfactory estimation of flow around the actual workstations is also of significance in the layout design. Flow assignment involves the decision of getting the products processed through particular machines on the job shop (Askin *et al.* 1996 & 1999). The assignment of products

to flow paths minimizes travel distance if there are several products with specified machine type routing to be processed (Venkatadri *et al.*, 1997). The objective of the flow assignment is to create a workstation layout that minimizes the capacity requirements and material travelling distances for a particular product. The flow assignment experiment and capacity analysis can be used to improve the layout repeatedly until a satisfactory layout is generated (Montreuil *et al.*, 1999). The results indicate that unrestricted product flows offer the best flow scores in a fractal layout.

## 3. OVERVIEW OF GA AND MATLAB

Many combinatorial optimization problems in manufacturing systems are very complex and can not be solved using conventional optimization techniques (Kamrani and Gonzalez 2003). Most fractal manufacturing layout problems are dynamic - they change with time, and the system is expected to self-adapt to unpredictable changes and uncertainties. To deal with such problems efficiently and effectively, different fault tolerant structures and adaptable methods are required in solving problems in these dynamic operational environments.

### 3.1. GA procedure

GA is an evolutionary algorithm, a simulation of natural evolutionary process technique that has been adopted for this study (Azadivar and Wang 2000). The GA approach is a powerful and broadly applicable stochastic search technique used to find approximate solutions in optimization problems (Holland 1975). Just like evolutionary algorithms, it allows systems to self-adapt to make up for unpredictable changes in the operational environment. It would take into account a wider range of possible solutions and further increase the probability of finding optimal solution by continuously iterating and optimizing the design of the fractal layout and flow assignment according to the performance of these parameters.

#### 3.1.1. Genetic Operators

Crossover and mutation are the two genetic operators that are applied probabilistically to create a new population of individual strings (Rajasekharan, 1998). Crossover is an important operation performed by GA for solving combinatorial optimization problem. Two of the individual strings in the initial population are selected randomly as two parents. A cut point is randomly chosen within the parent strings (Kamrani and Gonzalez 2003).

#### 3.1.2. Crossover

Crossover operation exchanges cross sections of the parents in order to form two offspring. As shown in (Figure 1), the two off-springs form new individual strings generated by combining the “head” of the first parent string with the “tail” of the second parent string and vice versa (Rajasekharan, 1998). The essential characteristic of crossover is the crossover rate (CR)

which is the ratio of number of off-springs produced in each generation to the population size. A higher CR allows deeper exploration of solution space and increases the chance of achieving accurate optimal results. If the CR is too high, it results in wastage of computational time (Kamrani and Gonzalez 2003).

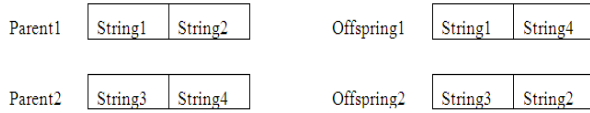


Figure 1: Crossover (Al-Sultan *et al.*, 1997)

Due to the unique hierarchical chromosome scheme used, a one-point crossover is used as in (Xiaodan Wu *et al.*, 2007). A cut point is randomly selected over the whole chromosome as shown in (Figure 2). Parent1 and Parent2 are the chromosome pair selected for the crossover operation. The “head” of Parent1 is replaced by “tail” of Parent2. Then Child1 is generated. On the other hand, the “tail” of the Parent1 replaces the “head” of the Parent2, and Child2 is then created.

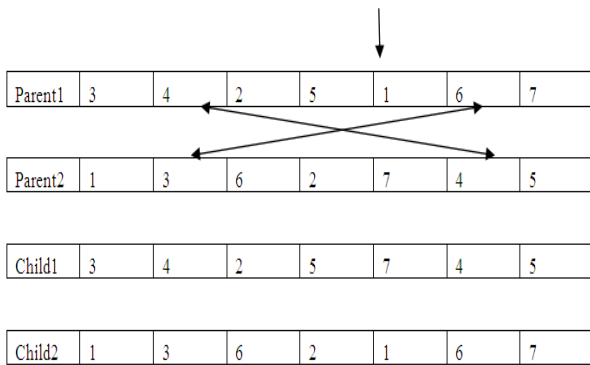


Figure 2: Numerical illustration of Crossover (Xiaodan Wu *et al.*, 2007)

### 3.1.3. Mutation

Mutation operation produces spontaneous random changes in certain chromosomes. Mutation play two roles that involve either replacing the genes lost from the population during the selection process, or providing the genes that were not present in the initial population (Kamrani and Gonzalez 2003). Mutation is designed to prevent premature convergence and to explore a new solution space (Xiaodan Wu *et al.* 2007). But the mutation operation alters and mutates one or more genes within the chromosomes of an individual rather than across a pair of chromosomes. There are two kinds of mutation proposed by (Xiaodan Wu *et al.* 2007), which are group mutation (Figure 3) and inverting mutation (Figure 4). Group mutation is for exchanging genes of the same group for the same layer at same time while inverting mutation involves exchanging the genes from the randomly chosen loci of the parent. Both genes are chosen randomly for the operation of mutation. From a theoretical perspective, if the length of the chromosome for inverting mutation is long, the chances of finding the optimal solution in the near-optimal area is low. However, the group mutation can help to enhance the GA's ability of exploiting and

converging rapidly to a promising region (Xiaodan Wu *et al.*, 2007). (Al-Sultan and Fedjki 1997) illustrated in (Figure 3) that group inverting mutation begins with a selection of a parent, and randomly dividing into two strings. The two strings are then exchanged to get a new offspring. Group inverting mutation involves two steps - a random cut of the selected parent is generated and the two chosen strings are then exchanged to obtain a new offspring.

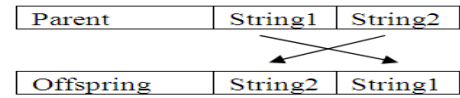


Figure 3: Group Mutation (Al-Sultan and Fedjki 1997)

There is yet another kind of inverting mutation by (Hicks 2006), that involves the selection of the two points randomly and then the genes between those points are placed in reverse order. This inverting mutation is shown in (Figure 4). The other genes in other positions are also copied directly from the parent to the child. In an inserting mutation, a gene is selected at random. The gene is taken off from the chromosome and then inserted back in a random position (Parames, 2001).

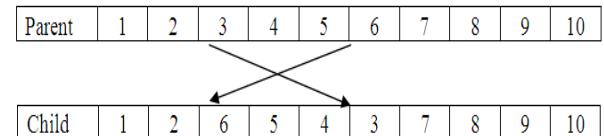


Figure 4: Inverting Mutation (Azadivar and Wang 2000)

### 3.1.4. Stopping Criteria

Two stopping conditions are employed to stop the GA from iterating continuously (Parames, 2001). First, if the number of iterations exceeds the predefined fitness value, GA would stop the operation immediately. The other stopping condition is that the value of the objective function does not change within the expected number of iterations. Once the algorithm has completed the given number of generations, the best value of the objective function is obtained. At that moment, GA would be terminated and display the layout configuration associated with the chromosome with the highest fitness value.

### 3.2. MATLAB R2008a

MATLAB is a high-level computer language for scientific computing and data visualization built around an interactive programming environment (Kiusalaas 2005). It integrates computing, visualization and programming in one user-friendly environment. In the course of this study, MATLAB R2008a is used to develop, implement, customize and create a user-friendly graphic interface (Kiusalaas 2005) for our fractal layout design. Its high-performance language essentially expresses the problems and solutions in mathematical notation. The model is then designed to fetch input of pair-wise comparison data of different

criterion and alternatives and process these data to an output of optimum score of the alternatives.

A set of general procedures are employed in the design of the fractal shop floor layout. There are several phases to the procedure;

1. Design and simulate the model of FrMS shop floor layout using MATLAB R2008a, determining the machine types and machine routing sequence.
2. Write MATLAB programming codes to reflect minimization of material travelling distances or flow distance score.
3. Apply Genetic Algorithm to continuously iterate and optimize the design of fractal layout and flow assignment according to the performance of the parameters.
4. Quality of the resulting layout is assessed and compared against Fractal cell layout according to (Venkatadri *et al.*1997)

#### 4. THE PROPOSED FRACTAL MANUFACTURING LAYOUT DESIGN

As a guide, the initial fractal manufacturing layout adopted for this study is the cellular manufacturing systems configuration proposed by (Co and Araar 1988) (Figure 5). This layout is re-designed and reconfigured from its initial cellular manufacturing layout using the GA optimization technique. Limitations of the cellular manufacturing layout include inflexibility due to a fixed set of part families, limited allowance for inter-cell flows, and long product life cycles which makes it incapable of performing in unstable environments. It also contains different types of machines which increases the product inter-cell and intra-cell travelling distances.

The design by (Co and Araar 1988) is modified and illustrates the process of constructing a fractal job shop. The example presented is a job shop with 15 distinct product types and 10 types of machine in the initial cellular layout. A total of 64 workstations are proposed by (Co and Araar 1988) in the 6 cells modified group layout design within a factory. But, each group cell contains uncertain numbers of machines. Montreuil *et al.* (1999) propounds that the grouping procedure implements a multi objective mathematical programming formulation with few surrogates;

- Minimize the difference between the assigned workload and capacity available.
- Maximize the number of products that are completed in each cell.
- Maximize the number of cells.

But, it is found that the objectives above are conflicting. The design for the group layout makes the job shop appear very much like a flow shop. But the group layout design suffers from the major disadvantage that requires too many workstation replicates (Montreuil *et al.*, 1999).

In this study, the GA approach lets us represent the entire group layout proposed by (Co and Araar 1988) as chromosomes. The modified group layout by (Co and Araar 1988) is shown in (Figure 5). MATLAB programming has made the representation of the machines in each cell easier. For instance, Cell1 can be represented as (1 5 2 6; 7 4 3 8; 9 10 3 5; 2 10 8 6; 1 5 9 10) in MATLAB codes. Cell4 is coded as (3 9 2 8; NaN NaN NaN 5) (where NaN means Not-a-Number in computing). Cell1 and Cell4 are combined using crossover operations. After the crossover, Cell1 is re-generated and it becomes one of the output cells for fractal manufacturing layout.

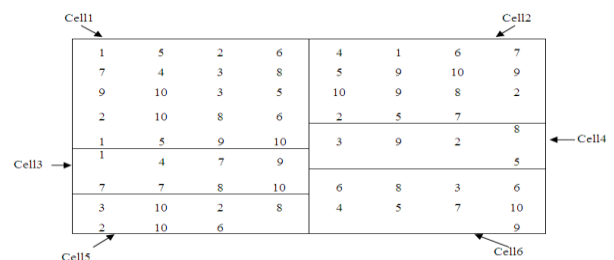


Figure 5: Modified group layout from (Co and Araar 1988)

The fractal manufacturing cell layout proposed by (Co and Araar 1988) has a number of characteristics and is shown in (Figure 6). All fractal cells are similar and contain roughly the same composition of machines. Similarity of fractal cells in terms of machine types and quantities enable high efficiency in controlling shop floor, high operational flexibility and high flexibility for factory expansion. Moreover, all fractal cells are independent and are also capable of processing all products routed to them. Furthermore, products are distributed evenly among fractal cells.

The design of fractal layout (Figure 6) contains three cells. This choice leads to a cell population of 10 workstations, which is within tractable standards of 5 to 15 machines in each fractal cell. It is not necessarily to limit the number of workstations to 30 machines in this case (Venkatadri *et al.*, 1997). But, by adding few more workstations congestion could be alleviated and flow efficiency could further be improved. Therefore, it is logical and reasonable to increase number of Machine 7 in the following approach in the fractal manufacturing layout that is proposed by Venkatadri *et al.* (1997) and Montreuil *et al.* (1999).

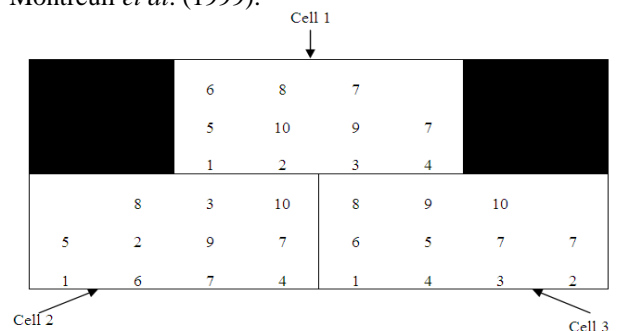


Figure 6: Fractal Manufacturing Layout

#### 4.1. Design Parameters

It is estimated that 10 types of machines are required in the fractal job shop. Machine requirement planning represents the beginning of the fractal layout process. This is carried out by computing the total number of hours required for processing the product demand (Montreuil *et al.*, 1999). There are 15 types of products that are required to be processed in the 3 fractal cells. Based on bottleneck analysis, the total demand for the fractal layout is estimated to produce 400 products that can be processed in the fractal system without violating aggregate capacity constraints and respecting product demands. The other design parameters that are used for the fractal layout modeling have to be defined and calculated as below:

Machine types in fractal job shop = 10

Product types in fractal job shop = 15

Total number of products demand = 400 products

Demand for fractal job shop =  $400/15 = 26.67$

Total machine processing times = 1108 minutes = 18.47hours

Machine processing times for processing the demand  
= 18.47hours x 26.67  
= 492hours

Total machine capacity (available hours) is 1297hours

Minimum number of machine required for fractal cell

= Machine capacity ÷ Machine processing times

= 1297 hours ÷ 492 hours

= 2.6 machines = 3 machines

Fractal decomposition is carried out using the procedures outlined in the section on cell creation design. The results of the calculation are shown on (Table 1). It can be shown that 3 machines are required for the 3 fractal cells. Therefore, it is feasible for each type of machine to be replicated or regenerated 3 times. The expected fractal layout contains 30 machines where each fractal cell has 10 machines.

Table 1: Number of replicates for fractal cell layout

Machine Type	1	2	3	4	5	6	7	8	9	10	Total
Number of Replicates	3	3	3	3	3	3	3	3	3	3	30

#### 4.2. Input Data

Tables (2, 3 & 4) contain the input data for the machine processing sequence, process times of products in minutes and machine capacity in hours respectively for each of the replicates (Co and Araar 1988). These data are written in Microsoft excel file and imported into MATLAB programming for the optimization process.

Table 2: Machine routing sequence for 15 types of product

Product Type	Machine Processing Sequence									
1	1	4	7	3	10	8				
2	3	9	2	8	5	6				
3	2	3	4	5	9	10				
4	1	7	8	10	2	3				
5	5	6	8	1	4	7	9			
6	5	2	6	4	1	7				
7	6	4	5	7	10	9				
8	1	3	5	6	8	10				
9	3	4	2	1	5	9	10			
10	8	10	2	4	6					
11	3	1	9	5	7					
12	1	9	10	2	7	8	3			
13	4	3	10	2	8	6				
14	4	2	8	5	1	6				
15	1	5	2	6	8	3	4	7	9	10

Table 3: Machine processing times for 15 types of product

Product Type	Machine Processing Times (Minutes)									
1	10	7	20	15	8	17				
2	10	15	15	15	10	5				
3	11	13	20	15	12	10				
4	9	17	9	8	10	20				
5	9	7	7	15	15	12	9			
6	7	6	13	10	8	8				
7	7	13	12	19	14	13				
8	12	11	18	11	13	10				
9	6	9	8	17	20	12	13			
10	12	18	7	5	6					
11	13	12	9	8	11					
12	7	13	17	6	11	12	5			
13	13	20	5	15	12	17				
14	7	12	20	9	18	8				
15	20	12	13	13	13	5	7	20	7	5

Table 4: Machine capacity for each replicate

Machine Type	Machine Capacity (Hours) for each replicate									
1	25	15	10	30						
2	16	29	15	25	30	20	28			
3	17	15	40	30	10					
4	18	19	17	28						
5	15	20	30	20	20	20	30			
6	18	20	15	15	10	15				
7	10	20	20	10	15	20	15	15	15	10
8	20	20	15	15	10	10	10			
9	18	17	20	30	40	30	20	17		
10	20	10	10	10	30	30	30	15	15	

#### 4.3. MATLAB dialog box

A dialog box is created as an interaction tool on MATLAB. The dialog box pops up to request for input data as shown on (figures 7, 8 & 9). These data are used to verify; the location of Microsoft Excel input file, sheet name of product sequence that is required for the modeling operation, and the sheet name of machines in fractal cell layout; the desired number of machines that are needed; number of rows and columns for each pair of initial cells that are required to generate each fractal cell; the cells required for crossover operation and the desired number of iterations needed for generating the final fractal manufacturing layout.

The input dialog box (Figure 7) for file location and sheet name in Microsoft Excel has been used to

ensure the location of the input data is identified and verified. The input dialog box (Figure 8) for desired number of cells is used to insert the number of cells that are required for the initial cell layout. The input dialog box (Figure 9) is for the number of iteration needed to determine number of replicates and analyze the output of the flow distance score.

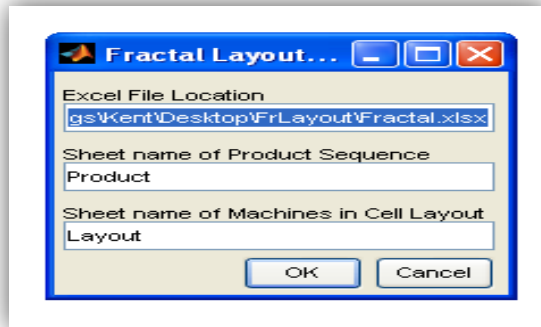


Figure 7: Input dialog box for file location and sheet name in Microsoft Excel

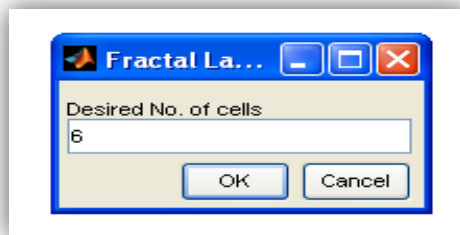


Figure 8: Input dialog box for desired number of cells

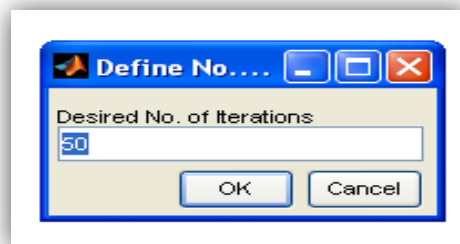


Figure 9: Input dialog box for number of iteration

#### 4.4. Facility layout problem (FLP)

The FLP is defined as “the determination of the relative locations for, and the allocation of the available space among a number of workstations” (Azadivar and Wang, 2000). The resources could be different sizes and the interactions between resources may vary. This is a major concern in developing a block layout that represents an optimal shape and arrangement of departments within a facility (Hicks 2006). The FLP is a combinational problem for which the optimal solution can be found for small problems. GA based search is one of the good method for dealing with problems of facility layout. In the GA approach to optimization, feasible solution to the problem is encoded in data

structures in the form of a string of decision choices that resemble chromosomes. GA maintains population of chromosomes or individuals that are created. The layout design is characterized by chromosomes’ fitness which is measured by its value of objective function. Offsprings are created through reproduction, crossover, and mutation (Balamurugan *et al.*, 2006). FLP is formulated as a quadratic set which covers linear integer programming problem, mixed-integer programming problem and graph-theoretical problem. Therefore, quadratic assignment problem (QAP) formulation has been popular in this kind of problems. But manufacturing practice normally requires particular layout configurations such as single row, multi-row or loop formations. These practical constraints place a huge restriction on the optimization process (Hicks 2006), but the GA based search is one good method of dealing with problems of facility layout.

The pick-up and delivery points position of each cell in our study are located on either one of the cell axes (Rajasekharan *et al.* 1998). In this model, the fractal cells are considered to be rectangular blocks with known dimension of (w, h) where w is width and h is height of each cell. After the crossover and mutation, the facility layout for FrMS for this model has a height, h of 3 rows and width, w of 4 columns. If the fractal cells are written as three rows and four columns in matrix form in MATLAB, then the Pick-up Point is (1, 1) and Delivery Point is (3, 4)(Figure 10).

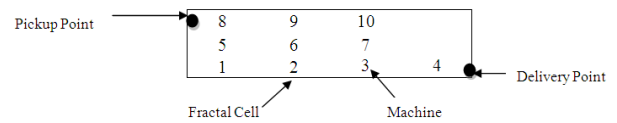


Figure 10: Facility layout problem for FrMS

Some logical assumptions are made for the facility layout problem. These include that the dimensions of the floor area on which the fractal cells are placed is given. The floor space for the flow path on the floor area is not considered. It is also assumed that the flow paths consist of segments that are horizontal and vertical to the walls of the floor (Hu *et al.*, 2007).

The fractal layout dimension, (3 x 4) is chosen because we are considering 10 machines in this study. Thus, we require at least 10 locations for the rectangular fractal cell layout. So, it is feasible to generate a facility layout with 10 machines and 2 spaces. This layout could reduce the material travelling distance by having multi-purpose machines in each fractal cell.

#### 5. IMPLEMENTING THE PROPOSED GENETIC ALGORITHM APPROACH

An iterative algorithm is implemented to optimize the layout and flow assignment according to the design parameters. The layout of each cell is refined using the implied flows between stations. The replicates are re-applied until the heuristic procedures could not find a better solution. The cells are continually iterated to obtain the optimal flow assignment and hence achieve



the optimum fractal layout (Montreuil *et al.*, 1999). The GA procedures - selection, crossover, row inverting mutation, column inverting mutation, and deleting mutation are embedded in the iterative procedure in order to generate the optimal material travelling distances. Hence the desired workstation layout that minimizes the material travelling distances and capacity requirements for product demand and mix is created. Each optimal fractal cell is selected based on the flow distance score. Thus, optimum fractal manufacturing layout is created by combining the three optimal fractal cells. The illustrations of the GA steps are presented by showing the first iteration of the fractal cell 1. Initial cellular layout is assumed to contain 6 cells. Fractal cell1 is generated by combining cell 1 and cell 4 by crossover operation. Cell 1 is shown as parent1 and cell 4 is illustrated as parent2 in MATLAB program codes. Chromosomes for each Parent are represented by the various kinds of genes. The genes are represented by the number 1 to 10 that signify that Machine1 to Machine10 are used. Parent1 is represented as (1 5 2 6; 7 4 3 8; 9 10 3 5; 2 10 8 6; 1 5 9 10), illustrated in 5 rows and 4 columns. Parent2, contains 2 rows and 4 columns as (3 9 2 8; NaN NaN NaN 5). The chromosome for each parent is represented in rows. This means that the chromosomes for Parent1 are (1 5 2 6), (7 4 3 8), (9 10 3 5) and so on. One of the chromosomes from Parent1 is chosen randomly. For instance, the first row chromosome for Parent1 has been selected for the crossover function. On the other hand, the 1<sup>st</sup> row chromosome for Parent2 also is selected to be combined with the chromosome of Parent1 as shown in (Figure 11). The continuous selection of the chromosomes for Parent1 and Parent2 generated 10 different Offspring after the crossover operation (Figure 11). Two Off-springs are generated from each iteration of the crossover. The Offspring1 that is created from selection and crossover with 5 chromosomes are selected for the upcoming mutation. Offspring2 is not used because there are only 3 chromosome lesser than Offspring1.

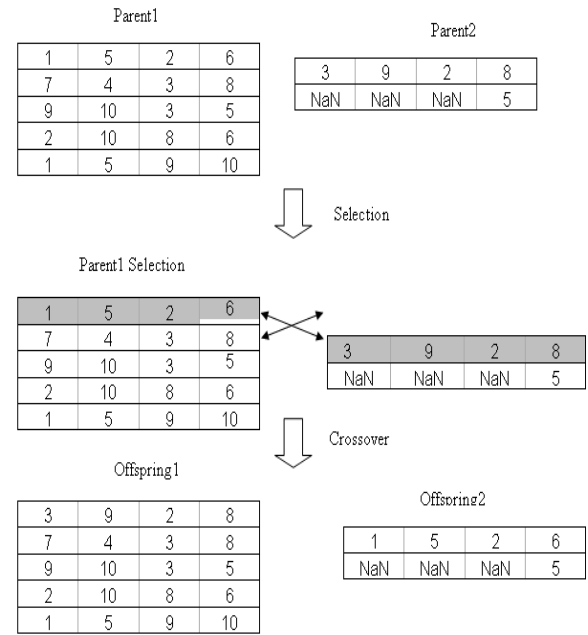


Figure 11: Selection and Crossover

Inverting mutation takes place after the crossover. The Offspring that is generated in the previous crossover is used as the Parent again in this inverting mutation operation. Initially, a cutting point is randomly introduced anywhere along the last row of the Parent. The cutting point indicates the row of the chromosomes for the inverting mutation. The last row of the chromosome is being mutated to the initial row based on the programming code “cirshift” - (mathscript function). The iterations of the row inverting mutation are replicated four times as shown in (Figure 12). For each offspring that is generated, three column inverting mutations take place. For column inverting mutation, chromosome is represented column by column. The cutting point is set in the last column of the chromosome. The column based chromosome is mutated and shifted from the last column to the first column. After this, the Parent is replicated by shifting its chromosomes in columns as shown in (Figure 13). For each Parent that is obtained from the previous mutation step, the entire inverting mutation is expected to replicate 12 times.

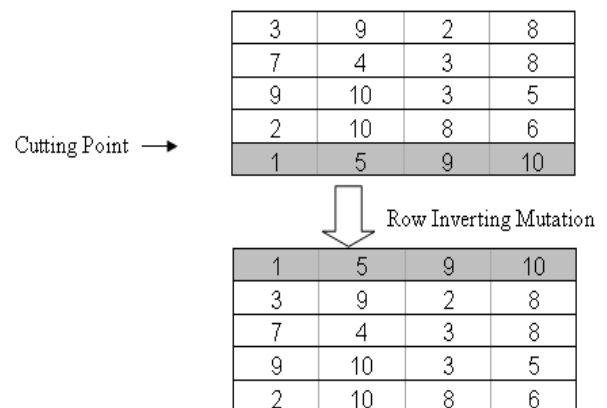


Figure 12: Row Inverting Mutation

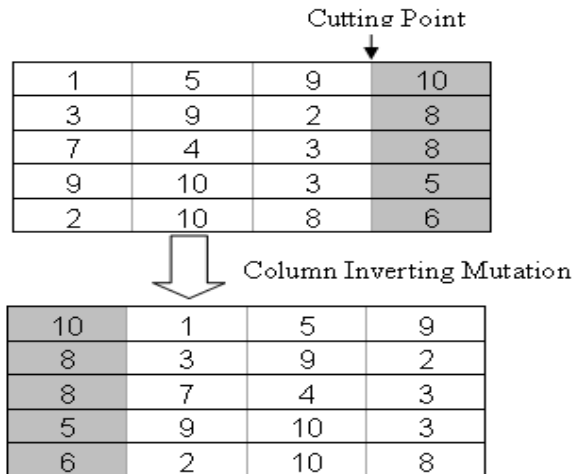


Figure 13: Column Inverting Mutation

After inverting mutation, the Child is generated and transformed to be the Parent again for deleting mutation as shown in (Figure 14). On completion of the previous mutation, the process of deleting mutation is simplified by just deleting the last two rows of the five chromosomes in the Child.

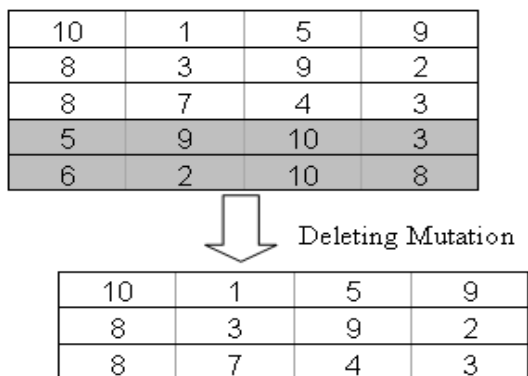


Figure 14: Deleting Mutation

Replacement is the last step in the process of generating fractal cell layout as shown in (Figure 15). In fact, each fractal cell requires 10 machines where no duplicated machines or missing machines are allowed. This is because duplicated machines will increase the material travelling distance. Minimum flow distance score is the requirement for fractal cells.

As a result, machine3, machine8 and machine9 are grouped as duplicated machines that required to be replaced by missing machines. The MATLAB codes are programmed to search the missing machines. The missing machine in this scenario is machine6. Thus, machine6 replaces one of the duplicated machines.

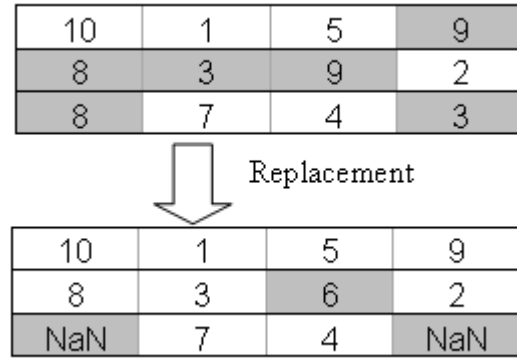


Figure 15: Replacement

The fractal cell layout that is generated after Replacement can be represented as (10 1 5 9; 8 3 6 2; NaN 7 4 NaN). From the Facility Layout Problem (FLP) that was discussed in the previous section, materials are moved into the cell through Pick-up Points and moved out from the cell through Delivery Points as shown in (Figure 16). The Pick-up Point is at (1, 1) while the delivery Point is at (3, 4).

The fractal cells are capable of processing all 15 types of product. Therefore, the materials to be produced need to be processed in specified machine routing sequence. For instance, materials that are used to produce Product1 need to be processed by machine1, machine4, machine7, machine3, machine10, and machine8 in continuous sequence. Each location of machines is represented on (x, y) coordinates. Before the materials are processed in machine1, they have to be carried into the fractal cell through the Pick-up Point. After processing in all the machines within the fractal cells, the final product1 gets delivered to the shipping department through Delivery Point as shown in (Figure 16).

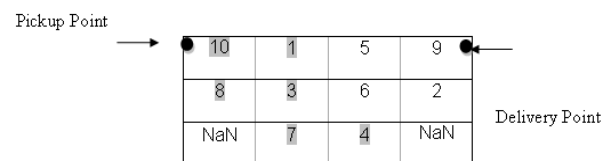


Figure 16: Material Routing sequence for Product1

Then the flow distance score is calculated based on the mathematical solution in MATLAB which is represented as:

$$\text{Distance} = \text{abs}(\text{buffer1}(1) - \text{buffer2}(1)) + \text{abs}(\text{buffer1}(2) - \text{buffer2}(2)) \quad (1)$$

The abs is representation of absolute. The absolute value allows the distance to the left (negative value) and distance to the right (positive value) to be counted into the total distance. Buffer1 and buffer2 is the matrices of data that are being stored in temporary memory.

## 6. OUTPUT RESULTS AND DISCUSSIONS

The computational result of product travelling distances within the fractal cells indicates the flow scores of fractal layout. Flow score is computed and represented as the product travelling distances.

The optimal fractal layout with the minimum flow distance scores is selected by MATLAB and displayed. These output data are used to draw the graphs of flow scores with different generations and flow scores with different product ranges. The GA search for an optimal solution yielded results from 100 iterations and the output is converted into the final fractal cell layout representing the fractal manufacturing layout. The material travelling distances for each of the three fractal cells work out as follows in terms of flow distance scores;

Flow distance score for Cell 1 = 205

Flow distance score for Cell 2 = 217

Flow distance score for Cell 3 = 197

Overall flow distance score for the final fractal manufacturing layout through the proposed GA = 619 and this is shown on (Figure 17).

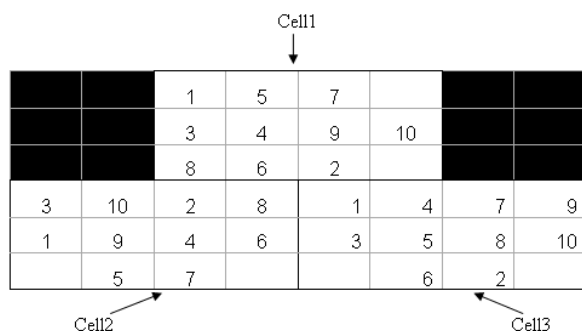


Figure 17: Final Fractal Manufacturing Layout A

Comparatively, the fractal layout according to (Venkatadri *et al.*1997) has machine requirements similar to our final layout requirements with the following flow distances;

Flow distance score for Cell 1 = 251

Flow distance score for Cell 2 = 252

Flow distance score for Cell 3 = 257

Overall flow distance score for Final Fractal Layout according to (Venkatadri *et al.*1997) is = 760 and that is shown on (Figure 18).

This shows that the flow distance score obtained from the proposed GA approach is lesser at 619 than that of (Venkatadri *et al.*1997).

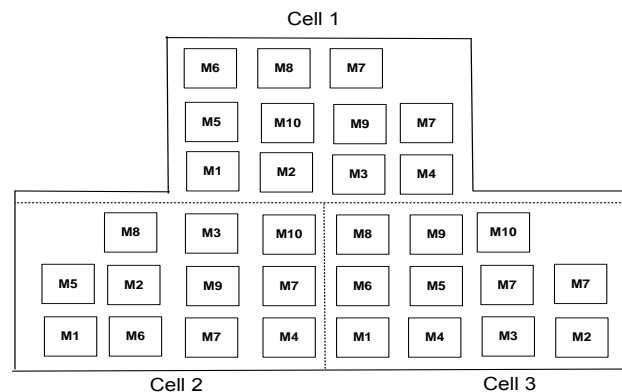


Figure 18: Fractal cell layout according to (Venkatadri et al.1997)

Ascertaining or working out the optimal number of iterations in each cell for our proposed GA approach aided in producing the right flow distances and involved plotting flow distance score against iterations as shown on figures (19), (20) and (21) for cells 1, 2 & 3. These plots signify the optimal flow distances at 205, 217, and 197 for cells 1, 2, & 3 respectively.

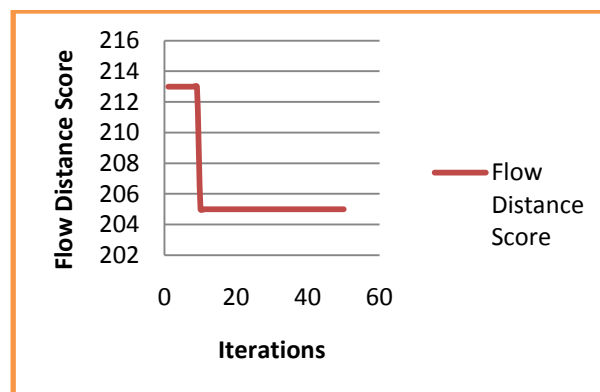


Figure 19: Flow distance score for fractal cell 1

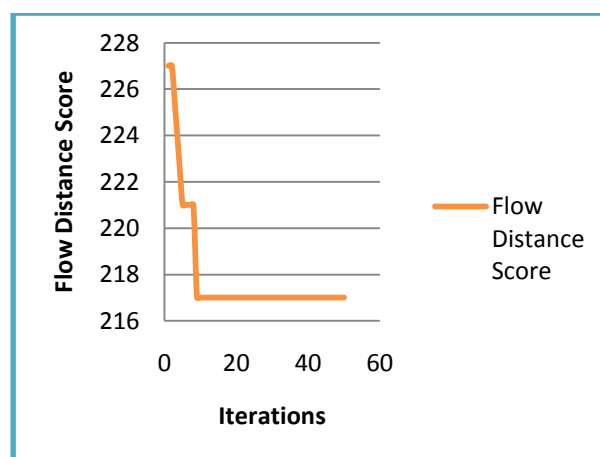


Figure 20: Flow distance score for fractal cell2

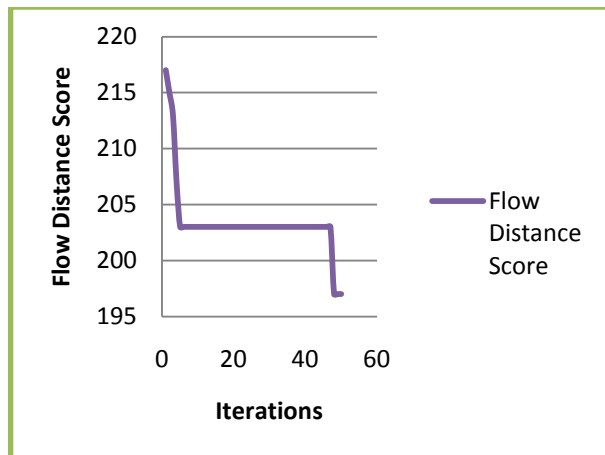


Figure 21: Flow distance score for fractal cell3

## 7. CONCLUSION

The GA approach has been applied in the design of the fractal manufacturing shop floor layout. This algorithm was used to search for the optimal fractal cell layout for efficient and effective material/ product movements within the shop floor. Fundamentally, the decision of how to distribute/assign products to cells as evenly as possible to aid responsiveness to uncertainties in manufacturing and easy control of resources was seen to be very important to the design, implementation and final outcome of the experimentation. The model implemented using MATLAB managed the scenario quite well and handled the mathematical formulations, swapping and deleting matrices etc. quite efficiently. Overall, the computational results indicated that unrestricted product flows offer the best flow scores in a fractal layout.

## REFERENCES

- Al-Sultan, K. S., and Fedjki, C. A., 1997, A Genetic Algorithm for the Part Family Formation Problem, *Production Planning & Control*, Vol. 8, No.8, pp. 788-796.
- Askin, R. G., Ciarallo, F. W. and Lundgren, N. H., 1999, An empirical evaluation of holonic and fractal layouts. *International Journal of Production Research*, 37, 5, 961-978.
- Askin, R. G., Lundgren, N. H. and Ciarallo, F., 1996, A material flow based evaluation of layout alternatives for agile manufacturing. In R. J. Graves, L. F. McGinnis, D. J. Medeiros, R. E. Ward and M. R. Wilhelm (eds.), *Progress in material handling research* (Braun-Brumfield), pp. 71-90.
- Azadivar, F., and Wang, J., 2000, Facility Layout Optimization using Simulation and Genetic Algorithms, *International Journal of Production Research*, Vol. 38, No. 17, pp. 4369-4383.
- Balamurugan, K., Selladurai, V., and Ilamathi, B., 2006, Design and Optimization of Manufacturing Facilities Layouts, *Institution of Mechanical Engineers*, Vol. 220, Part B, pp. 1249-1257.
- Chase, R. B. and Aquilano, N. J., 1992, *Production & operations management - A life cycle approach*, 6th Ed. (Irwin: Boston, Massachusetts, USA).
- Co, H. C. and Araar, A., 1988, Configuring cellular manufacturing systems. *International Journal of Production Research*, 26, 9, 1511-1522.
- Gau, K. Y. and Meller, R. D., 1999, An iterative facility layout algorithm. *International Journal of Production Research*, 37, 16, 3739-3758.
- Hicks, C., 2006, A Genetic Algorithm Tool for Optimization, Cellular or Functional Layouts in the Capital Goods Industry, *International Journal of Production Economics*, Vol. 104, pp. 598-614.
- Holland, J. H., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Michigan.
- Hu, G. H., Chen, Y. P., Zhou, Z. D., and Fang, H. C., 2007, A Genetic Algorithm for the Inter-Cell Layout and Material Handling System Design, *The International Journal of Advanced Manufacturing Technology*, Springer, Vol. 34, pp.1153-1163.
- Kamrani, A. K., and Gonzalez, R., 2003, A Genetic Algorithm-Based Solution Methodology for Modular Design, *Journal of Intelligent Manufacturing*, Vol. 14, pp. 599-616.
- Kiusalaas, J., 2005, *Numerical methods in Engineering with Python*, Cambridge University press.
- Montreuil, B., Venkatadri, U. and Rardin, R. L., 1999, Fractal layout organization for job shop environments. *International Journal of Production Research*, 37, 3, 501- 521.
- Parames Chutima, 2001, Genetic Algorithm for Facility Layout Design with Unequal Departmental Areas and Different Geometric Shape Constraints, *Thammasat International Journal Science and Technology*, Vol. 6, No. 2, pp. 33-43
- Rajasekharan, M., Peters, B. A. and Yang, T., 1998, A Genetic Algorithm for Facility Layout Design in Flexible Manufacturing Systems, *International Journal of Production Research*, Taylor & Francis Ltd., Vol. 36, No. 1, pp. 95-110.
- Saad, S. M. and Lassila, A.M., 2004, Layout design in fractal organizations, *International Journal of Production Research*, Vol. 42, Vol. 17, p. 3529-3550.
- Venkatadri, U., Rardin, R. L. and Montreuil, B., 1997, A design methodology for fractal layout organization. *IIE Transactions*, 29, 911-924.
- Xiaodan W., Chao-Hsien C., Yunfeng W., and Weili Y., 2007, A Genetic Algorithm for Cellular Manufacturing Design and Layout, *European Journal of Operational Research*, Vol. 181, No. 181, pp. 156-167.
- Wild, R., 1993, *Production and operations management*. 4th Ed. (Cassell: London)

# Mechanical Damage to Navy Beans as Affected by Moisture Content, Impact Velocity and seed Orientation

Feizollah Shahbazi<sup>a\*</sup>

<sup>(a)</sup> Lorestan University Khoram abad, I.R.IRAN

<sup>(a)</sup> [Email :feizollah.shahbazi@gmail.com](mailto:feizollah.shahbazi@gmail.com)

## ABSTRACT

The objective of this experiment was evaluate of impact damage to navy beans where seed moisture content, velocity of impact and seed orientation were independent variables. The study was conducted under laboratory conditions, using an impact damage assessment device. The results indicated that impact velocity, moisture content and seed orientation were all significant at the one percent level on the physical damage in seeds. Increasing the impact velocity from 5 to 15 m/s caused an increase in the mean values of damage from 0.17 to 32.88%. The mean values of seed damage decreased by 1.96 times, with increase in their moisture content from 10.03 to 14.98%. However, by a higher increase in the moisture from 14.98 to 24.89 %, the mean value of damage showed a non-significant increasing trend. It was found that the relationship between beans mechanical damage with moisture content and velocity of impact was non-linear and the percentage damage to seeds was a quadratic function of moisture content and impact velocity, respectively. Impacts to the end of the seed produced the higher damage (20.615) than side of the seed (11.14%)

Keywords: Navy bean, impact damage, moisture content, impact velocity, seed orientation.

## 1. INTRODUCTION

Bean seed quality is greatly affected by harvesting, cleaning, drying, handling and storage activities during the seed production process. In these operations, seeds are often subjected to impact forces repeatedly against metal surfaces predisposing them to mechanical damage. Navy beans and other large seeded legumes are especially vulnerable to rough treatment. These seeds are particularly delicate because of their seed anatomy. The mechanical resistance to the impact damage of seeds, such as navy beans, among other mechanical and physical properties, plays a very important role in the design of harvesting and other processing machines (Baryeh, 2002). The value of this basic information is necessary, because during operations, in these sets of equipment, seeds are subjected to impact loads which may cause mechanical damage. Damaged seed commands in lower value, storability problem, and

reduced seed germination and seedling vigor and subsequent yield of crops (Shahbazi and Khazaei, 2002). Impact damage of seeds depends on a number factors such as velocity of impact, seed structural features, seed variety, seed moisture content, stage of ripeness, fertilization level and incorrect settings of the particular working subassemblies of the machines. Among above factors, the seed moisture content and impact velocity are important factors influencing the damage (Keller et al., 1972; Hoki and Picket, 1973; Paulsen et al., 1981; Bartsch et al., 1979; Evans et al., 1990). Information relating the amount of navy bean seed impact damage to velocity of impact and seed orientation is limited. In light of above facts, the objectives of this study were to evaluate the impact damage to navy bean seed and determine the effects of impact velocity, seed moisture content and seed orientation on the percentage of physical damage to beansbiography.

## 2. MATERIAL AND METHOD

The tests were conducted under laboratory conditions. Each sample was impacted using an impact device shown in figure 1. Four steel impact tips (hammer), having a striking face 5 cm wide by 20 cm high, were mounted on a disk (40 cm diameter), rotating in the vertical plane. The impact point on the steel tips moved through a path having a radius of 30 cm. A horizontal slider and rail were mounted just under the disk and impact tips. The slider has 15 seed- supporting pedestals made of flexible plastic tubing. Seeds were held on the pedestals by gravity and the slider was moved toward the impact tips and seeds were impacted one-by-one. A cloth bag behind the machine caught the impacted seeds. The impact velocity of the tips was adjusted by changing the velocity of the electromotor through an inverter set. In this study, the effects of impact velocity (at 5, 7.5, 10, 12.5, and 15 m/s), seed moisture content (at 10.03, 12.65, 14.98, 17.49, 20.12, and 24.89% wet basis) and seed orientation (side and end) were studied on percentage of physical damage in beans. The factorial experiment was conducted as a randomized design with three replicates. For each impact test, 100 seeds were selected randomly from each sample and impacted by using the impact device. After each test,

damaged beans include the broken, cracked, and bruised and split beans were accurately identified by a handy lens and weighed. The percentage of beans damage was calculated as:

$$\text{Beans damage} = \frac{\text{Weight of damaged beans}}{\text{Weight of total beans (damaged + undamaged)}} \quad (1)$$

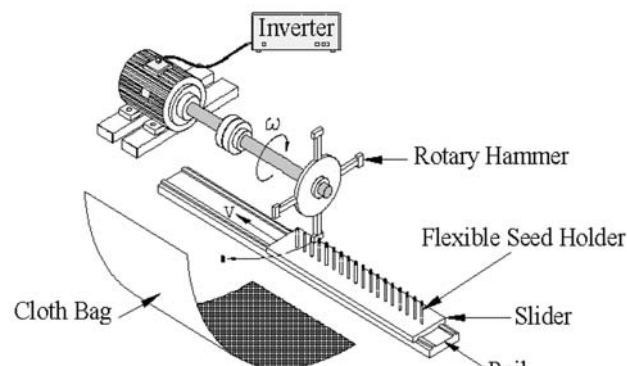


Figure 1: Diagram of the impact damage assessment device.

### 3. RESULTS AND DISCUSSION

The analysis of variance indicated that, Impact velocity, moisture content and seed orientation significantly influenced the percentage physical damage in beans, at 1% probability level. Impact velocity had the most influence; seed orientation and moisture content the least, respectively, within the range studied for these variables. The results of Duncan's multiple range tests for comparing the mean values of the damage to beans at different impact velocities is shown in figure 2a. It is evident that seed damage increased, as a quadratic function, with increasing impact velocity. For all the levels of impact velocity, the differences between the mean values of the damage are significant ( $P=0.05$ ). With increasing the impact velocity from 5 to 15 m/s, the mean value of the damage increased from 0.17 to 32.88%. In figure 2b the percentage damage to seeds is plotted against the velocity of impact. The figure reveals that, at all the seed orientations considered, the seed damage increases as the impact velocity increases. Due to the significant interaction effects between impact velocity and seed orientation, the rates of increase in damage are not the same for all levels of orientations. The effect of impact velocity on the damage is stronger at end orientation than at side. At end orientation, damage increased from 0.34 to 38.67% with increasing the impact velocity from 5 to 15 m/s. Corresponding damages are from 0 to 27.08 at side orientation.

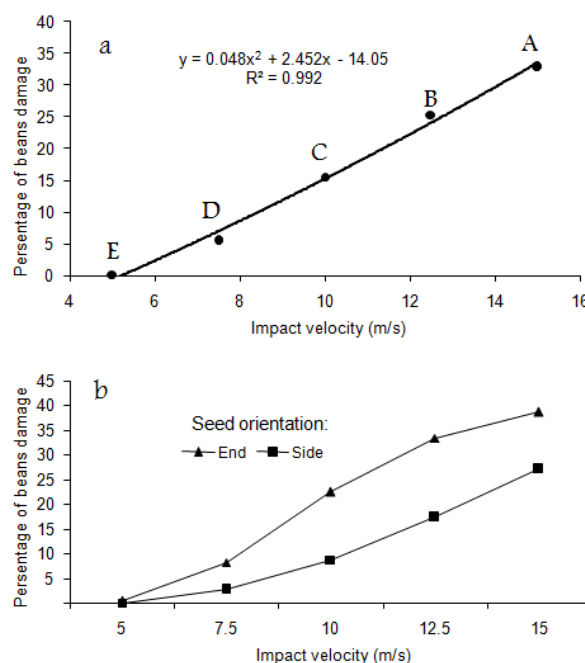


Figure 2. a: Effects of impact velocity on percentage damage to beans. Averages with the same letter have no significant difference at the 5% probability level. b: Beans damage variation with impact velocity at different seed orientations.

The results showed that the percentage of beans damage decreased, as a quadratic function, with increase in their moisture content (figure 3). With increasing the moisture content from 10.03 to 14.98%, the mean values of the percentage damage decreased from 27.09 to 13.79% (by 1.96 times) (figure 4). However, by a higher increase in the moisture from 14.98 to 25.89%, the mean values of damage showed a non significant increasing trend (figure 3). Figure 4 shows the beans damage variation with seed moisture content for various impact velocities. As follows from the relations presented in the figure 4, for all the impact velocities considered, the percentage of the beans damage decreases with increase in their moisture content but the rates of increase in percent damage to beans by decrease in their moisture content are not the same for all the levels of impact velocity. The effect of moisture content on the damage is stronger at higher impact velocities than at lower ones. At the critical range of the tests, when the moisture content decreased from 14.98 to 10.03%, the maximum rate of increase in the damage to beans is obtained for the impact velocity of 15 m/s, which is equal to 26.39% (from 28.87 to 55.26%).



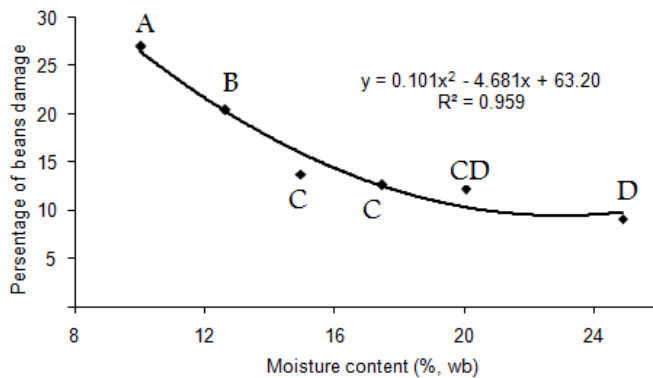


Figure 3: Effects of moisture content on percentage damage to beans. Averages with the same letter have no significant difference at the 5% probability level.

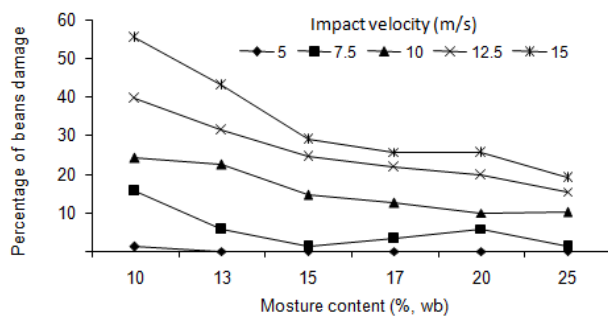


Figure 4: Beans impact damage variation with seed moisture content for different impact velocities

## REFERENCES

- Bartsch, J. A., C. G. Haugh., K. L. Athow, and R. M. Peart. 1979. Impact damage to soybean seed. ASAE Paper No. 79-3037. St. Joseph, MI: ASAE.
- Baryeh, E. A. 2002. A simple grain impact damage assessment device for developing countries. *J. Food Eng.* 56: 37-42.
- Evans, M. D., R. G. Homes, and M. B. McDonald. 1990. Impact damage to soybean seed as affected by surface hardness and seed Orientation. *Trans. ASAE* 33(1):234-240.
- Hoki, M and L. K. Picket. 1973. Factors affecting mechanical damage of navy beans. *Trans. ASAE* 16(6):1554-1557.
- Keller, D. L., H. H. Converse, T. O. Hodges, and D. S. Chung. 1972. Corn kernel damage due to high velocity impact. *Trans. ASAE* 15: 330-332.
- Paulsen, M. R., W. R. Nave, and W. R. Gray. 1981. Soybean seed quality as affected by impact damage. *Trans. ASAE* 24(6):1577-1582.
- Shahbazi, F, and J. Khazaei. 2004. Impact damage to wheat seed. Paper No. 03-149-A. 2004 CIGR International conference. Beijing, China 11- 14 October 2004
- Bantz, C.R., 1995. Social dimensions of software development. In: J. A. Anderson, ed. *Annual review of software management and development*. Newbury Park, CA: Sage, 502-510.
- Holland, M., 2004. *Guide to citing Internet sources*. Poole, Bournemouth University. Available from: <http://www.bournemouth.ac.uk> [accessed 15 July 2005]
- Das, A., 1992. Picking up the bills, *Independent*, 4 June, p. 28a.
- Agutter, A.J., 1995. *The linguistic significance of current British slang*. Thesis (PhD). Edinburgh University.



# HOW IMPORTANT IS PRICE LEADERSHIP IN THE UK FRESH FRUIT AND VEGETABLE MARKET?

Cesar Revoredo-Giha<sup>(a)</sup> and Alan Renwick<sup>(b)</sup>

<sup>(a)</sup>Land Economy and Environment Research Group, Scottish Agricultural College (SAC), Edinburgh UK

<sup>(b)</sup>Land Economy and Environment Research Group, Scottish Agricultural College (SAC), Edinburgh UK

<sup>(a)</sup>[cesar.revoredo@sac.ac.uk](mailto:cesar.revoredo@sac.ac.uk), <sup>(b)</sup>[alan.renwick@sac.ac.uk](mailto:alan.renwick@sac.ac.uk),

## ABSTRACT

The purpose of this paper is to analyse whether price leadership is present in the market for selected fresh fruits and vegetable and whether it plays any role in the pricing decisions of two major retailers in the UK: Tesco and Sainsbury's. This is done using weekly price data for six products sold in the supermarkets. The empirical methodology used in the paper consisted of three consecutive steps: first, a statistical analysis of the properties of the prices; second, a causality analysis of the series using Granger causality tests with the purpose of investigating the existence of price leadership; third, modelling the interrelationships of both supermarket prices by means of vector autoregressive models (VARs) for each product and simulating the retailers' interaction using impulse-response functions. Except in the case of tomatoes, where Tesco appears as the leader since 2001, the other results are mixed, probably reflecting the changes in the marketing of fruit and vegetables market during the period of study.

Keywords: supermarket pricing, perishable products, fruits and vegetables.

## 1. INTRODUCTION

As it is well known in economics, under perfect competition, the price of a product tends to change due to variations in the supply conditions (e.g., changes in the production cost, distribution costs, seasonality) and/or demand conditions (e.g., changes in preferences). However, when the number of suppliers in the market is reduced (and especially if the firms have market power) the pricing of products becomes more complicated as it may respond not only to the aforementioned economic forces, but also to strategic behaviour by the firms, namely price responses that take into account the behaviour of other firms operating in the market.

Supermarkets in the UK have substantively increased their presence in the retail market and now account for approximately 80 per cent of the total market. Their emergence as a force in the UK market for fruit and vegetable products since the 1990s has been well documented as well as major changes in the way they have been marketed (e.g., Fearne and Hughes, 2000; Hingley et al., 2006).

The increasing market share of supermarkets and allegations of non-competitive practices has led to several investigations by the UK Competition Commission (e.g., 2000, 2008, Wilson, 2003, Cooper, 2003, Wynne, 2008) to determine whether they were exercising market power along the supply chain (i.e., with respect to suppliers and consumers).

The evidence collected from the different analyses into whether or not supermarkets exercise market power has so far been inconclusive. If something has emerged from those reports (e.g., UK Competition Commission, 2000) is the fact that economic models do not seem to fit either the behaviour of supermarkets or the way they compete in the market place. However, despite this disappointing result, some stylised facts arose from the Competition Commission investigation including the fact that that supermarkets do consider prices set by their competitors as the principal driver in their own pricing (Competition Commission 2000, pp.135). As pointed out by Lloyd (2008) whether a firm operates a policy whose aim is to offer persistently low prices on a wide range of products or the more traditional policy reliant on discounts on a relatively narrow range of core products, the fact is that all major retailers routinely undertake detailed monitoring of competitor pricing, either via covert price collection or through market research companies who provide and process Electronic Point of Sale (EPOS) data on their behalf. Price surveys are typically conducted on a weekly basis and cover the entire products range.

The purpose of this paper is to analyse whether price leadership is present in the market for selected fresh fruits and vegetable and whether it plays any role in the pricing decisions of two main retailers in the UK: Tesco and Sainsbury's. Fruit and vegetables are chosen for analysis because they are they are a less complex line of products, in the sense that they are perishable, with short shelf duration and their supply chains comprise fewer agents. In addition it is an interesting category which has seen a mix of concentration but also fierce competition and innovation in the retail market over the past 15 years.

As regards the structure of the paper, in the next section, a brief overview of the price leadership

literature is provided, followed by the empirical work. Finally, conclusions are presented.

## 2. PRICE LEADERSHIP

Price leadership models have a long tradition in industrial organisation economics. Many economists such as Forchheimer (1908), Nichol (1930), Stigler (1947), Markham (1951), Lanzillotti (1957) and Bain (1960), Ono (1982), Rotemberg and Saloner (1990), Cooper (1996) have described various types of price leadership models. These models can be classified, as proposed by Scherer and Ross' (1990), into three types: dominant, collusive, and barometric price leadership.

The dominant type describes industries in which a firm (the leader), having the largest market share, establishes its price leadership position with the other minor firms being followers. The collusive type is similar to the dominant type, but with a group of firms exercising the leadership in term of setting prices which are followed by other minor firms. In these two cases (dominant and collusive) the price level is rather more monopolistic than competitive. In contrast in the barometric type the price is set around the competitive level. Within the price leadership models, the barometric type is considered as a benign form of price leadership, where the leader's advantage comes from its higher efficiency in getting relevant information for pricing the products.

Despite the fact that a leader-follower setting is inherently dynamic, with a firm (i.e., the follower) responding to the signals given by another firm (i.e., the leader), the economic literature is relatively scant in terms of the actual dynamics when analysing data. From an empirical perspective Lloyd (2008) introduces the concepts of tactical and strategic price leadership, which are associated with causality observed in the price data. Tactical leadership occurs when the prices of a firm is found to cause (in the Granger sense) the prices of another firm (Granger, 1969). Strategic leadership occurs when the price of a firm affects or determines in the long term the price of the other firm. Note that these two types can match the collusive or dominant leadership concepts found in the literature.

In the case of barometric price leadership, the literature provides more empirical clues about what might be observed in the data. Thus, according to Ono (1982), it can be characterised by: occasional switching between firms in the role of price leader; the occurrence of upward price leadership only in response to increased industry costs or demand; occasional and sometimes substantial time lags in the price response of follower firms and; occasional rejection by the rest of the market of price changes initiated by the price leader.

## 3. EMPIRICAL WORK

The empirical methodology used in this paper consisted of three consecutive steps: the first step was to analyse the statistical properties of the prices, particularly to test the stationarity of the series (i.e., the presence of unit roots). As the series were found stationary in levels, the

next step of the empirical work consisted of a causality analysis of the series using Granger causality tests (considering two separate periods and applying them recursively) with the purpose of investigating the existence of price leadership. The final step consisted of modelling the interrelationships of the prices in both supermarkets for each product by means of vector autoregressive models (VARs) and simulating the retailers' interaction using the impulse-response functions (Hamilton, 1994, Alvarez-De-Toledo et al., 2008).

### 3.1. Data

The data used consisted of two sets of prices: retail and wholesale prices (as indicative of costs). They were collected from the magazine *Grower* (Nexus Media Limited), a weekly British magazine specialising in horticulture, from their section 'supermarket price guide' for two supermarkets: J. Sainsbury and Tesco.

The prices published by the *Grower* were collected by the Market Intelligence Services (MIS), which is a market research company with experience in monitoring retail prices at the major supermarkets, convenience stores and discounters around the UK. The company has an experienced team of collectors who visit the stores and collect the prices independently.

The sample available for this study consisted of 10 years of weekly basis price data covering from July 1996 to March 2007, i.e., approximately 559 observations for tomatoes, Bramley's apples, white cabbages, cucumbers, iceberg lettuce and round lettuce. It should be noted that although MIS collects the data, which comprises a range of prices for each product within the week, the magazine *Grower* only publishes only the modal prices.

Wholesale prices were also collected from the *Grower*, however, the source was the UK Department for Environment, Food and Rural Affairs (Defra) and they correspond to the weekly UK average for several markets for produce class 1.

### 3.2. Data statistical analysis

Table 1 presents descriptive statistics for the price series. Note that the highest variation in the data (measured by the coefficient of variation, i.e., ratio of standard deviation to the mean) corresponded to wholesale tomato prices (approximately 39 per cent). In all cases, the variation in retail prices was below that of the wholesale prices.

Table 1 also presents results concerning the stationarity of the price series (i.e., whether they possess a unit root). This is important for two reasons: first, to avoid obtaining results based on models that reflect spurious correlations, and second, because these results indicate the methodology to follow (i.e., vector autoregressive models, VARs if the series are stationary or vector error correction models, VECs, if the series are non-stationary, Hamilton, 1994). As show by Table 1 all the series are stationary in levels.

Table 1: Descriptive statistics of the variables used within the analysis

	Units	Mean	St. Dev.	Min	Max	Skewness	Kurtosis	Season 1/ Season 2/	Unit roots 2/
J. Sainsbury retail price of tomato	p/kg.	118.8	21.9	77.0	199.0	1.0	4.6		I(0)
Tesco retail price of tomato	p/kg.	117.8	21.2	77.0	199.0	1.1	5.0		I(0)
Wholesale price of tomato - class 1	p/kg.	74.6	29.1	29.2	206.5	1.5	5.7	No in winter	I(0)
J. Sainsbury retail price of Bramley's apple	p/kg.	131.5	20.0	77.0	169.0	-0.5	2.7		I(0)
Tesco retail price of Bramley's apple	p/kg.	124.2	16.7	86.0	152.0	-0.6	2.0		I(0)
Wholesale price of Bramley's apple - class 1	p/kg.	56.0	14.6	19.0	135.9	1.6	6.8	All year	I(0)
J. Sainsbury retail price of white cabbage	p/kg.	55.8	9.1	35.0	89.0	1.8	7.2		I(0)
Tesco retail price of white cabbage	p/kg.	51.3	9.1	33.0	86.0	0.7	4.6		I(0)
Wholesale price of white cabbage - class 1	p/kg.	24.7	6.4	16.0	50.3	1.5	5.6	All year	I(0)
J. Sainsbury retail price of cucumber	p/unit	59.0	14.0	35.0	99.0	1.0	3.8		I(0)
Tesco retail price of cucumber	p/unit	56.7	13.4	35.0	99.0	1.1	4.2		I(0)
Wholesale price of cucumber - class 1	p/unit	34.1	9.8	17.0	81.7	1.5	6.8	No in winter	I(0)
J. Sainsbury retail price of Iceberg lettuce	p/head	60.0	14.6	29.0	129.0	0.9	4.6		I(0)
Tesco retail price of Iceberg lettuce	p/head	58.7	14.0	29.0	119.0	1.0	4.7		I(0)
Wholesale price of Iceberg lettuce - class 1	p/head	30.6	10.1	10.1	85.2	1.9	9.1	No in winter	I(0)
J. Sainsbury retail price of Round lettuce	p/head	35.7	6.7	19.0	49.0	-0.4	2.7		I(0)
Tesco retail price of Round lettuce	p/head	33.1	6.7	19.0	49.0	0.0	2.9		I(0)
Wholesale price of Round lettuce - class 1	p/head	23.4	6.2	11.9	53.8	0.9	5.1	All in year	I(0)

**Notes:**

1/ Indicates whether the fruit or vegetable is produced all the year.

2/ Augmented Dickey Fuller test of the null hypothesis that the series has a unit root. I(0) indicates that the series is stationary in levels. The tests were carried out at 1 per cent significance.

### 3.3. Granger causality analysis

To test the presence of price leadership we used Granger's test of causality (Granger, 1969) as in Lloyd (2008). Table 2 presents Granger causality tests for each product, for the entire sample and splitting the sample into two periods, namely, July 1996 to December 2000 and January 2000 to March 2007. The two periods considered were approximately based on the emergence of supermarkets as a major force in the UK retail market (Hingley et al., 2006). The level of significance chosen for rejecting causality was 1 per cent.

As shown in Table 2, the results indicate significant changes in causality across products between the two periods, therefore, it is expected that the results for the entire sample are masking part of the competition story between the retailers. Thus, in the case of tomatoes the entire sample indicates causality from Tesco to Sainsbury; however, this only reflects the result arising from the second sample period.

With respect to Bramley's apples, the entire sample test indicates causality from Sainsbury to Tesco, although the two sub sample tests indicate that there is no causality. For white cabbage the sub-samples tests indicate double causality, although the entire sample test shows that the causality goes from Sainsbury to Tesco. In the cases of cucumbers, Iceberg and Round lettuce the entire sample test indicates double causality despite the fact that the sub-samples tests indicate causality in only one direction.

Due to the fact that the results from Table 2 were not conclusive as regards of the causality, we decided to run the Granger test recursively (i.e., estimating the test by adding consecutively each observation). The results of the tests (i.e., the value of the F tests) are presented graphically in Figure 1 (solid lines) together with their p-values (dashed lines) in panels a to f.

The recursive Granger tests allow us to complement Lloyd's (2008) analysis and postulate that if one observes unidirectional causality from one retailer to the other, then it can be considered a case of tactical price leadership, whilst if the causality changes over time from one retailer to the other, it can be considered a case of barometric price leadership based on Ono (1982).

The results of the recursive tests help to understand the findings of the two period tests. As shown in the different panels of figure 1, the causality is not constant over time, although the most common case after year 2000 is the presence of double causality and not of a leader-follower relationship.

After 2000 two cases seem to favour the tactical price leadership model: tomatoes and Bramley's apples. For tomatoes (panel a), the figure indicates lack of causality (i.e., no relation at all between both prices) for the period before 2001; however, since 2001 the causality from Tesco to Sainsbury becomes stronger with the additional data points, showing Tesco as a clear "tactical leader" in that market.

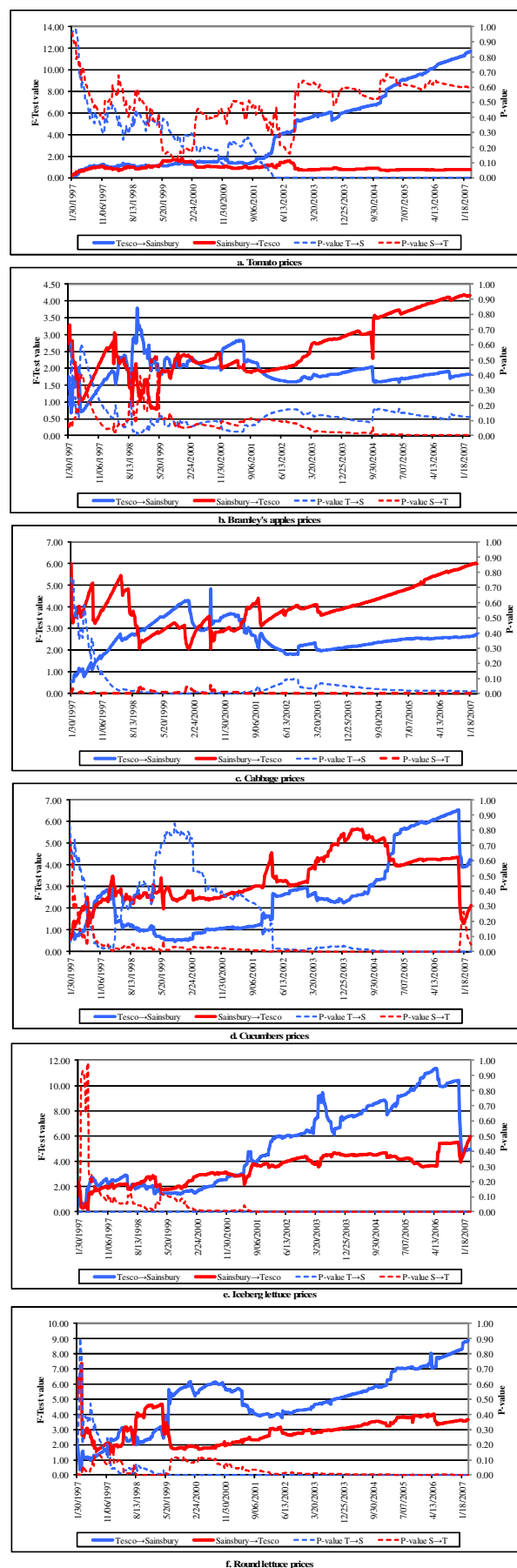


Figure 1: Recursive Granger Causality Tests between J. Sainsbury and Tesco Prices (F tests and P-values).

Table 2: Granger causality tests for different samples

Product	Causality	Entire sample			Jul-96 - Dec-00			Jan-00 - Mar-07		
		Obs.	F test	Signif.1/	Obs.	F test	Signif.1/	Obs.	F test	Signif.1/
Tomato	Tesco → Sainsbury	553	11.7	*	228	1.7		325	18.7	*
	Sainsbury → Tesco		0.8			1.1			1.9	
Bramley's apples	Tesco → Sainsbury	553	1.4		228	2.4		325	0.4	
	Sainsbury → Tesco		3.1	*		1.4			1.4	
White cabbage	Tesco → Sainsbury	553	2.8		228	3.6	*	325	3.6	*
	Sainsbury → Tesco		6.0	*		3.0	*		4.6	*
Cucumbers	Tesco → Sainsbury	536	6.5	*	228	1.1		308	12.1	*
	Sainsbury → Tesco		4.4	*		2.5			2.6	
Iceberg lettuce	Tesco → Sainsbury	536	10.4	*	228	2.8		308	10.8	*
	Sainsbury → Tesco		5.5	*		3.0	*		2.8	
Round lettuce	Tesco → Sainsbury	553	8.9	*	228	5.6	*	325	5.2	*
	Sainsbury → Tesco		3.7	*		2.1			1.6	

**Notes:**

1/ "\*" denotes that the null hypothesis is rejected at 1 per cent of significance. "→" indicates the direction of the tested causality.

The results from Bramley's apple prices (panel b) show the opposite situation and since about 2003 Sainsbury becomes the tactical leader. Note, however, that the situation is not as strong as for tomatoes.

The other products reflect stories of double causality but with different starting periods, namely: cabbage since 1998, cucumbers since 2003, iceberg lettuce since 2000 and round lettuce since 2001. Before these years the situation portrayed in the figures is more complex, with sudden changes in leader/follower roles, which may fit the notion of barometric leadership. Note that it is difficult to track the reasons behind these changes in roles and they might be related to all the transformations that supermarkets went through during the period under study.

The fact that in many cases the relationship between the supermarket prices is one of double causality motivated the next step in the methodology, i.e., the estimation of VARs models to study the interactions between the supermarkets in terms of pricing.

### 3.4. Vector autoregressive models

To study the interaction of the retailers' prices for each one of the products we proposed six VARs models. The generic structure of the VAR model for a product is presented in equation (1), where  $P_{1t}$  and  $P_{2t}$  are the prices at time period  $t$ , for supermarket 1 and 2,  $W_t$  is the wholesale price for the product, which operates as a coincident indicator (Granger causalities test found it causing retailer prices but not the opposite), the  $\varepsilon$  are error terms, which are white noise, and the  $\alpha$  are parameters. The results of the estimation are shown in Table 3.

$$\begin{bmatrix} P_{1t} \\ P_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} + \alpha_{11}W_t \\ \alpha_{20} + \alpha_{21}W_t \end{bmatrix} + \begin{bmatrix} \alpha_{12} & \alpha_{13} \\ \alpha_{22} & \alpha_{23} \end{bmatrix} \begin{bmatrix} P_{1t-1} \\ P_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (1)$$

The exact structure of the models varied for each product, with some of them including trends (all of them where estimated with monthly seasonal dummies). In all  $t$  cases the relevant wholesale prices were included, being statistically significant for all products except for Round lettuce.

The number of lags in each model was selected based on the Akaike and Schwartz criteria (Hamilton, 1994). Where these two criteria failed to indicate the same optimal number of lags, a decision was taken based on the properties of the residuals, which are supposed to be independent and identically distributed. We used the Breusch–Godfrey serial correlation Lagrange multiplier test (Breusch, 1978, Godfrey, 1978) to study the presence of autocorrelation in the series, which was rejected in all the cases.

In addition, we computed the inverse roots of the autoregressive characteristic polynomial to verify whether they were within the unit circle and therefore that all the studied models were dynamically stable.

The results indicated that all the models were dynamically stable.

The next step was to use the estimated models to compute both the impulse-response functions and the variance decomposition for each product (i.e., the effect of a shock in one of the prices in one of the supermarkets on the variance of the price of the same product in the other supermarket). To do this we used the Cholesky decomposition of the error matrix with the series ordered according to their causality (using Granger causality tests) due to the fact that one should expect correlation between the error terms of the VAR equations. The impulse-response functions are not presented in the paper but they are available from the authors.

Table 4 presents six sub-tables, one for each VAR model. Within each sub-table the top panel indicates the decomposition of the variance of the “exogenous” variable and the lower panel presents the decomposition of the variance of the “affected” variable.

To aid interpretation of Table 4, let us concentrate on the case of tomatoes. As the VAR system in the first period responds to a shock in the Tesco price, the variance of this price is only explained by its own shock and not by feedback from the Sainsbury price (see upper panel). However, for the Sainsbury price, 54.9 per cent of its variance is explained by Tesco's price shock and 45.1 per cent by its own price. It is interesting to note that whilst a significant part of the variance in Sainsbury's tomato price is explained by Tesco, the opposite is not true.

Based on the variance decomposition it is possible to classify the results into three cases: first, when a shock in a supermarket price affects the other supermarket price but only a small feedback is received from it (e.g., tomatoes, cucumber, and Iceberg lettuce); second, when the feedback is relatively small for both supermarkets (e.g., Bramley's apple, white cabbage); and third, an intermediate case, when the feedback received from the reacting supermarket is more significant (e.g., Round lettuce).

The first of the mentioned cases could indicate some sort of clear leader follower situation. The second one would be one of “related to some degree but independent”, whilst the third case would represent a higher degree of interaction between the supermarkets.

### 3.5. Conclusions

The purpose of this paper has been to identify the presence of price leadership using data for selected fruits and vegetables from two supermarkets: Tesco and Sainsbury. Except in the case of tomatoes, where Tesco appears as the leader since 2001, the other results are mixed, probably reflecting the changes that have occurred in the fruit and vegetables market during the period of study.

Table 3: Bivariate VAR models for each product 1/2/

	Tomato (1)		Bramley's apple (2)		White cabbage (3)		Cucumber (4)		Iceberg lettuce (5)		Round lettuce (6)	
	R11	R21	R12	R22	R13	R23	R14	R24	R15	R25	R16	R26
R11(-1)	0.316 (0.068) [4.640]	0.147 (0.067) [2.196]	R12(-1)	0.008 (0.043) [15.271]	R13(-1)	-0.072 (0.044) [15.948]	R14(-1)	0.579 (0.047) [12.333]	R15(-1)	0.245 (0.076) [6.323]	R16(-1)	0.842 (0.047) [18.064]
R21(-1)	0.520 (0.071) [7.355]	0.645 (0.070) [9.259]	R12(-2)	0.137 (0.055) [3.467]	R13(-2)	0.154 (0.054) [0.616]	R24(-1)	0.223 (0.047) [4.727]	R15(-2)	0.073 (0.082) [0.799]	R16(-2)	-0.052 (0.044) [-1.168]
Intercept	8.890 (2.420) [3.673]	12.429 (2.382) [5.217]	R12(-3)	-0.057 (0.056) [-1.027]	R13(-3)	0.006 (0.044) [1.870]	Intercept	2.845 (1.279) [2.225]	R15(-3)	-0.297 (0.076) [-0.509]	R26(-1)	0.141 (0.034) [4.198]
W1	0.141 (0.018) [7.761]	0.155 (0.018) [8.668]	R12(-4)	-0.033 (0.053) [-0.618]	R23(-1)	0.702 (0.049) [2.622]	W4	0.207 (0.031) [6.759]	R25(-1)	0.499 (0.078) [3.768]	R26(-2)	0.020 (0.035) [0.575]
			R12(-5)	0.105 (0.044) [2.398]	R23(-2)	0.022 (0.059) [-2.039]	Trend	0.006 (0.002) [3.114]	R25(-2)	-0.050 (0.083) [-0.222]	Intercept	1.722 (0.465) [3.700]
			R22(-1)	0.080 (0.045) [1.767]	R23(-3)	0.109 (0.048) [0.189]			R25(-3)	0.211 (0.079) [-0.365]	W6	0.002 (0.001) [0.249]
			R22(-2)	-0.033 (0.054) [-0.598]	Intercept	3.634 (1.543) [4.326]			Intercept	8.420 (1.942) [2.831]	Trend	0.002 (0.001) [2.647]
			R22(-3)	-0.005 (0.054) [-0.099]	W3	0.012 (0.037) [3.925]			W5	0.292 (0.044) [6.369]		
			R22(-4)	0.020 (0.054) [0.359]	Trend	-0.001 (0.001) [-2.305]						
			R22(-5)	-0.019 (0.045) [-0.412]								
Intercept	3.860 (1.854) [2.083]	3.410 (1.778) [1.918]										
W2	0.055 (0.022) [2.570]	0.034 (0.021) [1.640]										
Trend	0.007 (0.002) [3.387]	0.002 (0.002) [1.290]										
Observations	412	412	540	540	554	554	412	412	303	303	542	542
Adj. R-squared	0.84	0.84	0.93	0.91	0.74	0.79	0.83	0.80	0.73	0.71	0.93	0.87
F-statistic	725.66	695.75	643.30	472.84	201.75	262.84	501.15	401.62	119.07	107.26	1177.63	597.33
Log likelihood	-1500.15	-1493.71	-1644.92	-1622.40	-1610.98	-1547.09	-1264.36	-1297.35	-1016.42	-1005.81	-1080.68	-1243.24
Mean dependent	117.78	117.00	131.48	124.06	55.61	51.18	57.37	55.45	55.56	54.51	35.76	33.14
S.D. dependent	23.25	22.49	20.15	16.76	8.83	8.71	12.69	12.56	13.57	12.62	6.70	6.66

**Notes:**

1/ Standard errors are presented in parenthesis under the coefficients and t statistics are presented in brackets. All the regressions were carried by ordinary least squares.

2/ The name of the variables is as follows: Rij (e.g., R11) denotes the retail price of supermarket i (1=Sainsbury, 2= Tesco) of product j (where j is given with the heading of the regression). Wj indicates the wholesale price of product j. All the regressions included seasonal dummies (monthly), which are not presented to save space.



Table 4: Variance decomposition for each product VAR model 1/

Tomato (1)			Bramley's apples (2)			White cabbage (3)		
Variance Decomposition of R21			Variance Decomposition of R12			Variance Decomposition of R13		
Period	St dev.	R21	Period	St dev.	R12	Period	St dev.	R23
1	9.1	100.0	1	5.2	100.0	1	4.5	100.0
4	13.5	98.8	4	7.8	99.3	4	6.4	99.2
16	14.7	98.6	16	10.3	96.8	16	7.4	99.2
Variance Decomposition of R11			Variance Decomposition of R22			Variance Decomposition of R23		
Period	St dev.	R21	Period	St dev.	R12	Period	St dev.	R23
1	9.3	54.9	1	4.9	3.6	1	4.0	5.2
4	14.0	77.1	4	6.8	7.8	4	5.7	7.2
16	15.3	80.6	16	9.6	11.8	16	7.3	21.2
Cucumber (4)			Iceberg lettuce (5)			Round lettuce (6)		
Variance Decomposition of R14			Variance Decomposition of R15			Variance Decomposition of R16		
Period	St dev.	R14	Period	St dev.	R15	Period	St dev.	R26
1	5.2	100.0	1	7.0	100.0	1	1.8	100.0
4	7.6	93.4	4	9.9	95.2	4	3.2	91.4
16	8.1	91.1	16	10.2	93.3	16	4.8	68.6
Variance Decomposition of R24			Variance Decomposition of R25			Variance Decomposition of R26		
Period	St dev.	R14	Period	St dev.	R15	Period	St dev.	R26
1	5.7	29.3	1	6.8	46.6	1	2.4	9.7
4	7.8	42.9	4	9.4	59.0	4	3.8	18.6
16	8.2	46.3	16	9.7	57.5	16	5.2	24.7

**Notes:**

1/ The name Rij (e.g., R11) denotes the retail price of supermarket i (1=Sainsbury, 2= Tesco) of commodity j (where j is given with the heading of each variance decomposition).

The results of the recursive causality tests show that if one considers the entire sample available for each product, then one should conclude that there is no evidence of a price leadership situation in any. However, since it is possible to find leaders when considering sub-periods, then the evidence would support in most of the cases the presence of barometric price-leadership.

If instead of considering the entire sample, one considers sub-periods then, it is possible to find cases of price leadership (tactic) such as in the case of tomatoes since 2001 or in a lesser extend Bramley apples since also since 2003. All the other cases, show double causality indicating some sort interaction (e.g., by observing the other's prices) when pricing their product.

To provide a further analysis of the interrelationships between supermarkets VAR models were estimated to produce the impulse-response and the variance decomposition analyses. These analyses allowed us study the effect that movements (i.e., generated through a shock) on the price of one supermarket have on the price of the same product of the other supermarket. The results of the variance decomposition seem to indicate three cases depending on the degree of feedback received between the supermarkets' prices: the first case corresponds when a change in the price of one supermarket affects the other supermarket's price but received only small feedback from it. This was found in tomato, cucumber, and Iceberg lettuce. The second case, takes place when the feedback is relatively small for both supermarkets and this was found in the Bramley's apple and white cabbage cases. The third case, an intermediate situation, shows that the feedback received from the reacting supermarket is significant as was the case for Round lettuce.

## REFERENCES

- Alvarez-De-Toledo, P., Crespo Marquez, A., and Núñez, F., 2008, Introducing VAR and SVAR predictions in system dynamics models', *International Journal of Simulation and Process Modelling*, 4(1), 7-17.
- Bain, J. S., 1960, Price leaders, barometers, and kinks. *Journal of Business*, 33, 193-203
- Breusch, T. S., 1978. Testing for Autocorrelation in Dynamic Linear Models, *Australian Economic Papers*, 17, 334-55.
- Cooper, D., 2003. Findings from the Competition Commission Inquiry into Supermarkets. *Journal of Agricultural Economics*, 54(1), 127-143.
- Cooper, D.J., 1996, Barometric price leadership, *International Journal of Industrial Organization* 15, 301-325
- Fearne, A. and Hughes, D., 2000, Success factors in the fresh produce supply chain: insights from the UK. *British Food Journal*, 102(10), 760-772.
- Godfrey, L. G., 1978, Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables, *Econometrica*, 46, 1303-1310.
- Granger C.W.J. (1969) 'Investigating Causal Relationships by Econometrics Models and Cross-Spectral Methods' *Econometrica*, 27, 424-438.
- Forchheimer, K., 1908, Theoretisches zum unvollständigen Monopol. In: *Das Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft* (G. Schmoller, ed.). Leipzig.
- Lanzillotti, R. F. (1957). Competitive price leadership-a critique of price leadership models. *Review of Economics and Statistics*, 39, 55-64.
- Nexus Media Ltd (several issues). Grower, Kent.
- Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.
- Hingley, M., Lindgreen, A. and Casswell, B., 2006, Supplier-Retailer Relationships in the UK Fresh Produce Supply Chain. *Journal of International Food & Agribusiness Marketing*, 18(1/2), 49-86.
- Lloyd, T., 2008, Price Leadership in UK Food Retailing: Time Series Representation and Evidence. Paper prepared for the 82nd Annual Conference of the Agricultural Economics Society, held at the Royal Agricultural College, Cirencester, UK on 31st March to 2nd April 2008. Available from: <http://econpapers.repec.org/RePEc:ags:aes008:36862> [accessed 15 June 2010].
- Markham, J. W., 1951, The Nature and Significance of Price Leadership. *The American Economic Review*, 41(5), 891-905.
- Nichol, A. J., 1930, *Partial Monopoly and Price Leadership*. Philadelphia: Smith-Edwards.
- Ono, Y., 1982, Price Leadership: A Theoretical Analysis *Economica*, New Series, 49(193), 11-20.
- Rotemberg, J.J. and Saloner, G., 1990, Collusive Price Leadership, *The Journal of Industrial Economics*, 39(1), 93-111.
- Sheldon, I. and Sperling, R., 2003, Estimating the Extent of Imperfect Competition in Food Industry: What have we learned?, *Journal of Agricultural Economics*, 54(1), 89-109.
- Scherer, F.M. and D. Ross, 1990, *Industrial market structure and economic performance*, 3rd ed. Houghton Mifflin, Boston.
- Stigler, G. J., 1947, The kinky demand curve and rigid prices. *Journal of Political Economy*, 55, 432-449.
- UK Competition Commission, 2000. Supermarkets: A report on the supply of groceries from multiple stores in the United Kingdom. Available online: [http://www.competition-commission.org.uk/rep\\_pub/reports/2000/446super.htm#full](http://www.competition-commission.org.uk/rep_pub/reports/2000/446super.htm#full). [accessed 26 June 2010]
- UK Competition Commission, 2008,. The supply of groceries in the UK market investigation. Available online: [http://www.competition-commission.org.uk/rep\\_pub/reports/2008/fulltext/538.pdf](http://www.competition-commission.org.uk/rep_pub/reports/2008/fulltext/538.pdf) [accessed 26 August 2010]
- Wilson, P., 2003. Papers from a mini-symposium on Competition in the Food industry: An

Introduction, *Journal of Agricultural Economics*, 54(1), 85-87.

Wynne, Catherine, 2008, Supermarkets under scrutiny, BBC News, Friday, 15 February 2008.

#### **AUTHORS BIOGRAPHY**

**Cesar Revoredo-Giha** is an economist with a PhD from the University of California at Davis. He has been Team Leader of Food Marketing at the Land Economy and Environment Research Group in SAC since June 2008. Before joining SAC he worked at The World Bank, the University of Georgia, and the University of Cambridge. He specialises in the industrial organization of food markets, international trade and econometrics.

**Alan Renwick** is an economist with a PhD from Newcastle University. He has been Head of the Land Economy and Environment Research Group at SAC since September 2004. Prior to this, he spent 12 years in the Centre for Rural Economic Research in the Department of Land Economy, University of Cambridge. His main area of work is in policy evaluation, particularly involving commodity regimes and agri-environmental policy.

# SIMULATIONS OF LIQUID FILM FLOWS WITH FREE SURFACE ON ROTATING SILICON WAFERS (RoWaFlowSim)

Markus Junk<sup>(a)</sup>, Frank Holsteyns<sup>(a)</sup>, Felix Staudegger<sup>(a)</sup>, Christiane Lechner<sup>(b)</sup>, Hendrik Kuhlmann<sup>(b)</sup>,  
Doris Prieling<sup>(c)</sup>, Helfried Steiner<sup>(c)</sup>, Bernhard Gschaider<sup>(d)</sup>, Petr Vita<sup>(e)</sup>

<sup>(a)</sup>Lam Research AG, Villach, Austria

<sup>(b)</sup>Institute of Fluid Mechanics and Heat Transfer, Vienna University of Technology, Austria

<sup>(c)</sup>Institute of Fluid Mechanics and Heat Transfer, Graz University of Technology, Austria

<sup>(d)</sup>ICE Strömungsforschung GmbH, Leoben, Austria

<sup>(e)</sup>Department Mineral Resources and Petroleum Engineering, University of Leoben, Austria

<sup>(a)</sup>[markus.junk@lamresearch.com](mailto:markus.junk@lamresearch.com), <sup>(b)</sup>[christiane.lechner@tuwien.ac.at](mailto:christiane.lechner@tuwien.ac.at), <sup>(c)</sup>[prieling@fluidmech.tu-graz.ac.at](mailto:prieling@fluidmech.tu-graz.ac.at),  
<sup>(d)</sup>[bernhard.gschaider@ice-sf.at](mailto:bernhard.gschaider@ice-sf.at), <sup>(e)</sup>[petr.vita@unileoben.ac.at](mailto:petr.vita@unileoben.ac.at)

## ABSTRACT

For application in the semi-conductor industries, the flow on a rotating disk (wafer) is studied numerically. A systematic validation of available variations of the Volume-of-Fluid (VoF) scheme for this two-phase flow problem is done. To make larger parameter variations of this problem with a moving fluid supply possible, a code based on the thin-film approximation is developed. To better understand possible physical mechanisms for the removal of nano-particulate contaminations from a wafer, another code to study the detachment of submicron particles exposed to shear flows is developed and validated. An immersed boundary method with direct forcing is implemented. The particle-wall interaction is treated with a soft contact model.

Keywords: two-phase flow, thin-film approximation, particle removal, immersed boundary method

## 1. INTRODUCTION

Lam Research AG develops and manufactures single-wafer wet-processing solutions for the semiconductor industry. Their core technology is based on spin processing tools. In a spin process chamber, specific cleaning, etching and stripping chemicals are applied onto the semiconductor substrates (wafers). During the process, the wafer is placed on a chuck with the side to be processed facing up. A nitrogen cushion protects the bottom side from any contact from any contamination. While the chuck and with it the wafer rotate, chemical mixtures can be applied to the wafer by means of a dispenser to form a liquid layer on the surface of the wafer. Dependent on the process step, the dispenser is either in a fixed position (centre or off-centre) or is moving across the wafer. The continuously supplied chemicals are spun off the wafer and collected in drain levels, which surround the chuck. Multiple drain levels are arranged on top of each other making several processing steps with different chemicals in one chamber possible. For this, the vertical position of the chuck can be changed. During the entire process, an air flow through the process chamber from the top prevents

the chemical vapours from escaping the chamber into the surrounding environment. The air is exhausted through a duct in the side wall of the process chamber.

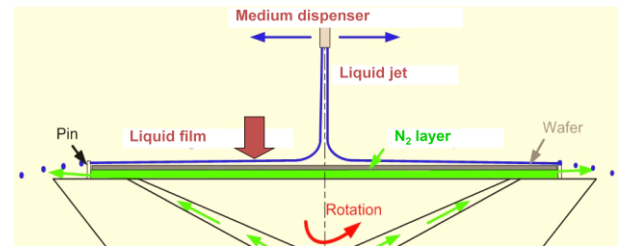


Figure 1: Sketch of a chuck, showing the moveable dispenser and the liquid jet, the N<sub>2</sub> cushion protecting the lower side of the wafer, and the ejection of droplets at its edge.

Semi-conductor industry has to face a number of challenges and continuously improve existing and develop new methods for processing due to decreasing dimensions (32nm and beyond) of the semi-conductor devices. One of these is the removal of nano-particulate contamination from wafer substrates without damaging high aspect ratio structures. Understanding the removal mechanisms in more detail is consequently an essential step in developing a reliable and efficient physical cleaning technique suitable for the 32 nm technology that can replace today's chemical cleaning technique. To meet these challenges, numerical simulations and especially computational fluid dynamics (CFD) simulations are getting more and more important. Lam Research is increasing its efforts to improve the simulation capabilities by close cooperation with universities. The cooperative project RoWaFlowSim (funded by the Austrian Research Promotion Agency FFG in the framework of the ModSim program) is part of these efforts and addresses flow related issues which are of high importance in the development process.

One subproject is related to thin film flow evolution and disintegration. A better understanding of liquid film flow, formation of droplets and ejection from the wafer is essential for the development of the

products, because they have a big impact on the quality and on the economic efficiency of the wafer processing. The formation of droplets and their movement in the airflow can lead to contamination on the wafer and to cross-contamination of the process media when the droplets enter the wrong drain level or are entrained in the airflow that is exhausted from the process chamber.

A second subproject is related to the transport of near-wall particles within the viscous sublayer of a laminar or turbulent boundary layer. The objective is to improve the understanding of particle adhesion and detachment by means of modelling and simulation. This will give an estimate of the minimum forces required to remove particles from a substrate, and lead to a better understanding of the larger scale flows, which are required for selective removal of contaminants, without damage to nanostructures. Understanding the basic effects will help to select physical mechanisms and to generate flows leading to jetting, micro-streaming, shock waves etc. (Fig. 2).

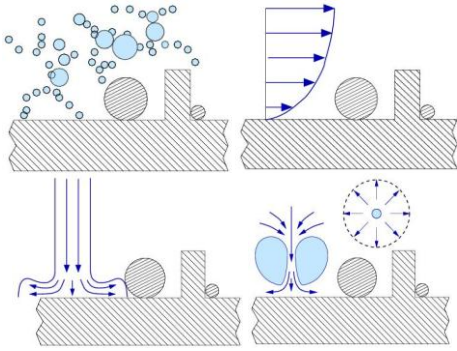


Figure 2: Physical mechanisms to generate controlled perturbations for particle removal: acoustic cavitation, acoustic streaming, droplet impact, laser bubbles.

## 2. FLOW ON A ROTATING WAFER

### 2.1. Numerical scheme

The work in this part of the project started with a systematic study of the suitability of several variations of the Volume-of-Fluid method (VoF), which is the method of choice for this kind of flows. The implementations available in the commercial code ANSYS Fluent and in the open-source code OpenFOAM were considered for this study. For a first comparison the numerical results of axisymmetric calculations with experimental results from the literature, a test case with simplified inflow conditions was chosen (Thomas, Fagri, and Hankey 1991).

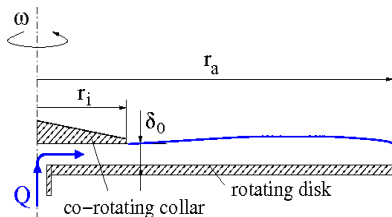


Figure 3: Inflow conditions for the first test case. The vertical jet is replaced by a horizontal inflow.

After comparing simulation results with the different available VoF schemes (Gschaider, Vita, Prieling, and Steiner (2010)), the HRIC scheme (in Fluent) and the Inter- $\gamma$  scheme (in OpenFOAM) were selected as the most appropriate interface tracking schemes for further use in the project. Other schemes showed either too high smoothing of the interface or a physically questionable strong waviness of the instantaneous solutions and caused problems in the outer area of the disk where the film height becomes very thin. Therefore, these schemes are not used in the further work.

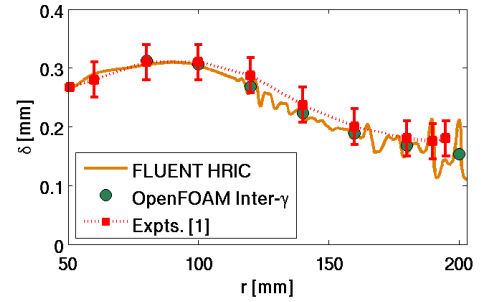


Figure 4: Comparison of the computed values for the instantaneous film thickness with experimental data for the case with horizontal inflow (water, flow rate 700l/min, rotation rate 200rpm).

### 2.2. Validation

For the more realistic case with a vertical inflow from a dispenser, the time-averaged numerical results always showed very good agreement with analytical solutions assuming a thin film approximation (Sisoev, Matar, and Lawrence 2003; Kim, and Kim 2009) for the axisymmetric case and in most cases also with the experimental data, if available. The simulations could also confirm that the profiles of the velocity components (based on 4<sup>th</sup> order polynomials) in the analytical solutions match very well those obtained in the 3D simulations sufficiently far away from the impingement region (Fig. 5). Discrepancies in some cases between the predicted time-averaged film heights and the corresponding experimental data may be partly attributed to uncertainties in the measured data.

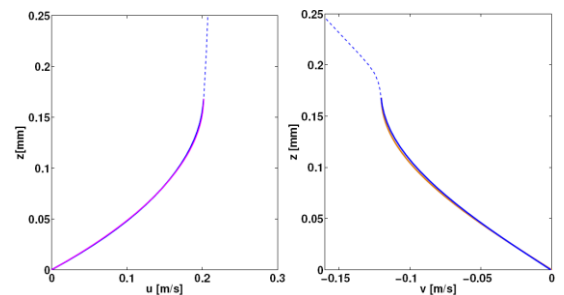


Figure 5: Profiles of the radial and azimuthal velocity at a radial position  $r=40\text{mm}$  (water, flow rate 0.36l/min, rotation rate 200rpm).

Even in the case with axisymmetric boundary conditions, i.e. with central impingement, strong

disturbances in the circumferential direction especially in the outer part of the wafer are observed. This asymmetry in the flow, and the fact that in the real process tools, the liquid is often applied to the wafer from an off-centre positioned or even moving dispenser, make three-dimensional simulations necessary. The high computational costs require a carefully adapted design of the numerical mesh to resolve the motion of the very thin liquid structures. Simulations (with mesh sizes between 1 and 5 million cells) showed fairly good agreement (by visual comparison with high-speed images obtained in the laboratory) for the case with the lower rotational speed, where the 3D wavy structures in the outer region of the disk are relatively large (Fig. 6). For higher rotation rates these structures tend to break up into irregular small-scale wavelets, whose spatial resolution would require unaffordable mesh sizes. Nevertheless, the simulations capture the average mean values very well in all considered cases (Fig 7).

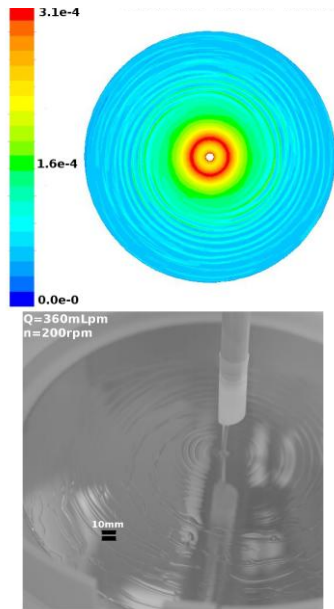


Figure 6: Surface waves in an axisymmetric setup with centred position of the dispenser in simulation (contours of film thickness) and experiment (water, flow rate 0.36l/min, rotation rate 200rpm)

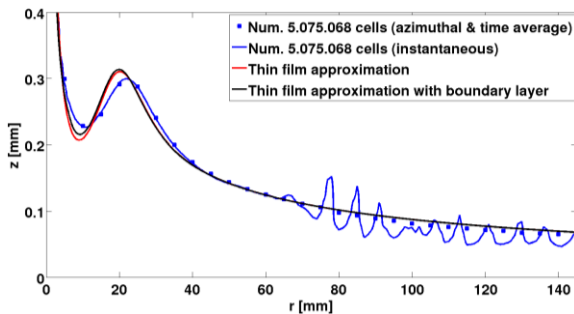


Figure 7: Comparison of computed instantaneous and averaged values of the film thickness (3D, Fluent) with results of an analytical solution based on a thin film approximation (water, flow rate 0.36l/min, rotation rate 200rpm).

The resulting computing times are still prohibitively high for systematic studies in the industrial use, even if a coarse base mesh is used and a dynamic grid adaption is applied where necessary. Thus, the applicability of 3D simulation is constrained to a small number of test cases and to a very limited range of operation conditions with relatively low rotational speeds. The costly 3D simulations are still highly valuable for the evaluation of some basic assumptions made in approximate solution approaches reducing the spatial dimensionality, as will be shown in the next section. Moreover, particular complex phenomena like a discontinuous wetted surface can only be captured by 3D simulations. This is seen from Fig. 8, where the 3D results reflect the formation of dry spots at the centre very well.

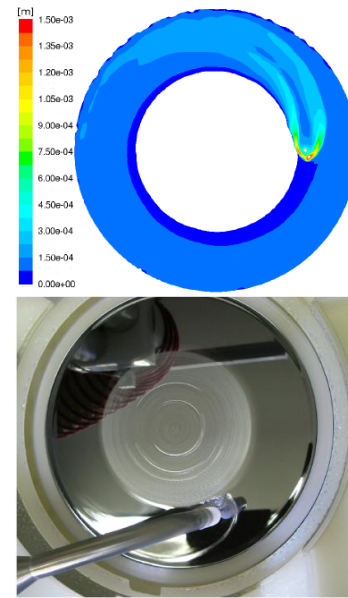


Figure 8: Occurrence of dry spots in a setup with static off-centre position of the dispenser (water, flow rate 0.3 l/min, rotation rate 60 rpm) in simulation (contours of film thickness) and experiment (note the reversed orientation of the rotation).

### 2.3. Thin film approximation

To make larger parameter studies of the flow on the wafer more feasible, it was decided to develop a numerical tool on a computationally less costly alternative, termed “2.5D” approach. This concept basically represents an integral method, reducing the simulation of the liquid flow to a 2D problem by solving the equations of motion in the thin film approximation (shallow water equations), neglecting the vertical momentum equation. This is justified by the small thickness of the liquid film compared to the dimensions of the wafer, and it seems to be a good candidate for most of the relevant operating conditions. The above mentioned analytical solutions are based on the same approximation; however, an analytical solution is impossible for off-centre or moving dispenser.



The dependent variables in this approach are the film height  $h$  and the vertically averaged film velocity  $\bar{\mathbf{u}}$ . For the vertical velocity profile, a 4<sup>th</sup> order polynomial ansatz is chosen. The coefficients of the polynomial

$$f(\xi) = \mathbf{a}_0 + \mathbf{a}_1\xi + \mathbf{a}_2\xi^2 + \mathbf{a}_3\xi^3 \quad (1)$$

with  $\xi = x_3/h$  being the normalized vertical coordinate and the coefficients  $\mathbf{a}_i(x_1, x_2)$  vectorial functions of the coordinates in the plane, are determined by the following set of boundary conditions at the wafer ( $\xi = 0$ ) and at the free surface ( $\xi = 1$ ):

$$\begin{aligned} f(\xi)|_{\xi=0} &= \mathbf{u}_{\text{Wafer}}, \quad \frac{\partial^2 f(\xi)}{\partial \xi^2} \Big|_{\xi=1} = 0, \quad \frac{\partial f(\xi)}{\partial \xi} \Big|_{\xi=1} = 0 \\ \int_0^1 f(\xi) d\xi &= \bar{\mathbf{u}} \end{aligned} \quad (2)$$

Applying the thin film approximation, the resulting form of the momentum equation based on this velocity profile is given by

$$\begin{aligned} \frac{\partial}{\partial t}(h\bar{\mathbf{u}}) + \nabla(\bar{\mathbf{u}}\bar{\mathbf{u}}) &= -\frac{1}{\rho}h\nabla(\rho|\mathbf{g}|h - \sigma\nabla^2 h) \\ &\quad - \nu \frac{1}{h} \frac{\partial f(\xi)}{\partial \xi} \Big|_{\xi=0} \end{aligned} \quad (3)$$

( $\sigma$ : surface tension,  $\rho$ : density,  $\nu$ : kinematic viscosity of the liquid,  $\mathbf{g}$ : gravity vector), which must be solved together with the continuity equation

$$\frac{\partial}{\partial t}h + \nabla(h\bar{\mathbf{u}}) = 0 \quad (4)$$

The programming of a flexible solver is based on an implementation of the Finite Area Method (FAM), a specialisation of the Finite Volume Method for film flows, by Tukovic and Jasak (2008) in a branch of OpenFOAM (OpenFOAM 1.6-ext, 2010).

The thin film approximation cannot be used in the vicinity of the impinging jet, where the flow is redirected and vertical momentum is converted to radial and circumferential momentum. Therefore, a combined approach must be used: the flow conditions close to the impingement region, which are the result of a fully 3D simulation, are imposed as a boundary condition. Outside this region, the flow is calculated based on the thin film approach. The full simulations have shown that the conversion of vertical momentum occurs in a rather confined area (typically less than 5mm radius, compared to 150mm wafer radius).

First tests of this concept assuming axisymmetric flow showed very good agreement with the results of the full 3D simulation with the VoF method. In the outer region, where the flow becomes non-axisymmetric and irregular, as indicated by the waviness in the instantaneous numerical results, good agreement with the averaged numerical results is still

observed. The described profile assumptions are also used for the more complex cases with off-centre impingement (Fig. 9). For a moving dispenser, a static mesh turned out to be more economical and to yield better results than re-meshing based on the momentary dispenser position. With this code, it becomes now possible to calculate this practically most important cases on a time frame which is reasonable for industrial use (e.g. 3D VoF simulation: computing time up to 3 weeks on 4 cores for 1s real time; thin-film method: about 2 hours for 10s real time on 1 core of the same computer).

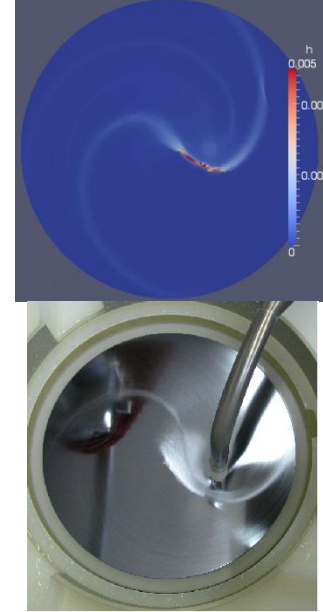


Figure 9: Simulation with the implemented thin-film approximation (contours of film thickness) for a test case with fixed off-centre position of the dispenser (offset 55mm, flow rate 2.8l/min, rotation rate 400 rpm).

## 2.4. Outlook

Present work in this part of the project concentrates on providing data for modelling the disintegration of the liquid film at the edge of the wafer and on the validation and the improvement of the implementation of the 2.5D concept. Extensions of this code, e.g. to calculate etch rates on the wafer, thus leading to a better modelling of the underlying industrial application, are under consideration.

## 3. PARTICLE MOTION

The objective of this sub-project is to numerically study the onset of particle motion under the influence of hydrodynamic forces and follow the particle after motion has set in. The idea is that physical mechanisms suitable to remove nano-particulate contamination from the substrate as e.g. acoustic cavitation generate local perturbations of the flow close to the surface of the wafer. Due to its small size, a submicron particle adhering to the substrate then essentially faces a simple



shear flow  $\mathbf{u}_0(z, t)$  which exerts hydrodynamic forces  $\mathbf{F}_D(t)$ ,  $\mathbf{F}_L(t)$  and a torque  $\mathbf{M}_D(t)$  on the particle as sketched in Fig. 10. By numerically exposing the particle to various profiles  $\mathbf{u}_0(z, t)$  with different shear rates  $\dot{\gamma}$ , criteria for effective particle removal shall be established.

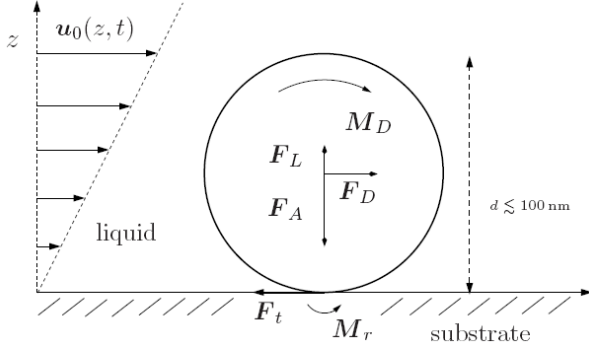


Figure 10: Model for particle removal: the nano-sized particle adhering to the substrate with adhesion force  $\mathbf{F}_A$  faces a shear flow  $\mathbf{u}_0(z, t)$ , which exerts drag  $\mathbf{F}_D(t)$ , lift  $\mathbf{F}_L(t)$  and a torque  $\mathbf{M}_D(t)$  on the particle.  $\mathbf{F}_t$  and  $\mathbf{M}_r$  denote tangential friction and rolling resistance.

### 3.1. Numerical method

In order to numerically simulate the onset of motion and follow the particle after motion has set in, a direct numerical simulation is performed where the finite sized rigid particle fully couples to the fluid. To this end we implemented an immersed boundary method, following Uhlmann (2005) and Taira and Colonius (2007), into OpenFOAM. The incompressible Navier-Stokes equations are solved on a Cartesian grid covering a domain  $D$  which includes the region  $P(t)$  occupied by the particle. The presence of the particle in the fluid is modelled with a volume force  $\mathbf{f}(\mathbf{x}, t)$  in the momentum equation,

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \nu \Delta \mathbf{u} + \mathbf{f}$$

$$\nabla \cdot \mathbf{u} = 0 \quad (5)$$

where  $p(x, t)$  denotes the pressure over density in the fluid and  $\nu$  is the kinematic viscosity.  $\mathbf{f}(\mathbf{x}, t)$  has to be determined numerically such that the fluid velocity  $\mathbf{u}(\mathbf{x}, t)$  satisfies the no-slip condition  $\mathbf{u}(\mathbf{x}, t) = \mathbf{U}(t) + \boldsymbol{\Omega}(t) \times (\mathbf{x} - \mathbf{X}(t))$  at the surface of the particle  $\partial P(t)$ , where  $\mathbf{U}(t)$  and  $\boldsymbol{\Omega}(t)$  denote the particle's translational and angular velocities and  $\mathbf{X}(t)$  is the centre of mass.

The momentum and the continuity equation are solved with icoFoam, OpenFOAM's standard solver for incompressible flow, which uses a PISO-like algorithm for pressure velocity coupling. During each pass in the corrector loop, before solving the Poisson equation for pressure, we require a corrected velocity to satisfy the no-slip condition on a set of Lagrangian marker points  $\mathbf{X}_l$ , which are regularly distributed on  $\partial P(t)$ . This condition yields a linear system of equations for a surface force  $\mathbf{F}(\mathbf{X}_l)$ . As the marker points  $\mathbf{X}_l$  are not

related to the Cartesian grid  $\mathbf{x}_i$ , interpolation and smearing operations are defined with the same regularized  $\delta$ -function (see Peskin (2002) and Beyer and LeVeque (1992)).

Given a numerical solution to the above problem at a time  $t_n$ , the particle's equations of motion

$$\frac{d\mathbf{X}}{dt} = \mathbf{U}, \quad M \frac{d\mathbf{U}}{dt} = \mathbf{F}_{\text{hydro}} + \mathbf{F}_{\text{body}} + \mathbf{F}_{\text{wall}}$$

$$I \frac{d\boldsymbol{\Omega}}{dt} = \mathbf{M}_{\text{hydro}} + \mathbf{M}_{\text{wall}} \quad (6)$$

are solved, where  $M$  and  $I$  denote the mass and the moment of inertia of the (spherical) particle.  $\mathbf{F}_{\text{body}}$  is a body force acting on the particle as e.g. gravity, which for a nano-sized particle is negligible compared to the adhesion force.  $\mathbf{F}_{\text{hydro}}$  and  $\mathbf{M}_{\text{hydro}}$  are computed directly from the surface force  $\mathbf{F}(\mathbf{X}_l, t_n)$  and the fluid velocity  $\mathbf{u}(\mathbf{x}_i, t_n)$ .  $\mathbf{F}_{\text{wall}}$  and  $\mathbf{M}_{\text{wall}}$  represent the force and torque the wall exerts on the particle and will be specified below.

The numerical implementation has been validated in 2D for standard test cases (Lechner and Kuhlmann (2010)). Validation in 3D is in progress. As an example, Fig. 11 shows the flow around a sphere in contact with the wall for a particle Reynolds number  $\text{Re}_p = \dot{\gamma}d/(2\nu) = 2$ , with  $\dot{\gamma}$  the shear rate and  $d$  the diameter of the particle. The above algorithm easily can be extended to handle particles that deviate from spherical shape.

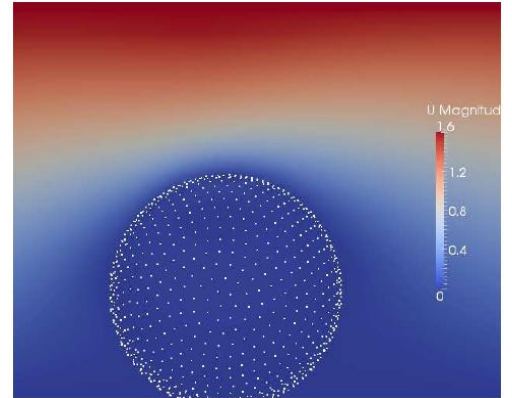


Figure 11: Direct numerical simulation of a (fixed) sphere in contact with a wall exposed to a linear shear flow. In this example the particle Reynolds number  $\text{Re}_p = 2$ . The no-slip boundary condition was enforced on the white points distributed on the surface of the sphere. The magnitude of the fluid velocity is shown on the plane  $y = \text{const}$  through the centre of the sphere with a scale  $V = \dot{\gamma}d$ .

### 3.2. Interaction with a wall

The hydrodynamic forces will cause motion of the particle in form of lift-off, sliding or rolling if they overcome the contact forces comprising adhesion, tangential friction, rolling and torsion resistance. The forces and torques describing the particle-wall interaction are given by

$$\begin{aligned}\mathbf{F}_{\text{wall}} &= \mathbf{F}_n + \mathbf{F}_t + \frac{RM}{I + R^2M} \mathbf{n} \times \mathbf{M}_r \\ \mathbf{M}_{\text{wall}} &= R\mathbf{n} \times \mathbf{F}_t + \frac{I}{I + R^2M} \mathbf{M}_r + \mathbf{M}_t\end{aligned}\quad (7)$$

where  $R$  denotes the radius of the particle and  $\mathbf{n}$  is the unit normal to the wall pointing from the fluid region outside. The normal part  $\mathbf{F}_n$  includes a repulsive force opposing elastic deformation and the attractive adhesion force. Depending on the elastic properties and the size of the particle a family of models ranging from the Johnson-Kendall-Roberts model to the Derjaguin-Muller-Toporov model (see e.g. Johnson and Greenwood (1997)) is appropriate to describe  $\mathbf{F}_n$ .

Tangential friction  $\mathbf{F}_t$  resists the relative motion of the contact area and the wall until static friction is overcome and sliding sets in. For (sub)-micron particles sliding friction is described by  $\mathbf{F}_t = -\tau_s A_c \hat{\mathbf{U}}_c$ , with  $\tau_s$  the shear stress of the contact,  $A_c$  the contact area and  $\hat{\mathbf{U}}_c$  the unit vector in direction of the relative velocity. Similarly the critical rolling resistance could be modelled as  $\mathbf{M}_r = -\tau_r R A_c \hat{\boldsymbol{\Omega}}$  with  $\tau_r$  the rotational friction coefficient and  $\hat{\boldsymbol{\Omega}}$  the unit vector in direction of the angular velocity (see e.g. Sitti (2004)).  $\tau_r$  and  $\tau_s$  have to be determined experimentally for the materials under consideration. The contact area  $A_c$  can be derived from the adhesion models if the work of adhesion is known.  $\mathbf{M}_t$  denotes the torsion resistance. Combining Eqs. (5) and (6) and assuming equilibrium, the limit of sliding or rolling friction would be reached if the hydrodynamic forces satisfy the following inequalities

$$\begin{aligned}\left| \frac{2}{7} \mathbf{F}_D + \frac{5}{7R} \mathbf{n} \times \mathbf{M}_D \right| &\geq \tau_s A_c \quad \Rightarrow \text{sliding} \\ |-\mathbf{M}_D + R\mathbf{n} \times \mathbf{F}_D| &\geq \tau_r A_c \quad \Rightarrow \text{rolling}\end{aligned}\quad (8)$$

As a first step for a stationary linear shear flow at very small particle Reynolds numbers  $\text{Re}_p$ , the expressions derived by O'Neill (1968) can be used to approximate drag and torque,  $|\mathbf{F}_D| \approx 16.031 \text{Re}_p \rho v^2$  and  $|\mathbf{M}_D| \approx 5.93 R \text{Re}_p \rho v^2$ , where  $\rho$  denotes the density of the liquid (note, that there is a typo in the paper concerning the torque, which misses a factor 1/2). This would yield the following conditions for the onset of sliding respectively rolling

$$\begin{aligned}0.345 \text{Re}_p \rho v^2 &\geq \tau_s A_c, \\ 21.96 \text{Re}_p \rho v^2 &\geq \tau_r A_c\end{aligned}\quad (9)$$

### 3.3. Outlook

At present, the 3D implementation is being validated. The next step will be the comparison with experimental data (onset of motion of well defined particles in a shear flow) which are being generated within another cooperative project by Lam Research AG.

## ACKNOWLEDGMENTS

The funding by the Austrian Research Promotion Agency (FFG) within the ModSim program (project number 819320) is gratefully acknowledged.

## REFERENCES

- Beyer, R.P., LeVeque, R.J., 1992. Analysis of a one-dimensional model for the immersed boundary method. *SIAM Journal on Numerical Analysis*, 29, 332-364.
- Gschaidner, B., Vita, P., Prieling, D., Steiner, H., 2010. Liquid coverage of a rotating disk. 5<sup>th</sup> OpenFOAM workshop, Göteborg (Sweden). [Slides](#)
- Johnson, K.L., Greenwood, J.A., 1997. An adhesion map for the contact of elastic spheres. *Journal of Colloid and Interface Science*, 192, 326-333.
- Kim, T.S., Kim, M.U., 2009. The flow and hydrodynamic stability of a liquid film on a rotating disc. *Fluid Dynamics Research*, 41, 1-28.
- Lechner, C., Kuhlmann, H.C., 2010. Development of a code for the direct numerical simulation of particles detaching from solid walls by hydrodynamical forces. In: M. Sommerfeld, ed. *Proceedings of the 12<sup>th</sup> workshop on Two-Phase Flow Predictions, Halle (Germany)*.
- O'Neill, M.E., 1968. A sphere in contact with a plane wall in a slow linear shear flow. *Chemical Engineering Science*, 23, 1293-1298.
- Peskin, C.S., 2002. The immersed boundary method. *Acta numerica*, 11, 479-517.
- Sisoev, G.M., Matar, O.K., Lawrence, C.J., 2003. Axisymmetric wave regimes in viscous liquid film flow over a spinning disk. *Journal of Fluid Mechanics*, 495, 385-411.
- Sitti, M., 2004. Atomic force microscope probe based controlled pushing for nanotribological characterization. *IEEE/ASME Transactions on Mechatronics*, 9, 343-349.
- OpenCFD, 2010. OpenFOAM 1.7.1 Available from: <http://openfoam.com>
- The OpenFOAM-Extend Project, 2010. OpenFOAM 1.6-ext. Available from: <http://www.extend-project.de>
- Taira, K., Colonius, C., 2007. The immersed boundary method: A projection approach. *Journal of Computational Physics*, 225, 2118-2137.
- Thomas, S., Fagri, A., Hankey, W., 1991. Experimental analysis and flow visualization of a thin liquid film a stationary and rotating disk surface. *ASME Journal of Fluids Engineering*, 113, 73-80.
- Tukovic, Z., Jasak, H., 2008. Simulation of free-rising bubble with soluble surfactant using moving mesh finite volume/area method. 6<sup>th</sup> International Conference on CFD in Oil & Gas, Metallurgical and Process Industries, SINTEF/NTNU, 10-12 June, Trondheim (Norway).
- Uhlmann, M., 2005. An immersed boundary method with direct forcing for the simulation of particulate flows. *Journal of Computational Physics*, 209, 448-476.

# PLC CODE PROCESSING FOR AUTOMATIC SIMULATION MODEL GENERATION

Gergely Popovics<sup>(a,b)</sup>, András Pfeiffer<sup>(a)</sup>, Botond Kádár<sup>(a)</sup>, Zoltán Vén<sup>(a,b)</sup>, László Monostori<sup>(a,b)</sup>

<sup>(a)</sup> Fraunhofer Project Center on Production Management and Informatics, Computer and Automation Research Institute, Hungarian Academy of Sciences

Kende u. 13-17, Budapest, H-1111, Hungary

<sup>(b)</sup> Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Budapest, Hungary

<sup>(a,b)</sup> [popovics@sztaki.hu](mailto:popovics@sztaki.hu), <sup>(b)</sup> [pfeiffer@sztaki.hu](mailto:pfeiffer@sztaki.hu), <sup>(a)</sup> [kadar@sztaki.hu](mailto:kadar@sztaki.hu), <sup>(a,b)</sup> [ven.zoltan@sztaki.hu](mailto:ven.zoltan@sztaki.hu),  
<sup>(a,b)</sup> [laszlo.monostori@sztaki.hu](mailto:laszlo.monostori@sztaki.hu)

## ABSTRACT

One of the most widespread techniques to evaluate various aspects of a manufacturing system is discrete-event simulation (DES) (Banks 1998, Law and Kelton 2000, O’Rielly and Lilegdon 1999). However, building a simulation model of a manufacturing system is a difficult task and needs great resource expenditures. Automated data collection and model buildup can drastically reduce the time of the design phase as well as support model reusability. Since most of the manufacturing systems are controlled by low level controllers (e.g., PLCs, CNCs) they store structure and control logic of the system to be modeled by a DES system. The paper introduces an ongoing research of PLC code processing method for automatic simulation model generation of a conveyor system of a leading automotive factory. Results of the validation process and simulation experiments are also described through a case study.

Keywords: PLC code, manufacturing, simulation, generic simulation model

## 1. INTRODUCTION

As majority of manufacturing simulation projects constrained on financing issues or on close deadlines, it is necessary to reduce the most time-consuming and expensive tasks of a simulation project. The design phase of a simulation project constituted by input data preparation and model building can be significantly reduced by using automatic data gathering and processing.

On the other hand, simulation is applied to long-term planning, design and analysis of manufacturing systems. These models are termed “throw away” or “stand-alone” models because they are seldom used after the initial plans or designs have been finalized. As opposed to the “traditional” use of simulation, Son et al. (2001) proposed that once the system design has been finalized, the simulation that was used for evaluation could be used as the basis for system control. In their concept simulation was created by using neutral system components, i.e., they made efforts to build simulation

models for Shop Floor Control System (SFC), generated automatically.

Data needed to build simulation model of a manufacturing system are available in production database or can be collected. Nowadays majority of the enterprises are installing automated manufacturing system (AMS) consisting of programmable logic controllers (PLCs). Subsequently, the topology and the control logic of the manufacturing system needed for a simulation model are inherently kept in these PLCs. Consequently, building of simulation models can be supported by data and control logic extracted from PLC codes.

The paper introduces an ongoing research of PLC code processing method that generates a structured dataset that can be used by manufacturing simulation software to automatically create and parameterize a model.

## 2. AUTOMATED SIMULATION MODEL BUILDING

As stated by Ryan and Heavey (2006) the most commonly used rule of a simulation project is the so called “40-20-40 rule”. The rule states that time spent developing a simulation project can be divided as follows:

- 40% to requirements gathering,
- 20% to model translation,
- 40% to experimentation.

Time-consuming requirements gathering phase contains input data collection and preparation. Significant planning time reduction can be achieved by automating data gathering and preparation.

Several approaches have been used for automating simulation model buildup by automatic input data gathering and processing. Park et al. (2010) suggest a naming rule in PLC codes to automatically identify objects and control logic in code giving a basic data set to build simulation model. This approach needs a renaming process on PLC codes if naming rule suggested is not applied.

Bagchi et al. (2008) describe a discrete event simulator developed for daily prediction of WIP

position in an operational wafer fabrication factory to support tactical decision-making. Model parameters are automatically updated using statistical analysis performed on historical event logs generated by the factory, while “snapshot” (specified later) of current status of production is generated by using the manufacturing execution system (i.e., aggregated info of PLC).

The most widely spread applications of using PLC codes for generating simulation models aims of verifying PLC codes themselves. Han et al. (2010) propose a prototyping to improve limitations of existing control logic verification methods and ladder programming. The technique proposed by them supports functionality verification of PLC code on low control level. Contrarily PLC code process method proposed by the authors is for evaluating the effects of changing PLC codes on the overall system.

Several previous studies aimed of reducing the time needed by the development phase of a simulation project of a manufacturing system that highlights the importance of this topic.

Research is carried out to develop a method to reduce the time required for building simulation models. Wya et al. (2011) proposed a generic simulation modeling framework to reduce the simulation model building time. The proposed framework composed several software that contained information of layout and control logic of the modeled objects. According to this approach layout and control logic of the manufacturing system must be designed by the appropriate software.

### 3. NOVEL SOLUTION FOR REDUCING SIMULATION MODEL BUILDING TIME

#### 3.1. Data needed to build a simulation model

Simulation models related to manufacturing systems require several types of data.

Shop floor layout provides information on the physical structure. Basically the layout identifies elements of the system, their dimensions and internal distances. Elements controlled by PLC also have unique identification, so a list of elements can be retrieved from PLC code. Relations of elements also can be extracted from PLC codes, because there are references in the code between elements that are connected. Even it is not easy to reverse engineer them the topology of controlled system is incorporated in the PLC code.

Second indispensable information of manufacturing systems is the control logic of their elements. Control logic describes the response to be given on PLC’s output depending on the input. The control logic consists of structured methods so variables and object’s relationship can be transformed to the language of the simulation software.

It is also necessary to parameterize elements of the simulation model. Most of PLC codes do not operate with these kinds of parameters of controlled element however, AMSS usually apply Manufacturing

Execution System (MES) that stores status changes of controlled elements and timestamp of state change events. Possible states and parameters of elements can be retrieved by statistical evaluation accomplished on these data as shown on Figure 1.

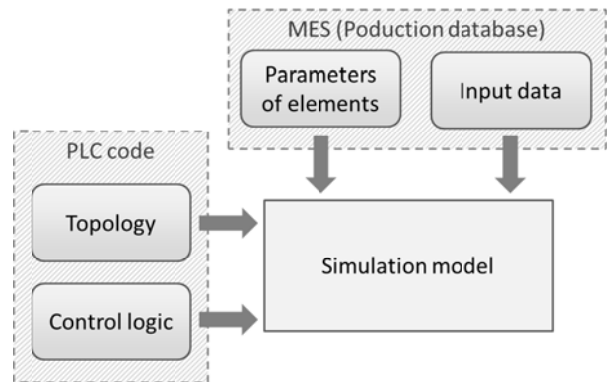


Figure 1: Data gathering for simulation model

#### 3.2. Data needed to run a simulation model

After building and parameterizing a simulation model it is ready to evaluate system performance in several production scenarios. Evaluations of the response given to cut-off scenarios are commonly used technique of testing maximum performance of a system. They are also called stress tests. Performance limits of a system can be determined by stored historical data of factory. Input parameters of cases when system run on maximum performance can be automatically detected by defining parameters that shows performance of the system and parallel recording the set of input parameters as well.

In order to evaluate behavior of systems in situations that have never happened before input parameters must be set manually. This type of experiments is so called “what-if-scenario” experiments.

Ramp up phase distorts results of the simulation run so it is necessary to ignore or avoid it. In order to achieve this goal simulation run must start from a steady state. A so called snapshot of real system is created and initial states and parameters of model are set according to states and parameters of snapshot. To analyze the behavior of a system in real time, snapshots are created periodically. This adaption of simulation model needs real time connection with production database hence it is an on-line application of the model.

#### 3.3. Proposed methods

As mentioned above, several types of data are needed to build up a simulation model. Each type of data is generated by different methods and processes.

##### 3.3.1. Variable and value identification of the PLC code

The base of data gathered from PLC code is in AWL format according to text-based IEC 61131-3 standard. A grammar definition language (ProGrammar, [www.programmar.com](http://www.programmar.com)) is applied to identify variables,

values and methods of PLC codes by defining rules and the nature of them. As PLC code is a strictly defined language this process can be applied on different codes as well.

The test PLC code as all the PLC codes consists of blocks. One block describes the behavior of one actuator (e.g. electric engine actuating the conveyor). Every block consists of two main parts (see Figure 2):

- In the first part the inputs of the PLC are stored in the appropriate memory field.
- The second part of the block is for function call. The arguments of function call are defined in this part, as well as the name of the function.

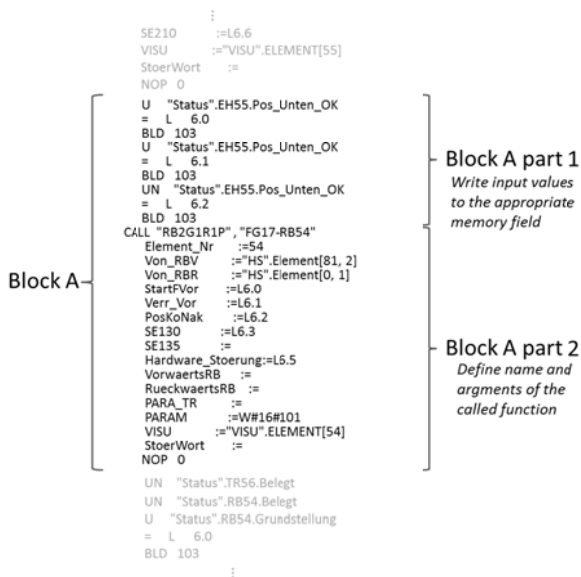


Figure 2: PLC code part

The well-defined syntactics of the code allows the identification of all inputs and parameters. As ProGrammar is a grammar definition language it is possible to create a language to identify the variables and parameters of the code by parsing the code with the defined language. Data stored in the code such as the connection of the objects can be gathered by this process, as all the blocks have references to the id of the connected objects. According to the direction of the material flow two type of reference exist:

- Reference to the next object (successor object)
- Reference to the previous object (predecessor object)

Possessing the information of the connections of the object it is possible to generate the system topology.

### 3.3.2. Topology graph generation

A technical computing software (Mathematica, [www.wolfram.com/mathematica](http://www.wolfram.com/mathematica)) is applied to visualize the results of the above process (see Figure 3). Results support the automatic buildup of simulation models.

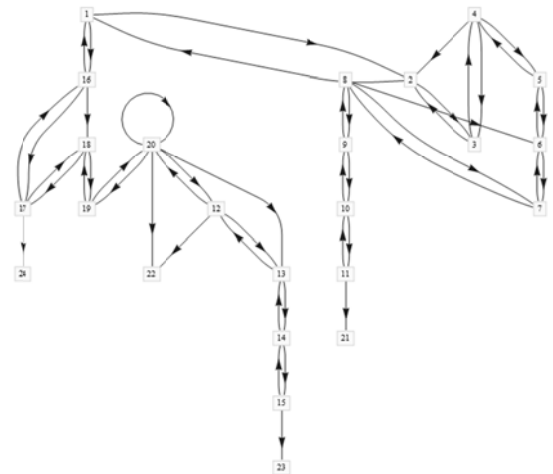


Figure 3: Part of topology of a conveyor system

As Mathematica is suited to call the parsing method of ProGrammar using a predefined command, it can be fed with data of connection among the parts of the conveyor system. After a conversion of data, the topology of the examined system can be generated by a graph creating command. The data received from ProGrammar can also be converted to a connection table form. Special simulation software (Tecnomatix Plant Simulation ver10) and its internal connection tables and methods are applied to generate the simulation object instances and run the simulation. (see Figure 4).

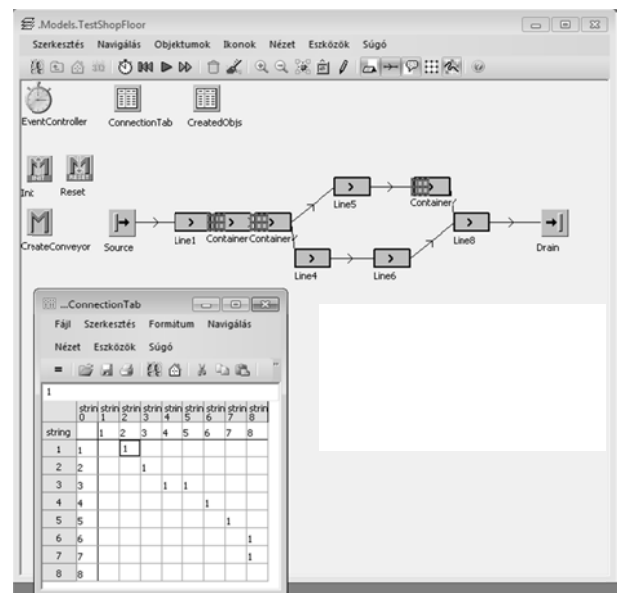


Figure 4: Screenshot of test run executed on simulation model

A production database is used by providing parameters of the elements and input data for the simulation model (see Figure 1). Directly and indirectly usable data are gathered from MES database, and transformed as well as processed to the format that simulation software is able to apply (Kádár et al. 2010, Pfeiffer et al. 2009., Pfeiffer 2007.).

#### 4. PRELIMINARY SIMULATION RESULTS

The proposed solution introduced in the previous section is tested on real data of complex conveyor system in a large-scale manufacturing factory by using a part of the real PLC code of the running system. Main drawbacks of currently applied simulation methods enlisted by decision makers for testing changes of either physical system or control system are as follows:

- specific building blocks, models of processes are redundantly implemented in separated systems, which results in more development time and costs;
- there is no integrated design and analysis environment or system where the overall system and processes can be analyzed;
- there is no factory-standard, uniform user interface;
- results of current simulations cannot be automatically adapted in real system;
- real-time daily use is not possible (manual changes required).

On the base of the above assumptions, the target of the research is to build a new simulation system in which the simulation model is an inherent part of the control system hence, input comes automatically from real execution system and support is provided to evaluation of simulation results.

This means the collection of data from PLC level, enabling reliable, automatically generated DES models, furthermore, manual intervention regarding simulation (data collection, modeling, parameterizing) can be reduced to a definitely low level.

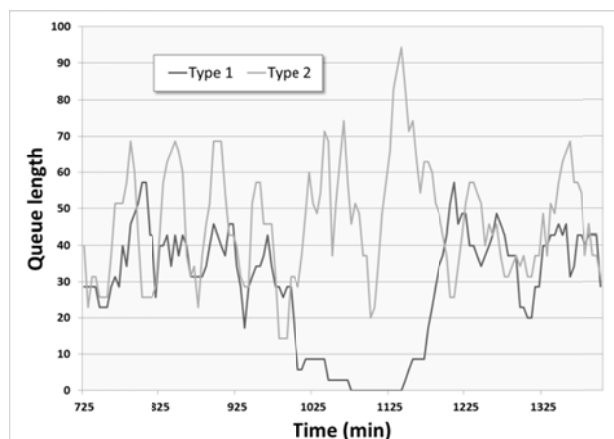


Figure 5: Number of waiting pallets on a critical section of conveyor system

Two main areas of use are distinguished when creating simulation from PLC level codes automatically:

- Off-line, i.e., planning level: Simulation is used for analyzing effects of changes in system, and fed with historical input data and relevant production data for parameterization.
- On-line, i.e., execution level: Simulation is applied for short-term “what-if” analysis. In case of certain deviation or disturbance has

occurred decision maker is able to test several predefined methods, scenarios for resolving problem.

Figure 5 shows results of a preliminary test run of number of waiting pallets on a critical section of conveyor system. Two types of pallets (Type1, Type2) were examined in this test.

Consequently, these results of simulation experiments can be easily adapted in real factory, thus enabling fast, flexible and smooth changes in factory, e.g., analyzing possible effects of changes in some part of PLC code of control system.

#### 5. CONCLUSIONS

This paper revealed a discrete-event simulation approach applied for decision support of control related production applications. Design phase is a significant part of a simulation project; hence reducing time of it heavily affects the effectiveness of the whole project. Automated data gathering supporting the buildup of simulation models is a possible solution to achieve this goal and is also a solution to create reusable models. Several approaches were studied in the topic and revealed that PLC codes store information needed to build up a simulation model. A new process for extracting topology and control logic data of system from PLC codes has been introduced. Data stored in production database were used to parameterize objects of the model and generating input for simulation experiments.

#### 6. ACKNOWLEDGEMENTS

Research has been partially supported by the Hungarian Scientific Research Fund (OTKA) grant “Production Structures as Complex Adaptive Systems” T-73376 and by National Office for Research and Technology (NKTH) grant “Digital, real-time enterprises and networks”, OMFB-01638/2009.

The research reported in this paper has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No: NMP2 2010-228595, Virtual Factory Framework (VFF).

#### REFERENCES

- Bagchi, S., Chen-Ritzo, C., Shikalgar, S.T., Toner, M., 2008. *A full-factory simulator as a daily decision-support tool for 300mm wafer fabrication productivity*, in Proceedings of the 2008 Winter Simulation Conference, S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds., pp. 2021-2029
- Banks, J., 1998. *Handbook of Simulation, Principles, Methodology, Advances, Application and Practice*. John Wiley & Sons Inc.
- Juyoung Wya, Sangwon Jeong, Byung-In Kim, Junhyuk Park, Jaejoon Shin, Hyunjoong Yoon, Sujeong Lee, 2011, *A data-driven generic simulation model for logistics-embedded assembly manufacturing*



lines. Computers & Industrial Engineering. Vol 60 Issue: 1. pp. 138-147.

- Kádár, B., Lengyel, A., Monostori, L., Suginishi, Y., Pfeiffer, A., Nonaka, Y. 2010. *Enhanced control of complex production structures by tight coupling of the digital and the physical worlds*, CIRP Annals - Manufacturing Technology, Vol 59, pp. 437–440.  
(<http://dx.doi.org/10.1016/j.cirp.2010.03.123>)
- Kwan Hee Han, Seock Kyu Yoo, Bohyun Kim, Geon Lee, 2010. *Rapid Virtual Prototyping of PLC-Based Control System*. ICAI'10 Proceedings of the 11th WSEAS international conference on Automation & information
- Law, A., Kelton, D., 2000. *Simulation modeling and analysis*, McGraw-Hill, New York.
- O'Reilly, J.J., Lilegdon, W.R., 1999. *Introduction to FACTOR/AIM*. In: Proc. of the 1999 Winter Simulation Conference, pp. 201-207.
- Park, Hyeong-Tae, Kwak, Jong-Geun, Wang, Gi-Nam, Park, Sang C., 2010., *Plant model generation for PLC simulation*. International Journal of Production Research Vol. 48, Issue 5, pp. 1517-1529.
- Pfeiffer, A., 2007. *Novel Methods for Decision Support in Production Planning and Control*, Thesis (PhD). Budapest University of Technology and Economics.
- Pfeiffer, A., Kádár, B., Szathmári, M., Popovics, G., Vén, Z., Monostori, L., 2009. *Self-building simulation tool for daily decision support in production control*, In proc. of the 7th International Workshop on Modelling & Applied Simulation, MAS 2009, 23-25 September, 2009, Tenerife, Spain, pp.: 246-254.
- Son, Y.J., Wysk, R.A. 2001. *Automatic simulation model generation for simulation-based, real-time shop floor control*, Computers in Industry 45. pp 291-308.
- Ryan, J., Heavey, C., 2006, *Process modeling for simulation*, Computers in Industry 57. pp. 437–450

## AUTHORS BIOGRAPHY

**Gergely Popovics**, graduated in 2008 at the Budapest University of Technology and Economics. Currently he is a research associate at the Engineering and Management Intelligence Laboratory (EMI) of the Computer and Automation Research Institute, Hungarian Academy of Sciences (SZTAKI), member of the Fraunhofer Project Center for Production Management and Informatics at SZTAKI (FhG Project Center). His current interest includes the simulation modelling of complex production systems and automatic identification technologies.

**András Pfeiffer**, earned his PhD in 2008 at the Budapest University of Technology and Economics. Currently he is a senior research fellow at EMI, project manager of FhG Project Center. His current interest

includes decision support in production planning and control, as well as the simulation and emulation modelling of complex production systems, self-building simulation systems.

**Botond Kádár** is a senior researcher at EMI, project manager of FhG Project Center. He obtained his MSc and Ph.D. degrees at the Budapest University of Technology and Economics, Hungary, in 1993 and 2002, respectively. His current interest includes production control, simulation and multiagent approaches for production engineering and manufacturing systems and he is involved in several research and development projects from these fields. Dr. Botond Kádár is author or co-author of 70 publications with over 120 citations.

**Zoltán Vén**, graduated in 2006 at the Budapest University of Technology and Economics. Currently he is a research associate at EMI, member of the FhG Project Center. His current interest includes reconfigurable manufacturing systems, as well as the simulation and emulation modelling of complex production systems, self-building simulation systems.

**Prof. László Monostori** acts as Deputy Director Research of SZTAKI, Head of the Engineering and Management Intelligence Laboratory, and Director of the FhG Project Center. He is also full time professor at the Department of Manufacturing Science and Technology, Budapest University of Technology and Economics. He is a Fellow and Council Member of the International Academy for Production Engineering (CIRP); Full Member of the European Academy of Industrial Management (AIM); Chairman of the Coordinating (CC) on Manufacturing and Logistics Systems, International Federation of Automatic Control (IFAC). He is Editor-in-Chief of the CIRP Journal of Manufacturing Science and Technology; Associate Editor of Computers in Industry, as well as the Measurement, and member of the editorial boards of other international scientific periodicals. For his research achievements published in more than 370 publications resulted in about 1800 independent citations and for his development activities – among others – the Dennis Gabor Prize was given to him in 2004. Prof. Monostori is a corresponding member of the Hungarian Academy of Sciences and member of the Hungarian Academy of Engineering.



# MODELLING AND ANALYSING THE IMPACT OF MANUAL HANDLING PROCESSES IN SEAPORTS

B. Scholz-Reiter<sup>(a)</sup>, M. Görge<sup>(b)</sup>, R. Matthies<sup>(c)</sup>

<sup>(a), (b)</sup> BIBA – Bremer Institut für Produktion und Logistik GmbH  
Hochschulring 20, 28359 Bremen  
Germany

<sup>(c)</sup> BLG Cargo Logistics GmbH & Co. KG.  
Neustädter Hafen, Schuppen 22 / Senator-Borttscheller-Straße, 28197 Bremen  
Germany

<sup>(a)</sup>[bsr@biba.uni-bremen.de](mailto:bsr@biba.uni-bremen.de), <sup>(b)</sup>[goe@biba.uni-bremen.de](mailto:goe@biba.uni-bremen.de), <sup>(c)</sup>[rmatthies@blg.de](mailto:rmatthies@blg.de)

## ABSTRACT

The handling of general cargo in seaports is a time-consuming and resource-intensive activity. In order to sustain competitive, ports have to offer efficient handling processes, leading to short vessel laytimes and fast handling operations. A possible solution approach is the implementation of novel handling technologies. This paper evaluates the potentials of a magnetic device for the handling of steel metal products in seaports. Due to the dynamics of that specific handling process and the interlinks to other actors in the process a simulative approach is applied for analyzing possible benefits concerning the handling frequency.

Keywords: material handling, seaport, process time variation, simulation

## 1. INTRODUCTION

The traffic of cargo is exposed to an intensive growth during the last years. Especially, the world-wide maritime traffic is strongly affected from this trend (Stopford 2009, Amerini 2007). Accordingly, efficient handling and warehousing processes are necessary to handle the increasing traffic and to stay competitive (Zondaga et al. 2010). Technical innovations are a driving force in the development of ports and naval processes. Especially general cargo, which cannot be handled with standard containers, comprise improvement potentials regarding technical automation. The handling of steel metal sheets is a classical example of such processes. This process is characterized by mechanical load handling devices like hooks, chains, ropes and belts, which are attached manually to the steel sheets. Thus, the handling of steel sheets is labour-intensive and inference-prone. Scholz-Reiter et al. 2010 identified four major weak points of these processes, and proposed the introduction of an innovative magnetic handling device. Especially, the process

stability can be improved by this new technology. It is well known, that automation of processes provides lower process variability and a higher degree of process reliability (Hopp and Spearman 2008, Groover 2008, Gudehus 2005). In order to evaluate the potentials of this novel technology in the early phases of planning quantitative analysis are necessary. Scholz-Reiter et al. (2010) stated that the introduction of an automatic magnetic handling device may reduce the variations of process times in the manual attachment. In this context it is assumed that the implementation of a magnetic device will lead consequently to higher performance of the total process. This paper investigates the connection between the process stability (in terms of reduced standard deviations of manual process times) and the performance of the total process with a discrete event simulation model. Therefore, two different statistical distributions are implemented and compared regarding stochastic variations process times. This paper will demonstrate the impact of these variations on the performance of the total process.

Therefore, this contribution is structured as follows: Section 2 gives a description of handling processes and possible potentials of novel handling technologies. Subsequently, section 3 discusses possible sources of stochastic variations in process times. The simulation approach is introduced in section 4. On this basis section 5 presents the results of numerical simulations experiments. Finally, section 6 gives a summary and provides an outlook with further research directions.

## 2. HANDLING OF STEEL SHEET PRODUCTS IN SEAPORTS

The handling of steel products is a classical general cargo process. Due to the size (4m x 12m) and the weight (up to 8.5t) of the steel sheets a containerized handling is not possible. The handling is done piece or bulk wise. In the conventional case the handling devices can be attached to multiple steel sheets as a steel ply at

ones. The number of sheets per handling operation depends on their weight. Normally, the crane, which is equipped with the mechanical handling device, can lift a weight up to 16t, which corresponds to two or three steel sheets (Scholz-Reiter et al. 2008).

The handling process can be divided into three parallel running sub processes: pick-up from storage, transfer to the bulk carrier and placing dunnage (Scholz-Reiter et al 2010). Figure 1 depicts the conventional handling process schematically.

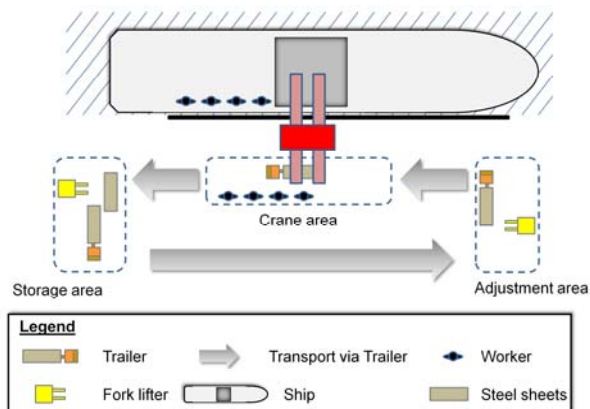


Figure 1: Scheme of the handling process (Scholz-Reiter et al. 2010).

The steel sheets are picked by a fork lifter in the storage area. By using a magnetic device the fork lifter is able to handle multiple sheets with one movement. It discharges the steel sheets on a special trailer which is designed to transport the steel sheet from the storage area to the crane area. Before handling with a crane, the steel sheets have to be adjusted in the so called adjustment area. Here a second fork lifter pushes the steel sheets in a predefined position. The crane can only pick up steel plies with a plain geometry. Thus, the process step of adjustment is necessary. Subsequently, the trailer transports the steel sheets directly to the crane area. In this area the crane lets down the mechanical handling device which consists of chains and four handling claws. Four workers attach these claws to the steel sheets. After the attachment, the workers have to press manually against the claws until the crane lifts the sheets slightly. This slightly lifting of the steel ply causes the correct mechanical force transmission between load and the handling device. After this step the workers leave the danger area near the hanging load. The crane waits until all workers are out of the danger area and turns the load into the ship. Inside the ship again four workers detach the material handling claws from the latest steel ply. After the detachment they place timber beams as dunnage on this ply. This step is indispensable for the unloading procedure in the port of destination. Without placing the dunnage there is no jacking point for further handling in this port.

Scholz-Reiter et al. (2010) identified four major weak points in this process and proposed a solution on the

basis of a magnetic handling device. Figure 2 presents these weak points and the corresponding potentials of the magnetic solution.

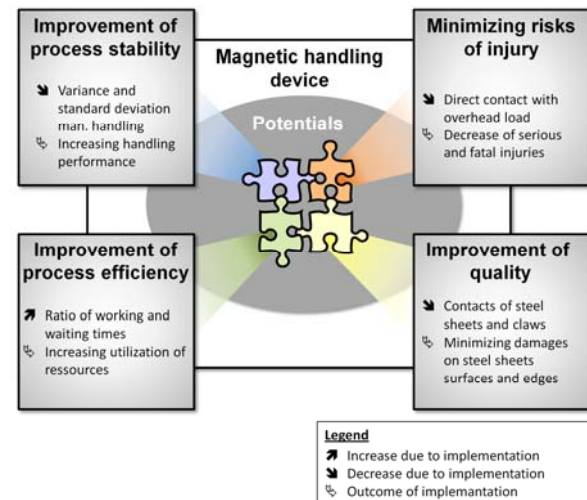


Figure 2: Weak points of the mechanic handling process and potentials of a magnetic handling technology.

The implementation of this technology minimizes the risk of serious injuries (e.g., bruises) during the attachment of the claws. Bruises are besides fractures a common type of injuries in the maritime sector (Ellis et al. 2010). Furthermore, it reduces the risk of scratches on the surface of the steel sheets. This effect helps to improve the product quality of the handled goods.

From an economical point of view the process stability and the process efficiency are crucial points. The manual attachment of the material handling claws is the bottleneck of the entire process. Due to the manual attachment there are deviations in process times. These deviations may affect the process times of subsequent steps. A result of these process delays is a reduced frequency of crane moves which corresponds to lower performance of the entire handling process. Table 1 confirms this assumption. It presents real process data concerning the handling process. The original data comprises data of 46 shifts. Table 1 presents the main performance indicators extracted from this dataset.

Table 1: Process performance

performance indicator	value
average amount of crane moves [moves/shift]	151.18
standard deviation [moves/shift]	41.43
average tonnage [t/h]	187.6
standard deviation tonnage [t/h]	64.97

In this context a crane move describes the complete cycle of the crane movement starting and ending at the same point (Arora and Shinde 2007). Table 1 shows a high proportion of standard deviation concerning the average amount of crane moves per shift. According to Hopp and Spearman (2008) the process variability can

be defined by the coefficient of variation (CV), which is the ratio of standard deviation and mean value. In the case at hand the CV is  $CV=0.274$  for the average amount of moves per shift. Regarding the tonnage this proportion is even bigger ( $CV=0.346$ ). In both cases the CV indicates that these processes underly strong variations.

It is assumed, that a reduction of standard deviations in this process step leads to higher handling quantities in the total process. Due to the strong interdependencies in this process (cf. Figure 1) an analytical analysis of the impact of standard deviations in the attachment processing times seems to be challenging. Thus, this objective paper proposes a simulation based approach for this analysis. In the following sources of variations in process times of manual attachment are discussed.

### 3. PROCESS STABILITY OF MANUAL PROCESSES

The problem of variability of processing times is not only limited to the material handling in seaports. Mapes et al. (2000) stated that the variability of processing times at single stations of production systems is a major driving force of the increasing complexity in planning and scheduling the overall production activities. In this context variability causes longer throughput times thwarts the delivery reliability, decreases the utilization of the system and lowers the productivity. The occurrence of variability in process times has different sources (Hopp and Spearman 2008):

- Natural variability: this category includes the inherent deviations which are caused by manual operations. It describes variations occurring in normal operations.
- Preemptive outages: unscheduled break downs of machines or resources belong to this category
- Nonpreemptive outages: downtimes of machines or resources which can be scheduled (e.g., maintenance activities)
- Rework: variability in the process output due to deficits in the product quality. This category also includes variations in the resources availability due to reworking activities.

In order to evaluate the potentials of the magnetic handling device against the current situation with the manual attachment especially the first category of the natural variability is of importance. This category comprises the characteristics of manual operations under normal operational conditions. Conventional approaches like the Methods-time measurement (MTM) methods aim at estimating mean values of manual activities, but they are not suitable to determine variability in these operations (Turek and Krengel 2007). Different authors propose modeling with statistic distributions like the normal distribution (e.g., Tiacchi and Saetta 2007), exponential distribution (e.g., Doerr

and Arreola-Risa 2000) or the weibull distribution (e.g., Buxey and Sadjadi 1976). Due to the characteristic of manual operations manual work studies showed that often distributions with a positive skew are observed. Turek and Krengel (2007) investigated empirically processing times of manual order picking activities. In this context they stated that the type and the shape of the underlying distribution (skewness) may depend on the process type. On this basis this contribution uses two types of distribution for the analysis. First, a normal distribution is used due to its comprehensibility. Second, a gamma distribution is used as a representative of a positive skewed distribution.

### 4. SIMULATION SCENARIO

The general scenario depicted in Figure 1 can be modeled as a closed queuing network with three service stations as indicated in Figure 3.

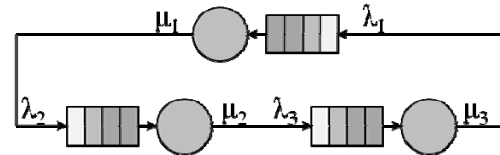


Figure 3: Scenario representation as queuing network.

All processes in the respective areas are modeled as service station with a certain service rate ( $\mu_1$ ,  $\mu_2$  and  $\mu_3$ ). The round course movement of the trailers in this model is given by the arrival rates ( $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ ), which directly depend on the service rates of the servers. The amount of jobs in closed queueing networks is fix and known in advance (Bocharov et al. 2004). In this case the amount of jobs represents the amount of trailers used in the process. In the case at hand, the service stations 2 and 3 represent the processes in the storage area and in the adjustment area and station 1 stands for the crane area. In order to evaluate the impact of randomness and standard deviations of the manual attachment, service rate  $\mu_1$  is modeled by statistical distribution. The remaining service rates are kept constant in the simulation study. Additionally, the service rate  $\mu_1$  covers the movement of the crane. This means that a succeeding trailer can only be served, if the crane completes an entire move. Thus, a recovery time is modeled for this station. This is done by defining a minimum cycle time  $C_m$  for the crane move.

A further constrains coming from the real process is the transport distances between storage and the crane area. These distances are considered by traveling time for passing theses distance, denoted by  $d$ . The duration of the non handling processes is cumulated in  $D$  and defined as follows:

$$D = d + \mu_1 + \mu_2 \quad (1)$$

In the real world process the amount of trailers is determined by the distances. Accordingly, the planning of the trailer operations aims at balancing the total

process times and increasing the utilization of the crane. Before starting the process, the amount of trailers ( $n$ ) is determined as follows:

$$n = \text{int}\left(\frac{D}{E(\mu_1) + C_m}\right) + 1 \quad (2)$$

In the case that the remainder of the division equals zero the amount of trailers is ideally attuned to the processing times of the crane. The simulation study focuses on three different degrees of trailer balancing: a balanced situation, an under-balanced situation and an over-balanced situation. All degrees are based on a predefined travel time  $D_0$  according to equation 1. The under- and over-balanced situations are generated by the introduction of a scale factor  $\delta$ , which compresses or stretches the predefined travel time  $D_0$ :

$$n = \text{int}\left(\frac{\delta \cdot D_0}{E(\mu_1) + C_m}\right) + 1 \quad (3)$$

Accordingly the situations are defined as follows:

- The balanced situation  $\delta=1$
- The under-balanced situation  $\delta<1$
- The over-balanced situation  $\delta>1$

In order to model variation in the process times of station 1, the processing rate  $\mu_1$  is set to different probability distributions: a normal and a gamma distribution are chosen. The simulation results, using both types of distribution will be compared.

A disadvantage of the normal distribution in this case is the modeling of the lower bound of the cycle time of the crane. A normal distribution has an infinite co-domain, which contradicts the modeling of a lower bound. However, this is taken into account by cutting off random values below this lower bound. The gamma distribution seems to be more suitable to the case at hand. This distribution is defined in the interval  $[0, \infty]$ . Hence, a lower bound of processing times can be defined by shifting the distribution to this lower bound. The distribution is described by two parameters  $\alpha$  and  $\beta$ , which determine the expectancy value and the variance. As mentioned before, the variance is seen as running variable and the expectancy value is kept constant. Thus the parameters used for building the corresponding gamma distributions are given by:

$$\alpha = \frac{E(\mu_1)^2}{Var(\mu_1)} = \frac{E(\mu_1)^2}{\sigma_{\mu_1}^2} \quad (4)$$

$$\beta = \frac{\alpha}{E(\mu_1)} \quad (5)$$

Figure 4 depicts exemplarily different density functions of the gamma distribution with a fixed expectation value and different variances.

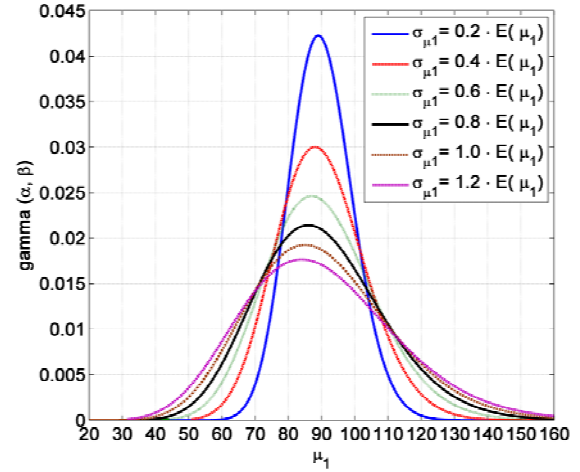


Figure 4: Density functions of the gamma distribution with constant mean value  $E(\mu_1)=90$  and varying standard deviation

Due to the mentioned extensions, like modeling with normal distribution or recovery times, a conventional modeling of the scenario as a standard closed queuing network and an analytical evaluation is not possible. Thus, the scenario described is implemented to a discrete event simulation model and analyzed by the evaluation of different simulation experiments.

## 5. SIMULATION AND RESULTS

### 5.1. Simulation setup

According to these preliminary considerations a set of simulation experiments is defined (Table 2). Thereby, the trailer balancing is defined by the variable  $\delta$ . An under-balance situation is represented by the choice of  $\delta<1$ . With  $\delta=1$  a balance situation is considered. By contrast, the choice of  $\delta>1$  an over-balance situation is modeled. In addition to this, the variance in every simulation run is varied as a percentage value of the predefined expectation value  $E(\mu_1)$  in steps of 1%.

Table 2: Simulation setup

Parameter / variable	Type	Value	Dimension
$D_0$	constant	600	seconds [s]
$C_m$	constant	210	seconds [s]
$E(\mu_1)$	constant	90	seconds [s]
$Var(\mu_1)$	variable	$\{0.01 \cdot E(\mu_1) \dots 1.2 \cdot E(\mu_1)\}$	seconds [s]
$\delta$	variable	$\{0.5, 1, 1.5\}$	[-]

The evaluation of a specific parameter constellation is done by calculating the number of moves per hour of a simulation run. In order to reduce stochastic effects, a certain parameter constellation is simulated  $10^3$  times.



The average number of moves per hour is used for the evaluation.

## 5.2. Simulation results

Figure 5 presents the results of the normal distribution. It depicts the average number of moves per hour for each combination of  $\delta$  and  $\sigma$ . Additionally, the black dotted plane in Figure 5 indicates the theoretical maximal amount of moves per hour on the basis of the expectation value and the recovery time of the crane.

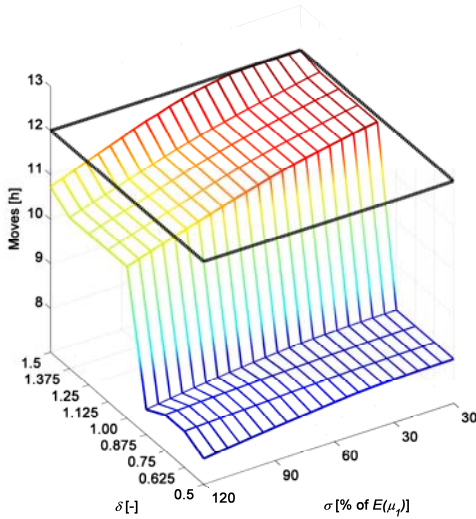


Figure 5: Simulation results of the normal distribution

As expected, in the under-balanced situation the performance is lower compared to the balanced and over-balanced situation. This effect is mainly caused by the lower number of circulating trailers, which lead to waiting times of the crane. Accordingly, the crane is underutilized and the average number of crane moves is lower compared to both other situations. However, the impact of rising standard deviations in the manual attachment process can be already observed in the under-balanced situation: The average amount of crane moves per hour decreases with an increase of the standard deviation.

This effect is even stronger in the balanced and over-balanced situation. In the balanced situation the performance of the process decreases on average about 0.07 moves per hour with an increase of the standard deviation of one unit. The biggest performance difference in the balanced situation is 12.22 %. A similar impact of the process variations can be observed in the over-balanced situation. For lower values of the standard deviation the process performance is near to its maximal capacity, but with an increase of the standard deviation the total performance of the process decreases like in the balance and under-balanced situation.

Similar effects can be observed for the second distribution type used in this study. Figure 6 presents the simulation results for the gamma distribution for

each combination of  $\delta$  and  $\sigma$ . In order to provide comparability with the previous simulation run the black dotted surface represents the maximal possible performance for this scenario. Similar to the results of the normal distribution, Figure 6 shows steps in the total performance depending of the chosen value of  $\delta$ . The impact of the standard deviation is comparable to figure 5. Only the results for  $\delta=1.5$  differ. From this point an additional trailer is required. In this particular case the performance decreases slightly with an increase of the standard deviation.

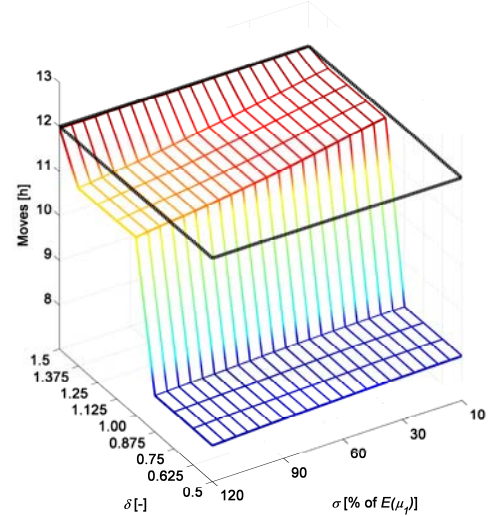


Figure 6: Simulation results of the gamma distribution

According to this result the impact of increasing variance can be compensated by an over-balance situation. Although an over-balanced operation compensates these effects and leads to a higher utilization of the crane, it causes to worse utilization of resources like trailers. In the case at hand the average utilization of the trailers for  $\delta=1.5$  is 49.59% compared to 62.78 for  $\delta=1$ .

Summarizing the results of the simulation study it can notice that the impact of process time variations is similar for both distributions (normal distribution and gamma distribution) in the balanced and under-balanced case. There are only differences for the highly over-balanced situation ( $\delta=1.5$ ). In this case the performance decreases less with an increasing standard deviation by applying a gamma distribution. A detailed view on the utilization of the trailers showed, that this compensations is won by the expense of low trailer utilization.

## 6. SUMMARY AND OUTLOOK

This paper presented an analysis of the impact of process variations on the performance of the specific handling process in seaports. A general model of the process has been introduced and implemented to a simulation model. The detailed analysis of the simulations results showed a characteristic curve of the

process performance depending on the process variations. For this purpose two different statistic distributions were under consideration. The results show the sensitivity of the process against variations in process times for both types of distribution. A comparison of the trailers utilization shows a potential tradeoff dilemma between sensitivity against variations and the trailers utilization for the gamma distribution. In general these results indicate that a reduction of variations in process times, i.e. induced by a magnetic handling device, leads to a higher process performance. This effect occurs especially in the balanced situation, which characterizes the normal operation mode. Further work will focus on a holistic evaluation of the implementation of a magnetic handling covering process related and an economic analysis.

## ACKNOWLEDGMENTS

This research is funded by the German Ministry of Economics and Technology (BMWi) based on a decision of the German Parliament. The project "Advanced stowage and handling technology for steel products (ISUS - Intelligente Stau- und Umschlagstechnologie für Stahlprodukte)" is situated in the funding programme Innovative Seaport Technologies II (ISETEC II).

## REFERENCES

- Amerini, G. 2007. Maritime Transport of Goods and Passengers 1997 - 2005. *Statistics in focus* 94/2007. London: Eurostat.
- Arora, K.C. and Shinde, V.V., 2007. *Aspects of materials handling*. New Delhi, Boston: Laxmi cop.
- Bocharov, P.P., D'Apice, C., Pechinkin, A. V. and Salerno, S., 2004. *Queueing theory*. Utrecht: VSP.
- Buxey, G. M. and Sadjadi, D., 1976. Simulation studies of conveyor-paced assembly lines with buffer capacity, In: *International Journal of Production Research*, 14: 5, 607-624.
- Doerr, K.H. and Arreola-Risa, A., 2000. A worker-based approach for modeling variability in task completion times. In: *IIE Transactions*, 32: 7, 625-636.
- Ellis, N., Bloor, M. and Sampson, H., 2010. Patterns of seafarer injuries. *Maritime Policy & Management*, 37(2), 121 - 128.
- Groover, M. P., 2008. *Automation, production systems, and computer-integrated manufacturing*. Upper Saddle River: Pearson/Prentice Hall.
- Gudehus, T., 2005. *Logistik: Grundlagen, Strategien, Anwendungen*. Berlin: Springer.
- Hopp, W.J. and Spearman, M.L., 2008. *Factory physics*. Boston: McGraw-Hill/Irwin.
- Mapes, J., Szejczewski, M. and New, C., 2000. Process variability and its effect on plant performance, In: *International Journal of Operations & Production Management*, 20(7), pp. 792-808.
- Scholz-Reiter, B., Hinrichs, U. and Tervo, J.T., 2008. Advanced handling of steel products in seaports using magnet systems. In: Bruzzone, A., Longo, F., Merkuriev, Y., Mirabelli, G., Piera, M. eds. *11th International Workshop on Harbour, Maritime and Multimodal Logistics Modeling and Simulation (HMS2008)*, 202-207.
- Scholz-Reiter, B., Tervo, J.T., Görges, M. and Matthies, R. 2010. Modelling and analysis of mechanical steel handling processes. In: Bruzzone, A., Chunha, G., Frydman, C., Giambiasi, N., Mekouar, K., Merkuriev, Y., Piera, M. eds. *13th International Workshop on Harbour, Maritime and Multimodal Logistics Modeling and Simulation (HMS2010)*. 73-78.
- Stopford, M., 2009. *Maritime Economics*, Taylor & Francis.
- Tiacci, L. and Saetta, S., 2007. Process-oriented simulation for mixed-model assembly lines. In: *Proceedings of the 2007 summer computer simulation conference (SCSC)*. Society for Computer Simulation International, San Diego, CA, USA, 1250-1257.
- Turek, K., Kregel, M., 2008. Modellierung der Dynamik manueller Operationen in logistischen Systemen. In: 4. Fachkolloquium der Wissenschaftlichen Gesellschaft für Technische Logistik an der Technischen Universität Chemnitz, Subbert, . (Ed.), pp. 50-59.
- Zondaga, B., Buccia, P., Gützkowb, P. and de Jonge, G., 2010. Port competition modeling including maritime, port, and hinterland characteristics. *Maritime Policy & Management* 37(3), 179-194.

## AUTHORS BIOGRAPHY

**Prof. Dr.-Ing. Bernd Scholz-Reiter** is head of the department Planning and Control of Production Systems at the University of Bremen and managing director of the BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen. He is Editor of the professional journals *Industrie Management* and *PPS Management*. Prof. Scholz-Reiter is a corresponding member of C.I.R.P., the International Institution for Production Engineering Research and also a member of the Scientific Advisory Board of Bundesvereinigung Logistik e.V. He is also vice president of the German Research Foundation (DFG).

**Dipl.-Wi.-Ing. Michael Görges** works as a research scientist at the BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen in the division Intelligent Production and Logistic Systems.

**Ralf Matthies** leads the Profitcenter West at BLG Cargo Logistics GmbH & Co.KG in the Neustädter Hafen in Bremen, Germany

# NUMERICAL ANALYSIS OF WIND INDUCED PRESSURE LOADS ON AN INTEGRATED ROOF-BASED PHOTOVOLTAIC SYSTEM

Marco Raciti Castelli <sup>(a)</sup>, Sergio Toniato <sup>(b)</sup>, Ernesto Benini <sup>(c)</sup>

<sup>(a)</sup> Department of Mechanical Engineering – University of Padua  
Via Venezia, 1 – 35131 Padova, Italy

<sup>(b)</sup> ESPE S.r.l.

Via Cappello, 12/A – 35010 San Pietro in Gu (PD), Italy

<sup>(c)</sup> Department of Mechanical Engineering – University of Padua  
Via Venezia, 1 – 35131 Padova, Italy

<sup>(a)</sup> [marco.raciticastelli@unipd.it](mailto:marco.raciticastelli@unipd.it), <sup>(b)</sup> [stoniato@espe.it](mailto:stoniato@espe.it), <sup>(c)</sup> [ernesto.benini@unipd.it](mailto:ernesto.benini@unipd.it)

## ABSTRACT

Wind loads on the roof of a civil structure inside an industrial area still represent a great challenge for structural engineers. The quantitative amounts of force and pressure coefficients depend on the shape, size and orientation of the building and on its interaction with the surrounding environment.

The static loads due to wind pressure on a 5 deg pitched roof-based integrated photovoltaic system have been estimated by means of a 2D numerical simulation of the flow field around the T&T building in Valdagno (Italy). A constant wind velocity profile, based on the maximum reference wind speed in the building site (peak gust speed worked out for 50 years return period) and on the local roughness coefficient, has been simulated.

The distribution of wind-induced loads as a function of the panel row position and the amplitude of the recirculation region downstream the building have been determined, allowing a numerical quantification of the effect of the building geometry on the pressure loads on top of the roof.

Keywords: CFD, wind, building, roof

## 1. INTRODUCTION AND BACKGROUND

Due to the growing demand for renewable energy sources, the manufacturing of solar cells and photovoltaic arrays has shown a considerable advancement in recent years [1]. As of 2010, solar photovoltaic generated electricity in more than 100 countries and, while yet comprising a tiny fraction of the 4.8 TW total global power-generating capacity from all sources, is the fastest growing power-generation technology in the world [2]. However, as pointed out by Cosoiu et al. [3], the structure of the photovoltaic (PV) panel is rather flexible, continuous and fragile,

sustained only by a thin framework. These features make it easily damageable by high winds and, in order to prevent such events, wind engineering experimental tests and numerical simulations are demanded if a more optimized and cheaper solution for a solar panel framework is required.

As focused by Krishna et al. [4], wind causes a random time-dependent load, which can be considered as a mean plus a fluctuating component. Strictly speaking, all civil structures experience dynamic oscillations due to the fluctuating component (gustiness) of wind, however, in short rigid structures, such oscillations are insignificant and the buildings can therefore be satisfactorily treated as being subjected to an equivalent static pressure. This approach is taken by most Codes and Standards. However, the response of a civil structure to high wind pressure depends not only on the geographical location and proximity of other obstructions to airflow, but also on the characteristics of the structure itself. As a matter of fact, most buildings present “bluff forms” to the wind, making it difficult to ascertain the wind forces accurately. Thus, the problem of bluff-body aerodynamics remains largely in the empirical/descriptive realm of knowledge. Furthermore, the flow patterns (and hence the wind pressure/forces) change with the Reynolds number, making the direct application of wind tunnel test results to real structures quite difficult [5]. Nevertheless, the limitations in accurate aerodynamic databases can be overcome by using advanced Computational Fluid Dynamics (CFD) codes, which can outflank the lack of experimental data thanks to their inherent ability to determine the aerodynamic components of actions through the integration of the Navier-Stokes equations. Performing CFD calculations can provide knowledge about the flow-field around the building in all its details, such as velocities, pressure, etc. Moreover, all types of useful graphical presentations, such as flow lines, contour lines



and iso-lines are readily available. As suggested by Jensen et al. [6], this stage can be considered as if an accurate wind-tunnel study or an elaborate full-scale measurement campaign had been conducted. The numerical prediction of wind loads on buildings as a branch of Computational Wind Engineering (CWE) was well introduced by Franke [7], who described the different simulation approaches with their corresponding basic equations and the necessary turbulence models.

Wind loading on solar panels depends on three basic elements: the wind speed, the height of the panel above the roof, and the relative location of the panel on the roof: in fact, pressure loads will depend on the location of the module on the roof (with different net pressure coefficients applied for those near the roof edge), whether the roof has a parapet and whether the PV support structure is open or fully clad. Moreover, as pointed by Dalglish [8], roof angle strongly affects the flow around a low-rise building: as a direct consequence, the flow field over a 5 deg pitched roof building is dominated by flow separation.

Hoxey et al. [9] performed both full-scale experiments and CFD calculations in order to investigate the main geometric parameters that affect wind loads on low-rise buildings. The study focused on the use of numerical simulations so as to assess the sensitivity of wind loads to changes in height, span and roof pitch.

More recently, Dagnew et al. [10] performed several comparison between CFD and experimental analysis, in order to assess the potential use of numerical wind load predictions approaches for practical use. The work revealed the suitability of CFD tools for preliminary assessments and detail explanation of complex building aerodynamic characteristics.

In order to develop a preliminary procedure to be used as a guidance in selecting the appropriate grid configuration and corresponding turbulence model for the prediction of the flow field over a two-dimensional roof architecture dominated by flow separation, Raciti Castelli et al. [11] recently tested the capability of several turbulence models to predict the separation that occurs in the upstream sector of the roof and the extension of the relative recirculation region for different vertical longitudinal positions, respectively from the upstream leading edge to the downstream bottom edge of a reference model building. Also spatial node distribution was investigated, in order to determine the best compromise between numerical prediction accuracy and computational effort. The numerical code proved *Standard k-ε* turbulence model to be quite accurate in predicting the flow-field features, especially after the recirculation region in the upstream portion of the model roof. On the basis of this preliminary study, in the present work the flow field over a 5 deg pitched roof-based integrated photovoltaic system was numerically investigated, in order to determine the distribution of wind loads as a function of

the panel row position on the roof and the amplitude of the recirculation region downstream the building.

## 2. THE CASE STUDY

The proposed numerical simulations were based upon the T&T building in Valdagno (Northern Italy). Figure 1 shows an aerial view of the building site, located inside an industrial area placed on a wind plain close to a low hilly area.



Figure 1: Aerial view of the T&T building site (the analyzed building is evidenced by the red rectangle)



Figure 2: Top view of the T&T building double 5 deg gabled roof enclosed by a continuous brick fence

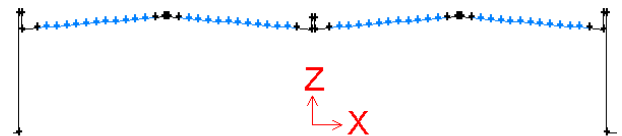


Figure 3: Main vertical section of the T&T building

Table 1: Main dimensions of the T&T building

$H_{\text{building}}$ [m]	12
$L_{\text{building}}$ [m]	60
$\alpha$ [deg]	5
$H_{\text{lateral fence}}$ [m]	1.6
$H_{\text{central fence}}$ [m]	1.05

As can be seen from Figures 2 and 3, the T&T building is a rectangular structure characterized by a double 5 deg gabled roof enclosed by a continuous lateral brick fence 1.6 m high and a central one 1.05 m high. Table 1 summarizes the main geometrical features of the analyzed building. A total number of 52 rows of PV panels are to be deployed on the top of the roof, 13 for each roof pitch.

### 3. MODEL GEOMETRY

In order to simulate a wind flow directed orthogonally into the face of the PV panels, a 2D simulation was performed. Table 2 summarizes the main geometrical features of the computational domain, which is also reproduced in Figure 4.

Table 2: Main geometrical features of the computational domain

$L_{\text{domain}}$ [m]	2400
$H_{\text{domain}}$ [m]	600
$L_1$ [m]	600

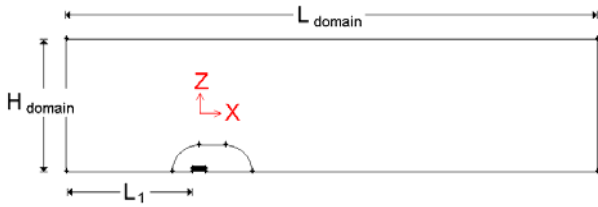


Figure 4: Main dimensions of the computational domain

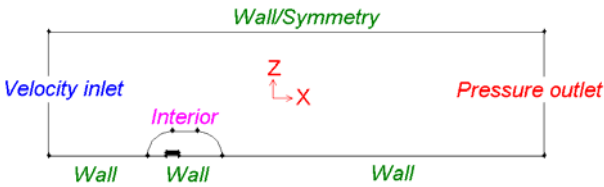


Figure 5: Boundary conditions of the computational model

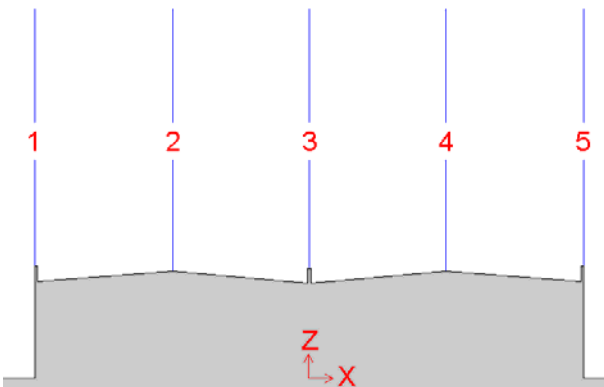


Figure 6: Displacement of the five reference positions (along the roof length) that were used for x-component mean velocity computation

The numerical model boundary conditions are represented in Figure 5. As suggested by Raciti Castelli

et al. [11] both “Wall” and “Symmetry” boundary conditions were tested for the upper portion of the computational domain, and their influence on the numerical results proved completely negligible.

Table 3: Normalized x-coordinates of the five reference positions with respect to the model building length (the origin of the coordinate reference system is located at the model building center, as evidenced from Figure 3)

Reference position No.	$x_{\text{ref},\text{norm}}$
1	-0.50
2	-0.25
3	0.00
4	0.25
5	0.50

The vertical profiles of the x-component mean velocity were computed at five reference positions along the roof length. Figure 6 shows the displacement of the reference positions, whose normalized x-coordinates with respect to the model building length, defined as:

$$x_{\text{ref},\text{norm}} = x_{\text{ref}}/L_{\text{building}} \quad (1)$$

are reported in Table 3.

### 4. SPATIAL DOMAIN DISCRETIZATION

An isotropic unstructured mesh was created around the model building. Considering their features of flexibility and adaption capability, unstructured meshes are in fact very easy to obtain, for complex geometries, too, and often represent the “first attempt” in order to get a quick response from the CFD in engineering work.

The same spatial grid resolution suggested by Raciti Castelli et al. [11] was adopted for the present calculations. In Table 4 the characteristic data of the adopted grid architecture are reported, as a function of the normalized grid resolution on the building, defined as:

$$Res_{\text{building}} = \Delta g_{\text{building}}/H_{\text{building}} \quad (2)$$

and as a function of the normalized grid resolution on outer computational domain, defined as:

$$Res_{\text{domain}} = \Delta g_{\text{domain}}/H_{\text{domain}} \quad (3)$$

Table 4: Characteristic data of the adopted grid architecture

$Res_{\text{building}}$ [-]	Growth factor [-]	$Res_{\text{domain}}$ [-]
0.025	1.25	0.07

For further details upon the validation procedure and the reliability of the adopted numerical settings (as

far as grid resolution and turbulence model are concerned), see [11].

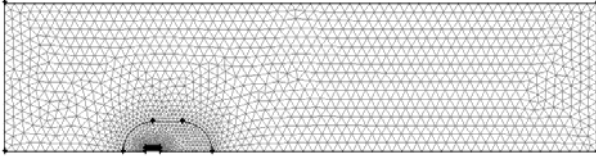


Figure 7: Main geometrical features of the adopted grid resolution

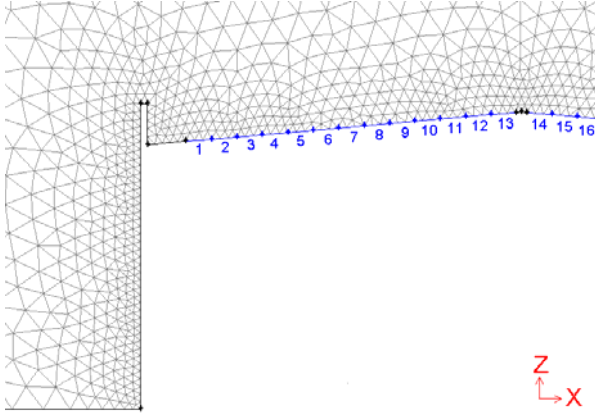


Figure 8: Main geometrical features of the adopted grid refinement near the building and numbering of the PV rows on the top of the roof, respectively from 1 to 52 starting from the building upwind leading edge (on the left)

Figures 7 and 8 show the main features of the adopted grid. Figure 8 shows also some of the numbered panel rows on the top of the roof, respectively numbered from 1 to 52 starting from the building upwind leading edge.

Table 5: Main coefficients adopted for the calculation of the maximum reference wind speed on the T&T building site, according to [12]

$v_{b,0}$ [m/s]	25
$a_0$ [m]	1000
$a_s$ [m]	60
$v_b$ [m/s]	25
$c_t$ [-]	1
$k_r$ [-]	0.22
$z_0$ [m]	0.30
$z_{min}$ [m]	8
$c_e(H_{building})$ [-]	1.91
$v_{b,max}$ [m/s]	34.5

## 5. DETERMINATION OF INLET WIND VELOCITY PROFILES

A constant velocity profile, based on the maximum reference wind speed in the building site (peak gust speed worked out for 50 years return period) was

computed. After determining from [12] the values of  $v_{b,0}$ ,  $a_0$ ,  $c_t$ ,  $k_r$ ,  $z_0$  and  $z_{min}$  for the building site, the reference wind speed was determined as:

$$v_b = v_{b,0} \quad (4)$$

being:

$$a_s \leq a_0 \quad (5)$$

and the coefficient of exposure for the building site was determined as:

$$c_e(H_{building}) = k_r^2 c_t \ln(H_{building}/z_0) [7 + c_t \ln(H_{building}/z_0)] \quad (6)$$

being:

$$H_{building} \geq z_{min} \quad (7)$$

The maximum reference wind speed for the building site was eventually determined as:

$$v_{b,max} = [v_b^2 c_e(H_{building})]^{0.5} \quad (8)$$

Table 5 summarizes the main coefficients adopted for the calculation of the maximum reference wind speed.

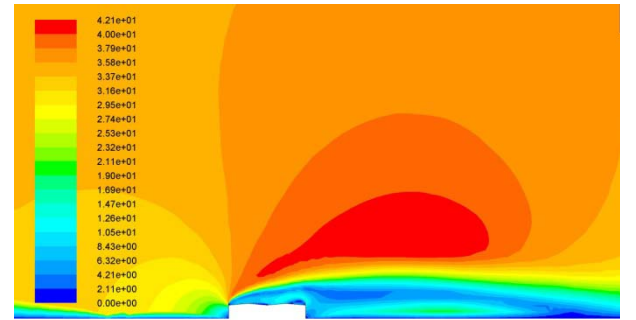


Figure 9: Contours of absolute velocity magnitude [m/s] on the surroundings of the T&T site

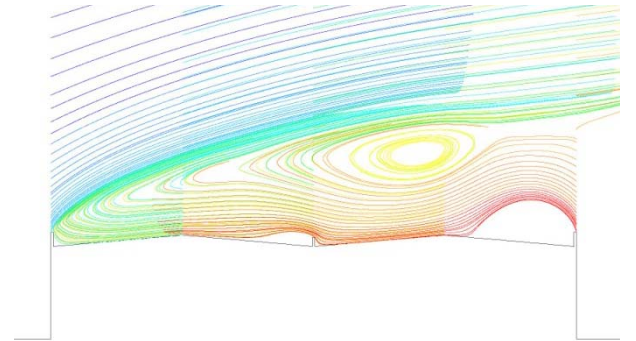


Figure 10: Visualization of absolute pathlines, colored by particle variables and starting from the five reference positions. A huge separation bubble associated to the vertical recirculation zone on the top of the building is visible



## 6. TURBULENCE MODEL AND CONVERGENCE CRITERIA

Simulations were performed using the commercial RANS solver ANSYS FLUENT®, which implements 2-D Reynolds-averaged Navier-Stokes equations using a finite volume-finite element based solver. A segregated solver, implicit formulation, was chosen for unsteady flow computation. The fluid was assumed to be incompressible, being the maximum fluid velocity on the order of 42 m/s. *Standard k-ε* model was used for turbulent calculations, as suggested from [11].

As a global convergence criterion, residuals were set to  $10^{-5}$ . The simulations, performed on a 8 processor, 2.33 GHz clock frequency computer, required a total CPU time of about 3 hours.

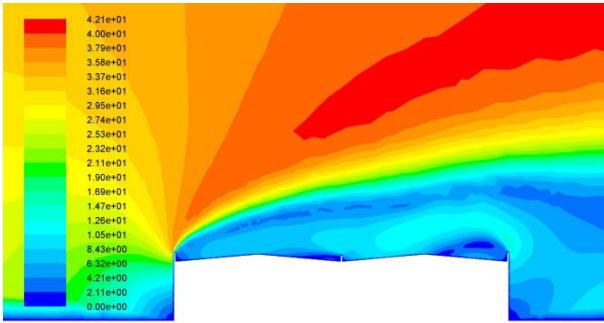


Figure 11: Contours of absolute velocity magnitude [m/s] close to the T&T building

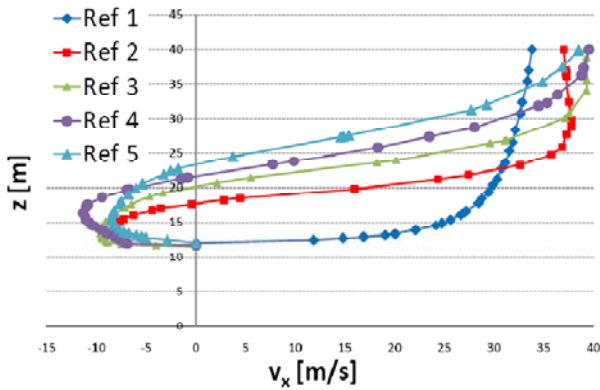


Figure 12: Computed x-velocity profiles for the five reference positions

## 7. RESULTS AND DISCUSSION

Figure 9 shows the contours of absolute velocity magnitude on the surroundings of the T&T site. As can be clearly seen, velocities close to the structure result much higher (up to 42.1 m/s) with respect to the constant (34.5 m/s) inlet velocity profile.

A recirculation zone on top of the roof and downstream of the building can be seen from Figure 10, representing the velocity pathlines colored by particle variables and also from Figure 11, showing the contours of absolute velocity magnitude close to the T&T building.

Figure 12 shows the computed x-velocity profiles for the five reference positions on the top of the T&T

building. Flow separation is clearly visible, as well as the recirculation zone. It can also be noticed that, while reference position No. 1 presents no recirculation, the amount of reverse flow increases starting from reference position No. 2 and shows a peak in correspondence of reference position No. 4 (violet line).

Figures from 13 to 15 show the horizontal, vertical and resultant forces for unit length on solar panel rows along the T&T building as a function of the PV panel row number.

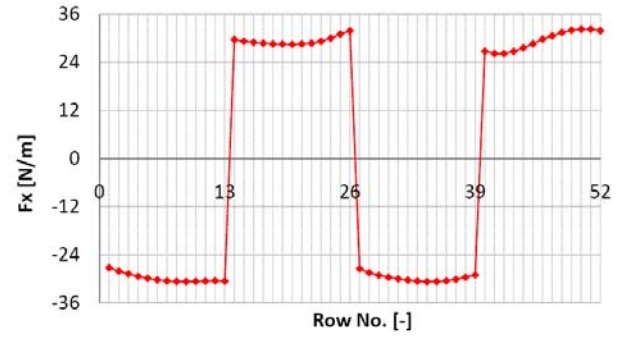


Figure 13: Horizontal force on PV panel rows for unit length along the T&T building

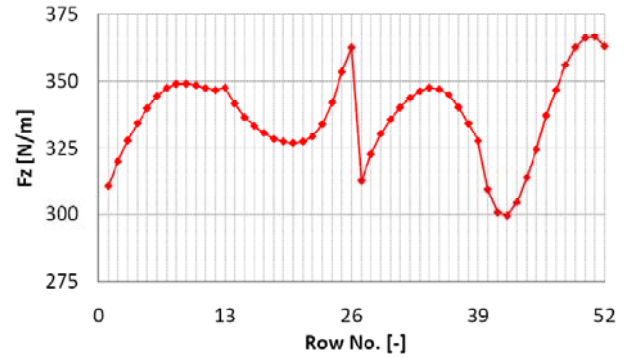


Figure 14: Vertical force on PV panel rows for unit length along the T&T building.

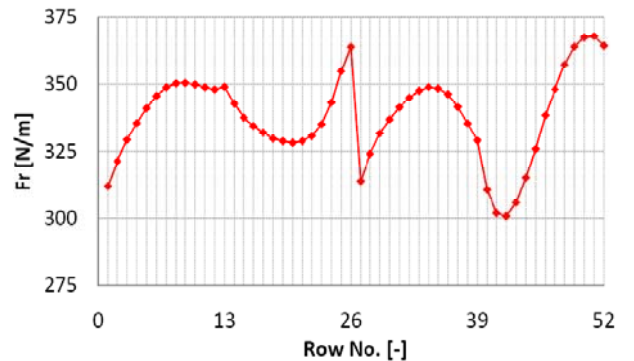


Figure 15: Resultant (total) force on PV panel rows for unit length along the T&T building

It can be noticed that, thanks to the continuous lateral brick fence on the whole perimeter of the roof, no downward force is registered on the PV panel rows. On the contrary, all panels are subjected to upwards thrusts, due to the low pressure in the recirculation zone

on the top of the building, as can be seen from Figures 14 and 16, showing the contours of static pressure close to the T&T site. It can be argued that the described phenomenon is related to a vertical suction occurring in the whole roof, due to a dramatic decrease of static pressure, as a consequence of the large separation bubble on the top of the building.

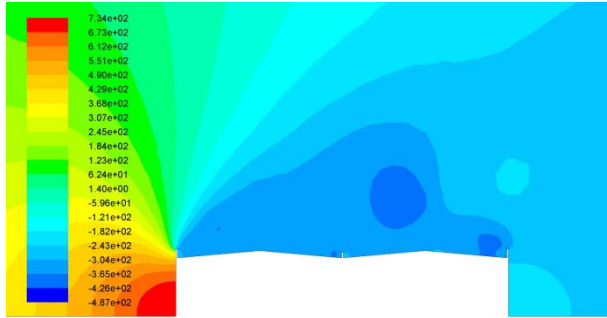


Figure 16: Contours of static pressure [Pa] close to the T&T building

As can be seen, the horizontal force is negligible if compared to the vertical one, being the latter nearly 10 times higher. A discontinuity in the horizontal force per unit length can be observed in correspondence of the top/bottom of the roof pitches. Also the vertical force presents a discontinuity in correspondence of the center of the building.

A periodicity in the vertical force distribution is registered, each period corresponding to a double-pitched portion of the roof. Moreover, the influence of the central brick fence on wind induced pressure loads appears to be negligible.

## 8. CONCLUSIONS AND FUTURE WORKS

A numerical model for the evaluation of wind induced pressure loads on a 5 deg pitched roof-based integrated PV system was presented. A constant wind velocity profile, based on the maximum reference wind speed in the building site (peak gust speed worked out for 50 years return period) and on the local roughness coefficient, was simulated.

A large recirculation zone was registered, causing reverse flow in the vertical x-velocity profiles on the top of the building.

It was proved that, thanks to a continuous lateral brick fence on the whole perimeter of the roof, no downward force is registered on the PV panel rows. On the contrary, all panels resulted subjected to upward thrusts, due to the low pressure in the recirculation zone on the top of the building. The influence of the central brick fence on wind induced pressure loads resulted negligible.

Further 3D work should be performed in order to investigate the influence of a lateral wind flow on the PV panel rows.

## ACKNOWLEDGMENTS

The present work was performed in cooperation with E.S.P.E. S.r.l., San Pietro in Gu (Italy) in order to develop a numerical model for the prediction of the turbulent flow field around the T&T building, Valdagno (Italy).

## NOMENCLATURE

$a_0$ [m]	reference height above sea level for T&T building site
$a_s$ [m]	height above sea level for T&T building site
$c_e(H_{\text{building}})$ [-]	coefficient of exposure for T&T building site and height
$c_t$ [-]	coefficient of topography for T&T building site and height
$F_x$ [N]	horizontal force on PV panel rows for unit length
$F_r$ [N]	resultant (total) force on PV panel rows for unit length
$F_z$ [N]	vertical force on PV panel rows for unit length
$H_{\text{building}}$ [m]	building height
$H_{\text{lateral fence}}$ [m]	lateral brick fence height
$H_{\text{central fence}}$ [m]	central brick fence height
$H_{\text{domain}}$ [m]	computational domain height
$k_r$ [-]	reference roughness coefficient at T&T building site
$L_1$ [m]	distance between computational domain inlet condition and building
$L_{\text{building}}$ [m]	building length
$L_{\text{domain}}$ [m]	computational domain length
$\text{Res}_{\text{domain}}$ [-]	normalized grid resolution on outer computational domain
$\text{Res}_{\text{building}}$ [-]	normalized grid resolution on the building
$v_{b,0}$ [m/s]	basic wind speed at T&T building site
$v_b$ [m/s]	basic wind speed at T&T building site and height above sea level
$v_{b,\text{max}}$ [m/s]	maximum reference wind speed at T&T building site
$v_m(z)$ [m/s]	average wind velocity profile at T&T building site
$v_x$ [m/s]	x-component of wind velocity on top of the T&T building
$x_{\text{ref}}$ [m]	x-coordinate of the reference position for x-component mean velocity computation
$x_{\text{ref,norm}}$ [-]	normalized x-coordinate of the reference position for x-component mean velocity computation
$z_0$ [m]	reference height of roughness coefficient for T&T building site
$z_{\text{min}}$ [m]	minimum reference height of roughness coefficient for T&T building site
$\alpha$ [deg]	gabled roof inclination with respect to the horizontal
$\Delta g_{\text{domain}}$ [m]	grid resolution on outer computational domain

$\Delta g_{\text{building}}$  [m]      grid resolution on the building

## REFERENCES

- [1] Lacey, S., April 26, 2011. *Creating Investor Certainty in Large-Scale Solar*, Inside Renewable Energy podcast.
- [2] September 2010. *Renewables 2010 Global Status Report*, REN21.
- [3] Cosoiu, B., Damian, A., Damian, R. M., Degeratu, M., June 11-13, 2008. *Numerical and Experimental Investigation of Wind Induced Pressures on a Photovoltaic Solar Panel*, 4th IASME/WSEAS International Conference on ENERGY, ENVIRONMENT, ECOSYSTEMS and SUSTAINABLE DEVELOPMENTS (EEESD'08), Algrave, Portugal.
- [4] Krishna, P., Kumar, K., Bhandari, N. M. *IS:875 (Part3): Wind Loads on Buildings and Structures – Proposed Draft & Commentary*, Document No. IITK-GSDMA-Wind02-V5.0 and IITK-GSDMA-Wind04-V3.0.
- [5] Bhandari, N. M., Krishna, P., Kumar, K., Gupta, A. *An Explanatory Handbook on Proposed IS 875 (Part3) – Wind Loads on Buildings and Structures*, Document No. IITK-GSDMA-Wind06-V3.0.
- [6] Jensen, A. G., Franke, J., Hirsch, C., Schatzmann, M., Stathopoulos, T., Wisse, J., Wright, N. G., 2004. *CFD Techniques – Computational Wind Engineering*, Proceedings of the International Conference on Urban Wind Engineering and Building Aerodynamics – Impact of Wind and Storm on City Life and Built Environment – Working Group 2, COST Action C14, Von Karman Institute, Rode-Saint-Genève (Belgium).
- [7] Franke, J., *Wind Effects on Buildings and Design of Wind-Sensitive Structures*, CISM International Centre for Mechanical Sciences, 2007, Vol. 493, pp. 67-103.
- [8] Dalgliesh, W. A., January 1981. *Wind Loads on Low Buildings*, Division of Building Research, National Research Council of Canada, Ottawa.
- [9] Hoxey, R. P., Robertson, A. P., Basara, B., Younis, B. A., *Geometric Parameters that Affect Wind Loads on Low-Rise Buildings: Full-Scale and CFD Experiments*, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 50, Dec. 1993, pp. 243-252.
- [10] Dagnew, A. K., Bitsuamalk, G. T., Merrick, R., *Computational Evaluation of Wind Pressures on Tall Buildings*, 11<sup>th</sup> Americas Conference on Wind Engineering, San Juan, Puerto Rico, June 22-26, 2009.
- [11] Raciti Castelli, M., Castelli, A., Benini, E., July 27-29, 2011. *Modeling Strategy and Numerical Validation of the Turbulent Flow over a two-Dimensional Flat Roof*, ICCFD 2011, Paris (France).
- [12] DM 14/01/2008 – *Norme tecniche per le costruzioni*, issued on the Italian “Gazzetta Ufficiale” on February 4, 2008.

## AUTHORS BIOGRAPHY

**Marco Raciti Castelli** is Research Associate at the Department of Mechanical Engineering of the University of Padua and Fluid Dynamic Specialist at ESPE S.r.l.

**Sergio Toniato** is Executive and Design Manager at ESPE S.r.l.

**Ernesto Benini** is Associate Professor at the Department of Mechanical Engineering of the University of Padua.

# MODELING OF DRYING PROCESS OF CANDIES OBTAINED WITH STARCH MOLDING TECHNIQUE

N. Baldino<sup>(a)</sup>, L. Seta<sup>(b)</sup>, F.R. Lupi<sup>(c)</sup>, D. Gabriele<sup>(d)</sup>, B. de Cindio<sup>(l)</sup>

<sup>(a)</sup>Department of Engineering Modeling, University of Calabria. Via P. Bucci Cubo 39c. 87036 Arcavacata di Rende (CS) Italy

<sup>(a)</sup>noemi.baldino@unical.it, <sup>(b)</sup>lucia.seta@unical.it, <sup>(c)</sup>francesca.lupi@unical.it, <sup>(d)</sup>d.gabriele@unical.it,  
<sup>(e)</sup>bruno.decindio@unical.it

## ABSTRACT

Starch molding is a widely used shape forming technique in the confectionery industry. Candies are deposited into cavities imprinted in dry granular starch. A drying stage follows deposition, which largely determines the productivity of the process and the candy quality.

In the present work a finite elements model was developed that quantitatively predicted the moisture diffusion from a single candy piece into the starch bed or air at a given temperature (50°C).

This simulated a simultaneous bidimensional heat and moisture transfer accounting for a variation in the candy physical properties as functions of local moisture content and temperature.

Model validity was attained for candy drying experiments performed both in the laboratory and monitored in the Silagum srl (Lamezia Terme (CZ)), which has contributed to this work. The model is a powerful tool for analyzing the behavior of an industrial drying process.

Keywords: starch molding, finite elements model, heat and moisture transfer, powerful tool

## 1. INTRODUCTION

Confectionery items such as starch gums and jelly candies have been extensively manufactured using the starch molding technique, which involves a series of operations including deposition of the hot candy syrup into the starch molds and drying by stoving. These operations are very critical in many aspects. The candy suspension (gel) which will be cast into starch molds, must be fluid enough at depositing to get a good shape definition and to prevent tailing. This often requires casting of candy suspension at a moisture content greater than desirable in the final product. For this reason the products need a drying step following depositing, with moisture exchange between the product and both its environment and the starch of the mold. This is a dynamic process involving the candy, molding starch and the air that surrounds them.

The molding starch actually is able to shape the candy and also to adsorb the candy humidity gently; for this reason its moisture content and temperature are very important process parameters. During the period

that the product is in contact with the molding starch, moisture content and temperature are constantly changing. The rate at which starch adsorbs moisture is comparatively slow and the dynamic relationship of the product, the starch and the atmosphere has been scarcely investigated (Sudharsan, Ziegler, and Duda 2004). In economic terms, the rate determines I, productivity of the process and the amount of time to spend in making these products.

It is well known that the understanding of the physical processes occurring during the candy production operations and in particular during the drying process, can lead to control the work parameters involved, improving product quality at a lower cost. Then it clearly appears that, unlike in most foods, water is a minor constituent of candies but one that has a major impact on the production process.

In the candy drying stage simultaneous physical phenomena are involved, heat transfer from the air to the product and mass-moisture transfer from the product to the air. This stage is controlled by many parameters: temperature and relative humidity of the air into the oven; initial temperature and moisture content of molding starch; shape of the candies; distance between candies in the mold.

This research proposes a theoretical model describing the transport phenomena involved in candy drying. The aim of the study was to determine the influence of some of the most important operating variables, namely drying air humidity and temperature, on the performance of the drying process of candies.

The main connotation of this study regarded the possibility of increasing the accuracy of the drying process modeling by the use of the finite elements method (FEM). With respect to most of the works published in the literature, the main innovation introduced in this study is represented by the model formulation. In fact, the latter must consider the simultaneous bidimensional heat and moisture transfer accounting for the variation in the candy physical properties as functions of local moisture content and temperature. The resulting system is an unsteady-state partial differential set of equations that has been solved by means of a FEM commercial software, *Comsol Multiphysics* v.3.5 (Comsol, Sweden).



The validation of the model was carried out by comparing simulated and experimental profiles of candy humidity over time during the drying process.

The experimental study was applied before to both model candies and later to industrial ones, prepared with the same recipe and dried in a laboratory and industrial oven respectively. In both cases the drying process was realized by molding starch technique using a cornstarch bed with fixed dimension and distribution of candies. The latter parameters are those employed by Silagum srl (Lamezia Terme (CZ)), a Calabrian candy industry which has contributed to this work, allowing us to monitor and to study its drying process. Industrial operating conditions, such as temperature (50°C) and humidity of candy, starch and air were reproduced in the laboratory to study model systems.

The main objective was to develop an accurate transport model that could be used to simulate the behavior of a real drying process and to define, over a wide range of drying conditions and of different types of candies, the “optimal” set of operating conditions in each specific situation. In this way, it might be possible to minimize expensive pilot test-runs and to have good indications on the characteristics and the quality of final products.

## 2. THEORICAL BACKGROUND

When a moist food is put in contact with dry and warm air, two different transport mechanisms simultaneously occur: heat transfer from air to the material and transfer of water from food to air. Within the solid material, conductive heat transfer and diffusive water transfer take place.

In this study water and heat transfers are computed through transient mass and energy balances, respectively, whereas the evaporation occurring at air-candy interface was considered by defining proper boundary conditions expressed in terms of heat and mass transfer coefficients estimated by well-known empirical correlations (Perry and Green 1984).

Ellipsoidal geometry was assumed for the candies immersed at the same distance in a starch dry bed. Then, a spherical coordinate system was considered including candy and starch because both the air and the mold starch are able to adsorb product humidity (Figure 1). Dimensions of candies and of starch bed were established with reference to the industrial drying process of candies (Silagum srl, Lamezia Terme (CZ)).

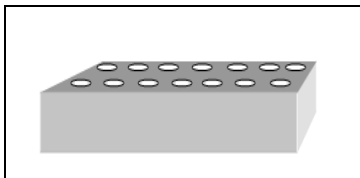


Figure 1: Sketch of mold starch with candies.

The model is based on the following hypothesis:

1. Heat transfer in the product was by conduction.
2. Mass transfer in the product was by diffusion.

3. Transport of water vapor within the dehydrated candy was negligible.
4. No food shrinkage occurred during drying.
5. Azimuth symmetry.
6. Candy and starch densities were constant (Sudharsan, Ziegler, and Duda 2004).
7. Candy and starch heat capacities were constant (Cabras and Martelli 2004).
8. Initial humidity and temperature of candy and starch were uniform within the materials.
9. Candy diffusion coefficients were function of moisture content within the materials.
10. Starch diffusion coefficient was constant in the characteristic humidity range (6-12%) (Sudharsan, Ziegler, and Duda 2004).
11. Immediate achievement of liquid-vapor equilibrium conditions at the interfaces air/candy and air/starch.
12. The drying air was supplied continuously to the product and its flow was parallel to its surface.

### 2.1. Mathematical model

The energy balance in the solids, candy and starch, based on Fourier's law, leads to:

$$\rho C_p \frac{\partial T}{\partial t} = \nabla \cdot (k(T, C) \nabla T) \quad (1)$$

where  $\rho$  is density (kg/m<sup>3</sup>),  $c_p$  is heat capacity (J/kg K),  $T$  is the temperature (K),  $t$  is time (s), and  $k$  is thermal conductivity (W/m K).

Mass balance, based on Fick's law, leads to

$$\frac{\partial C}{\partial t} = \nabla C \cdot (D \nabla C) \quad (2)$$

where  $C$  is water concentration (mol/m<sup>3</sup>),  $D$  is the effective diffusion coefficient of water (m<sup>2</sup>/s).

The drying model was applied to a central candy in the mold starch for which a symmetrical approach can be applied. Thanks to this approach two domains were considered: the starch and the candy.

Since the considered system is composed of two domains with different physical properties, Eqs. (1) and (2) modeling the drying process must be applied to each one by considering their specific properties.

Supposing that the material parameters  $c_p$ ,  $k$  and  $D$ , are determined by local water concentration and on food temperature, Eqs. (1) and (2) form a system of nonlinear partial differential equations. Initial conditions are straightforward since it is assumed that, at  $t=0$ , candy and starch concentrations and their temperatures are equal to their initial values, respectively.

Boundary conditions were applied to the border (B) of the system sketched in figure 2.

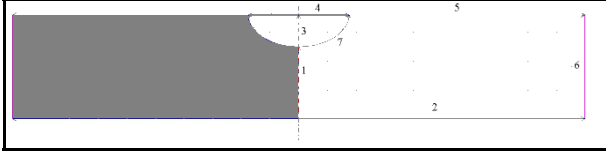


Figure 2: Simulation domain and boundary conditions for the 2-D FEM analysis.

The boundary conditions were applied to the surfaces of the two different domains identified by their normal  $\underline{n}$ . Room temperature and mass adiabatic conditions were assumed for the starch box (Boundary 2). The physical parameters of candy and starch are indicated with subscripts  $c$  and  $s$  respectively.

The boundary conditions relative to eq. (1) were:

- B-1 and B3:  $\frac{\partial T}{\partial r} = 0$  (3)

- B-2:  $T = T_\infty$  (4)

- B-4:  $-\underline{n} \cdot (-k_c \underline{\nabla} T) - \lambda N_{wc} = h(T_\infty - T_c)$  (5)

- B-5:  $-\underline{n} \cdot (-k_s \underline{\nabla} T) - \lambda N_{ws} = h(T_\infty - T_s)$  (6)

- B-6:  $-\underline{n} \cdot (-k_s \underline{\nabla} T) = 0$  (7)

- B-7:  $-\underline{n} \cdot (-k_c \underline{\nabla} T) = -\underline{n} \cdot (-k_s \underline{\nabla} T)$  (8)

where  $k$  is water latent heat of vaporization (J/mol),  $N_w$  is water diffusive flux on food surface (mol/m<sup>2</sup> s),  $h$  is heat transfer coefficient (W/m<sup>2</sup> K),  $T_\infty$  is gas temperature in the bulk,  $T_s$  and  $T_c$  are starch and candy surface temperatures respectively (all temperatures are measured as °K).

The boundary conditions relative to eq. (2) were:

- B-1 and B3:  $\frac{\partial C}{\partial r} = 0$  (9)

- B-2:  $\underline{n} \cdot (D_s \underline{\nabla} C) = 0$  (10)

- B-4:  $-\underline{n} \cdot (-D_c \underline{\nabla} C) = K_c(y - y_\infty)$  (11)

- B-5:  $-\underline{n} \cdot (-D_s \underline{\nabla} C) = K_c(y - y_\infty)$  (12)

- B-6:  $\underline{n} \cdot (D_s \underline{\nabla} C) = 0$  (13)

- B-7:  $-\underline{n} \cdot (-D_c \underline{\nabla} C) = -\underline{n} \cdot (-D_s \underline{\nabla} C)$  (14)

where  $K_c$  is mass transfer coefficient (mol/m<sup>2</sup>s),  $y_\infty$  is water bulk molar fraction in the air,  $y_s$  and  $y_c$  are water molar fraction evaluated in gaseous phase at the starch/air and candy/air interfaces respectively

The interfaces candy/air and starch/air were considered to be at the thermodynamic equilibrium. Thanks to this hypothesis it is not necessary to apply the thermodynamic equilibrium at the B-7 surface too. The equality of both temperature and water activity of starch and candies results a consequence of previous boundary conditions.

At the drying air material surface (i.e. B4 and B5) it is possible to calculate  $y_s$  and  $y_c$  values, from the thermodynamic equilibrium linking between water concentration in the gas phase and water concentration

on candy/starch surfaces. If it is assumed that the gas phase behaves as a mixture of ideal gases, the following equation applies:

$$Py = P^0 a_w \quad (15)$$

where  $P$  is air pressure (Pa),  $P^0$  vapor pressure (Pa),  $y$  the molar water fraction in the gas phase, and  $a_w$  is the activity of water of starch or candy respectively.  $P^0$  was easily calculated with the equation 16 (Perry 2007).

$$P^0 = \exp\left(73.649 - \frac{72582}{T} - 7.3037 \cdot \log(T) + 4.1653 \cdot 10^{-6} T^2\right) \quad (16)$$

To compute the equilibrium between drying air and both starch and candies, the same nonideal mixture behavior was assumed by using GAB isotherm that is valid in the 0.1-0.9 activity of water range and links water content to the activity of both materials considered:

$$\frac{U}{U_0} = \frac{CKa_w}{(1 - Ka_w)(1 - Ka_w + CKa_w)} \quad (17)$$

Where  $U$  is the moisture content in dried material (dry basis- Kg<sub>water</sub>/Kg<sub>dry solid</sub>),  $U_0$ ,  $C$  and  $K$  are the GAB constants characteristic of dried material. The latter parameters were calculated for candies and mold starch by using their desorption isotherms.

The isotherms were obtained by measuring the moisture content in dried materials with analyzer HB43 (Mettler Toledo, Germany) and the activity water with instrument Sprint TH500 (Novasina, Italy). Desorption isotherms data were fitted by GAB equation by using the software Table Curve 2D.

Fitting results for two types of candies and for mold starch studied were reported in Table 1.

Table 1: GAB constants

Sample	$U_0$	$C$	$K$
Candy <sub>lab</sub>	34.60	0.12	1.04
Candy <sub>ind</sub>	34.47	0.22	0.93
Mold Starch	12.56	34.53	0.62

Eqs. (3)–(15) were essential to describe properly the complex steps involved during the drying process. Eq. (5), in fact, was expressed in terms of water activity  $a_w$ , which is a distinctive parameter of each product being characterized by its own structure. Water activity, then, determines the strength of the bonds between food structure and water. Moreover,  $U$  and  $T$  through Eq. (15), determine the value of the food water molar fraction, and then the time evolution of the drying process. In particular, if  $U$  and  $T$  are high enough to have  $y > y_\infty$ , food drying starts immediately (Eqs. (11) (12)); heat is transferred by convection from air to food

whereas latent heat of vaporization is transported from the food to the water thus allowing its evaporation (Eqs. (5) (6)).

The physical properties  $\rho$  and  $c_p$  of candies and mold starch used in the model with reference Sudharsan, Ziegler, and Duda 2004, Cabras and Martelli 2004 respectively, are reported in Table 2 .

Table 1: Physical properties of candy and mold starch

Sample	$\rho$ (Kg/m <sup>3</sup> )	$c_p$ (J/Kg K)
Candy	748	2300
Mold Starch	700.2	1300

Thermal conductivity values of candies and starch as a function of  $U$  and  $T$ , are defined by Eq. (18) (Saravacos and Maroulis 2001):

$$K(U,T) = \frac{1}{1+U} \lambda_0 \exp \left[ -\frac{E_0}{R} \left( \frac{1}{T} - \frac{1}{T_{ref}} \right) \right] + \frac{1}{1+U} \lambda_i \exp \left[ -\frac{E_i}{R} \left( \frac{1}{T} - \frac{1}{T_{ref}} \right) \right] \quad (18)$$

Where  $\lambda_i$ ,  $\lambda_0$ ,  $E_i$  and  $E_0$  (W/m K) are parameters characteristic of candy and starch and  $T_{ref}$  is a reference temperature. The values of them used in the model are that reported in Saravacos and Maroulis 2001.

The diffusivity coefficient of water in starch can be considered constant in the operating range of humidity and temperature, and a value of  $10^{-9}$  m<sup>2</sup>/s was used in the model (Vagenas and Karathanos 1991, Sudharsan, Ziegler, and Duda 2004).

A dependence of diffusivity coefficient of water in the candies on the  $U$  and  $T$  values (equations not reported) was obtained experimentally measuring  $D$  coefficient of candies having different humidity by using NMR tests (Bruker, USA). These measurements were carried out at three different temperatures (50°C, 60°C, 70°C) on the candies dried in the laboratory oven (Termostabil K3), and on those produced industrially taken from Silagum oven. The trend  $D$  vs  $U$  registered for the latter was reported in figure 3.

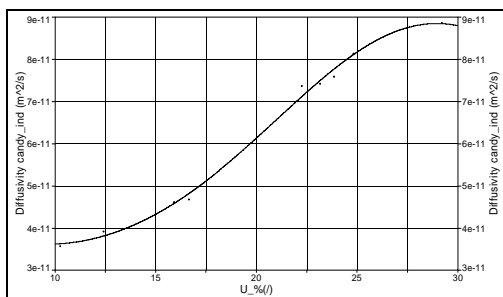


Figure 3:  $D$  vs  $U$  for industrial candies

Heat transfer coefficient  $h$  was determined as the ratio between the heat flux and the temperature difference among the oven environment and the product surface measured during the drying stage by flux sensor.

Mass transfer coefficients  $K_c$  at candy/air and starch/air were calculated with the Chilton-Colburn analogy, for flat slab and air tangential flow (Bird, Stewart, Lightfoot 1960).

### 2.1.1. Numerical solution

The above system of nonlinear partial differential equation has been solved by the finite elements method developed by the commercial package called *Comsol Multiphysics* v.3.5 (Comsol, Sweden).

The control volume concerned was discretised into 23937 and 16888 triangular finite elements for the laboratory system (figure 4) and industrial one respectively, with a thicker mesh close to the candy/starch and candy/air interfaces.

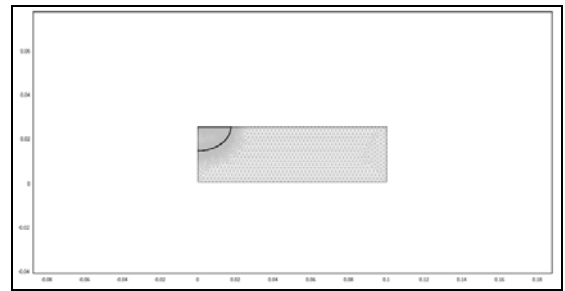


Figure 4: FEM mesh used to model laboratory candies

On 1.67 GHz and Intel Core 2 Duo PC running under Windows 2000, candy drying behavior over a time horizon of 48 h was simulated, on average, in about 15 min, using the time-dependent iterative nonlinear solver already implemented in the Chemical Engineering Module of the software.

It should be remarked that the proposed model does not need any parameters adjustment, but only the specification of a set of input variables that can be varied within a specific range of physical significance to simulate food drying behavior at different operating conditions. The above was tested simulating the humidity profile in the candy over time, varying air temperature in the range 18-50°C, at constant values of air relative humidity and starch humidity. Water activity profile at the candy/air and candy/starch interfaces were also obtained at the same operating conditions.

The water content concentration profile inside the candy was finally obtained from simulation, varying the relative humidity of the drying air in the range 0-100% and maintaining constant values of both air temperature and starch humidity.

## 3. RESULTS AND DISCUSSION

The simulation results expressed as the time evolution of candy average moisture content ( $U_b$  expressed on wet basis- Kg<sub>water</sub>/Kg<sub>wet solid</sub>), calculated as bulk integral, for the laboratory products were presented in figure 5. The

candy humidity profile was obtained as a function of air relative humidity ( $U_r\%$ ) at the air temperature  $T_\infty=50^\circ\text{C}$  and starch water activity  $a_{ws}=0.115$ . Both temperature and moisture profiles within the candy were also calculated as a function of drying time but were not here reported.

Figure 5 shows the effect of air humidity on the drying process. As expected, with increasing its value, the final candy average moisture content results higher due to a substantial reduction of the concentration gradient, which is the driving force of moisture transfer from the candy to the air.

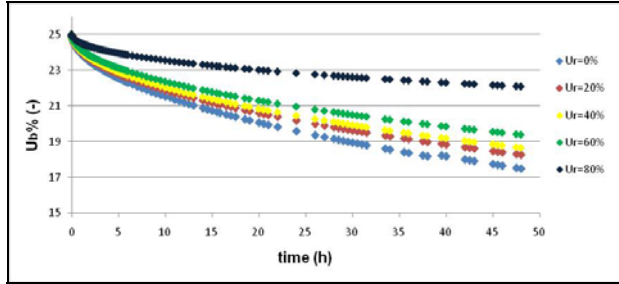


Figure 5: time evolution of laboratory candy  $U_b$  at different air relative humidity ( $U_r\%$ ) ( $T_\infty=50^\circ\text{C}, a_{ws}=0.115$ )

To confirm the above the time evolution of laboratory candy water activity ( $a_{wc}$ ) at the candy/air and candy/starch interfaces were also presented in the figures 6 and 7 respectively.

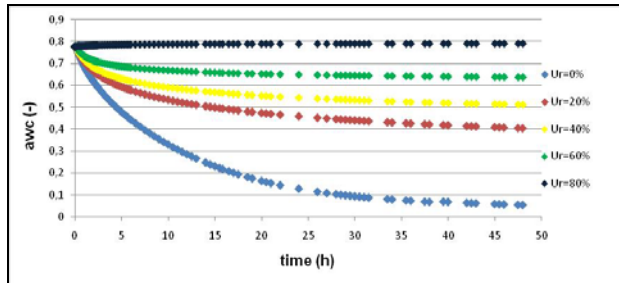


Figure 6: time evolution of laboratory candy  $a_{wc}$  at the candy/air interface at different air relative humidity ( $U_r\%$ ) ( $T_\infty=50^\circ\text{C}, a_{ws}=0.115$ )

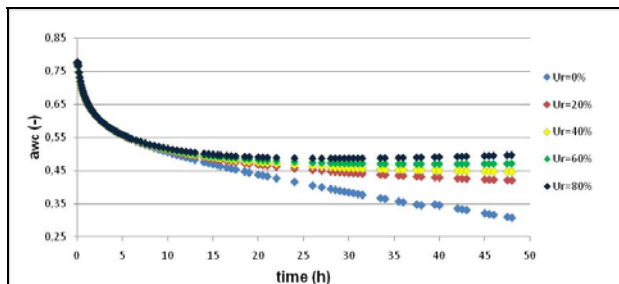


Figure 7: time evolution of laboratory candy  $a_{wc}$  at the candy/starch interface at different air relative humidity ( $U_r\%$ ) ( $T_\infty=50^\circ\text{C}, a_{ws}=0.115$ )

At both the interfaces the  $a_{wc}$  value after 48 hours of drying decreases with reducing air humidity, and

these effects, as expected, result more significant at the candy/air than the candy/starch interface, confirming more and more what was said before.

Moreover, it is important to note that the air humidity also indirectly affects the concentration gradient at the candy/starch interface, confirming the findings of Sudharsan, Ziegler, and Duda 2004.

As expected air relative humidity proved to be the most important operating parameter among those investigated, with which it is possible to modify the candy drying stage. Actually simulations carried out at different air temperature and starch water activity values showed little and negligible effects on the candy humidity profile.

#### 4. MODEL VALIDATION

Candy drying experiments were performed in the laboratory and some data were monitored in industry in order to be compared with the model theoretical predictions so as to verify their validity. Model predictions were obtained respecting the geometrical and operating conditions characteristic of both laboratory and industrial drying experiments.

Model validity was attained in the both cases investigated. Figure 8 shows the comparison between the experimental results and the model predictions for industrial candies only, obtained without any parameter adjustment.

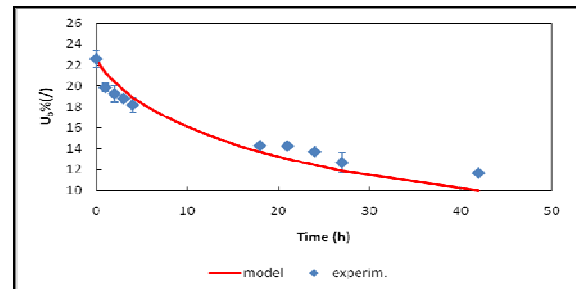


Figure 7: comparison between model prediction and experimental data evolution of industrial candies moisture content

A remarkable agreement is observed especially during the initial drying hours when experimental data and theoretical predictions overlap and relative errors never exceed 5%. Later on, a deviation can be observed, which can be ascribed to the lack of accounting for a formation of candy dry skin in the model. This sort of crust begins to appear typically in the final stage of the candy drying process from a region of low moisture content near the surface and is an additional resistance of moisture transport from the product to the air (Sudharsan, Ziegler, and Duda 2004). For this reason, probably, the model predicted a candy humidity reduction higher than the real one, just over last hours of drying process.

## 5. CONCLUSION

A theoretical analysis of the drying process of candies, produced with the starch molding technique, was presented. With respect to most works published in the literature, the main innovation introduced by this study lies in the model formulation. This, in fact, simulated the simultaneous bidimensional heat and moisture transfer accounting for the variation of both air and candy physical properties as functions of local values of temperature and moisture content. The numerical solution of unsteady-state heat and mass balance equations, performed by the finite elements method gave a good agreement between experimental results and model predictions, mostly during the initial drying hours. When candy dry skin effects occur a deviation between theoretical predictions and experimental results can be observed. The model that contains no adjustable parameter, is a powerful tool for analyzing the behavior of industrial drying ovens, as it could be used, over a wide range of drying conditions and of different types of candies, to individuate the “optimal” set of operating conditions in each specific situation. In this way, it might be possible to minimize expensive pilot test-runs and to have good indications on the characteristics and the quality of the final products.

The proposed model is general since it is capable of describing both the transport of free and bounded water within the food and the evaporation/condensation phenomena that, as a function of the actual driving forces, may occur at the air/candy and starch/candy interfaces. It is very flexible and can be easily adapted to different geometries with no *a priori* limitation for irregular or complex-shaped systems. Work is, however, in progress for improvement of the model features, in particular by seeking an independent and more reliable determination of heat and mass transfer coefficients and accounting for candy dry skin during the drying process.

## ACKNOWLEDGEMENT

The University of Calabria and the Food Science and Engineering Interdepartmental Center of University of Calabria and L.I.P.A.C., Calabrian Laboratory of Food Process Engineering (Regione Calabria APQ-Ricerca Scientifica e Innovazione Tecnologica I atto integrativo, Azione 2 laboratori pubblici di ricerca mission oriented interfiliere) are gratefully acknowledged.

The factory Silagum srl (Lamezia Terme (CZ) Italy), is also acknowledged for its kind support in the experiments.

## REFERENCES

Perry, R. H., and Green, D. (1984). *Perry's Chemical Engineers' Handbook*. New York, USA: McGraw-Hill.

Sudharsan, M.B., Ziegler, G.R., Duda, J.L. (2004) Modelling diffusion of moisture during stoving of starch-molded confections. *Institution of Chemical Engineers, TransICHEME, Part C, Food and Bioprocesses Processing*, 82 (C1), 60-72

Cabras, P., Martelli A. (2004). *Chimica degli alimenti* Piccin-Nuova Libreria, pp287-288

Bird, R.B., Steward, W.E., Lightfoot, E.N. (1960). *Transport Phenomena*. John Wiley & Sons, London, UK.

# BENCHMARKING REAL-WORLD JOB-SHOP SCHEDULING USING A GENETIC ALGORITHM WITH A SIMULATION APPROACH

Dr.-Ing. Peter Steininger

Steinbuch Centre for Computing (SCC)  
Karlsruhe Institute of Technology (KIT)

[steininger@kit.edu](mailto:steininger@kit.edu)

## ABSTRACT

Although production scheduling has attracted the research interest of production economics communities for decades, a gap still remains between academic examples and real-world problems. Genetic Algorithms (GA) constitute techniques which have already been applied to a variety of combinatorial problems. I intend to explain the application of a GA approach to bridge this gap for job-shop scheduling problems, by minimizing the makespan of a production program or increasing the due-date reliability of jobs.

Simulation is a useful tool in problem solving. Here repetitive runs of simulated models or computed solutions through algorithms are applied. For job-shop and resource-constrained project scheduling, problems trying to bridge the gap between computed solutions and the feasibility of the simulation occur. I would like to explain the application of this special GA for job-shop and resource-constrained project scheduling. Possible goals for scheduling problems include minimizing the time required of a production program, or increasing the due-date reliability of jobs, or possibly other objectives which can be described in a mathematical expression. The approach focuses on integrating a GA into a commercial software product, namely Microsoft Project 2003, and verifying the results with the simulation.

Keywords: Job-shop scheduling, Genetic algorithms, Job-shop Benchmarks, Real-world scheduling problems, Simulation.

## 1. INTRODUCTION

### 1.1. Characteristics of job-shop scheduling problems

Many jobs in industry and elsewhere require a collection of tasks or activities to be completed whilst satisfying temporal, resource and precedence constraints. Temporal constraints refer to the time requirement that some tasks or set of tasks have, where they must be finished before or after a certain point in time. Resource constraints dictate that two tasks requiring the same resource cannot be carried out simultaneously. Whereas Precedence constraints refers to the technologically imposed carrying out of the tasks within a job or production order. The objective is to create a schedule specifying when each task or activity is to commence (or finish) and what resources it shall use in order to

satisfy all the constraints while pursuing an objective. The overall goal is to complete all tasks to an acceptable standard within the least time possible (makespan), whilst taking into consideration such aspects as minimizing the mean tardiness and number of jobs. This is also referred to as the job-shop scheduling problem (JSP).

The JSP is a very important and well defined scheduling problem. It is a basic model, which may be extended through use of additional characteristics like buffers, transportation, setup time, time lags, etc., allowing practical scheduling problems to be modeled more precisely. In its general form, it is NP-complete, meaning that there is probably no efficient procedure for exactly finding the shortest schedule for arbitrary instances within this class of problem.

BAGCHI (p. 109) references the JSP as follows: "Within the great variety of production scheduling problems that exist, the job shop scheduling problem is one that has generated the largest number of studies. It has also earned a reputation for being notoriously difficult to solve. Nevertheless, the JSP illustrates at least some of the demands imposed by a wide array of real world scheduling problems... Attempts to tackle the multi objective job shop are still relatively few."

A JSP is usually solved using a heuristic algorithm which takes advantage of special properties of a specific class of instances. This can be regarded as a loophole to reduce the complexity of a given problem.

### 1.2. Formal problem description

An instance of the JSP consists of a set of  $NOA_i$  activities within jobs  $i$  and  $NOM$  machines  $j$ . Each job consists of a number of activities so that we may count the total number of activities  $NOA$  as follows:

$$NOA = \sum_{i=1}^{NOM} NOJ_i \quad (1)$$

Here each job has a fixed number and a sequence of activities. Each activity requires a certain amount of time implementing a single machine for its entire duration. An activity must be finished before each activity following it can commence, with each job utilizing a different machine. Two activities cannot be scheduled at the same time if they both require the same machine. In all we need to find a feasible schedule



which minimizes some objective function, whilst reducing makespan. Where makespan expresses the overall completion time of all activities, see STEININGER (2007, pp. 26). This produces a complexity function for the JSP expressed as

$$O((NOJ!)^{NOM}) \quad (2)$$

In order to find the best schedule for a problem instance, we could enumerate and evaluate all possible schedules. The number of feasible schedules to be enumerated would be the result of function (2).

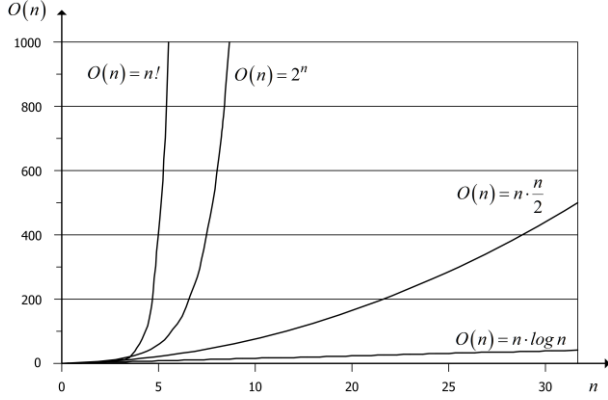


Figure 1. Selected complexity functions (STEININGER, p. 82).

Figure 1 illustrates the dimensions of selected complexity functions, where  $n$  represents the number of problem elements, which are determined here by the number of activities and machines. These graphs illustrate how the complexity of a JSP can be much larger than some other well-known problems, such as "Permutation problem" which is  $O(n) = n!$ , "Towers of Hanoi" which is  $O(n) = 2^n$ , and "Quicksort" which is  $O(n) = n \cdot \log n$ .

### 1.3. Classification of scheduling problems

Classes of scheduling problems can be specified in terms of the three-field classification approach initially introduced by CONWAY, MAXWELL and MILLER (2003) and extended by GRAHAM (1979) and BRUCKER (2004). Of course this depicts a continuous developmental process open to scrutiny and reevaluation. This three-field classification is described as  $\alpha | \beta | \gamma$ , where  $\alpha$  specifies the machine environment,  $\beta$  specifies the job characteristics and  $\gamma$  represents either the objective function or a combination of objective functions. Using the three-field classification to specify the problem instance of the JSP we are examining, the following taxonomy is noted:

$$J, NOM | NOJ, intree, t_{ij} | C_{max} \quad (3)$$

Formula (3) describes a class of scheduling problems as JSP ( $J$ ) with a fixed machine count of  $NOM$  and a predefined and fixed number of  $NOA$  activities. The precedence routing of activities in each order is predefined and fixed as a directed graph with operation times

( $t_{ij}$ ). These are expressed as integer values for each activity.

The three-field classification denotes  $\gamma$  as the objective function or a combination of objective functions. In formula (3)  $\gamma$  specifies the "traditional" objective function ( $C_{max}$ ). This depicts our primary objective which is to limit the makespan using the least time possible for the schedule of all  $NOA$  activities using  $NOM$  machines.

## 2. MODELLING OF JSP SCHEDULING PROBLEMS

### 2.1. Formal problem representation

Even slightly different job-shop problems require completely different encodings in order to find a good solution. Thus, choosing an efficient representation is a very important component when solving a JSP. However, deciding upon a relevant representation for a scheduling problem is as difficult as choosing a good search algorithm for a decision problem. Not all algorithms work equally efficient in a specific problem representation. In order to describe the representation technique developed for our solution, a simple job-shop scheduling problem example has been used in Table 1.

Table 1. Example of a production schedule problem with 3 jobs, routing information  $S_j$  for the jobs  $i$ , 3 machines  $j$  and operation times  $t_{ij}$  for each task of a job in time units (TU).

Job $j$	$S_j$	$t_{ij}$ [TU]		
		Machine 1	2	3
1	(3,2,1)	3	5	1
2	(1,2,3)	3	2	1
3	(2,1,3)	1	2	5

The scheduling problem can be represented by a graph as shown in Figure 2. In addition to the activity nodes ( $i, j$ ), it contains a source node  $a$ , and a sink node  $e$ , both with no durations (operation times), and two dotted nodes called  $r_2$  and  $r_3$  which describe a later start of jobs 2 and 3 imposed relative to job 1.

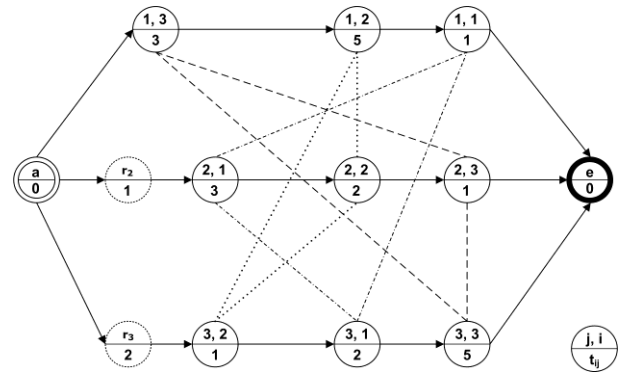


Figure 2. Network representation of a JSP based on Table 1 (STEININGER, p. 64).



The directed arcs running from the source node  $a$ , through each activity node  $(i, j)$  to the sink node  $e$  describes the technological sequence of activities based on the routing  $S_j$  in Table 1. Each node shows the job number  $i$ , the machine needed  $j$  and the operation time  $t_{ij}$ . There are also undirected arcs in the network, which reference all possible sequences of an activity in a given job on a specific machine. Such a representation is termed a disjunctive network.

## 2.2. Data representation and problem reduction

Care must be taken when adopting such a graphical representation into a data structure, especially for the JSP in an area with hundreds of activities, thousands of sequences and millions of possible actions carried out with specific machines.

A data structure which is very efficient in the use of storage (due to the size of a practical problem) as well as in time, can also be considered as depicting a network. GALLO and PALLOTTINO (1982) introduced the so-called "Forward Star" data structure, which is the most efficient portrayal of all existing network data structures for representing a network. The "Forward Star" data structure uses three arrays to describe a (directed) network. First we have an array termed *from*, whose index represents all nodes of the network. The value of an index field references the index of the second array named *succ*, which is the index of a node to connect, referencing *from*. The third array is labeled *cost* and reports the cost of a specific arc connecting two nodes.

The "Forward Star" data structure allows for a perfect implementation of the activity order of any JSP. An efficient implementation in storage and time and a reduction of the initial problem described by the  $\alpha | \beta | \gamma$  three-field classification is achieved/possible/occurs. Using the "Forward Star" data structure our problem is reduced to the following taxonomy (4):

$$J, NOM | NOJ, chains, t_{ij} | C_{max} \quad (4)$$

Care must also be taken when adopting the representational scheme and the associated operators for an effective algorithm. Here with the application of traditional problem solving with GAs, the chromosomes are implemented as binary vectors. Such an algorithm is an excellent choice for problems in which a point naturally maps into a chromosome of zeros and ones. Unfortunately, this approach of zeros and ones cannot be implemented for real-world engineering problems such as JSP, because of the amount of information needed to represent coding of the JSP. Therefore, we have to find a way to integrate the "Forward Star" data structure into a GA.

## 3. GENETIC ALGORITHMS

### 3.1. General principle

The term Genetic Algorithm describes a set of methods, which can be used to optimize complex problems. As the name suggests, the processes employed by GAs are

inspired by natural selection and genetic variation. Thus GA uses a variety of possible solutions to a problem and applies repetition in order to modify them. These iterations mimic those in nature in such a way that subsequent populations are fitter and more adapted to their environment compared to their predecessors. As generations progress over time, they become better suited to their environment and provide an advantageous solution in a given time.

Since their development in the late 1980's GAs have been used to find solutions to many types of problems. A unique characteristic of a GA is that the fundamental algorithm is unaware of the problem it is optimizing. All that is required is that the parameters entered into a model can be efficiently transformed to and from a suitable GA chromosome format. Therefore GA optimization can be applied to many types of complex problems.

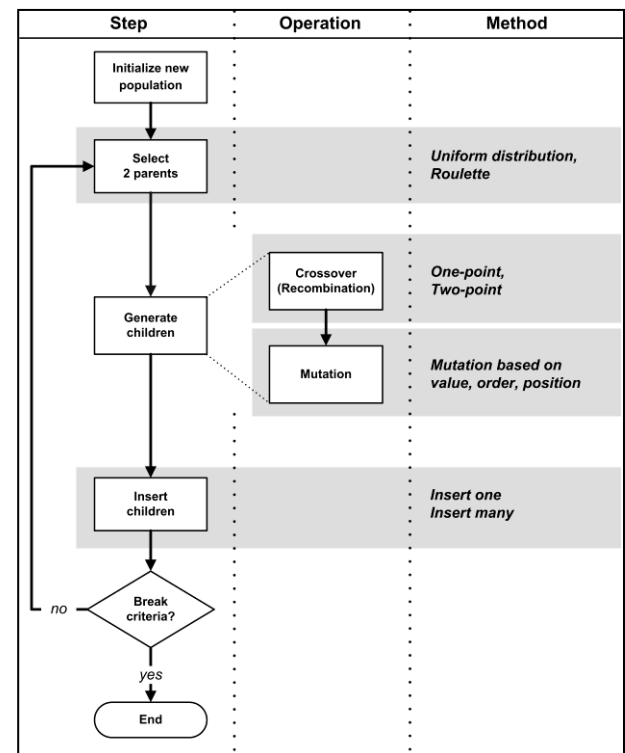


Figure 3. Principal flow of a Genetic Algorithm (following GOLDBERG (1989, pp.59)).

The flow-chart stages of the GA are as follows: First, an initial population of randomly generated sequences of the activities in the schedule is created. These individual schedules form chromosomes which are subject to evolution. Once an initial population has been formed, "selection", "crossover" and "mutation" operations are performed repeatedly until the fittest member of the evolving population converges to a near-optimal fitness value. Alternatively, the GA may run for a user-defined number of iterations.

The size of the population is user-defined and the fitness of each individual, in this case a schedule, is calculated according to a fitness function. In our case this is the makespan or an additive combination of different

goals. It is also possible to use a fitness function on other calculated values like mean tardiness, maximum tardiness, number of tardy jobs and so on. Combinations of different functions in one fitness function are also possible. The schedules are then ranked according to the value of their fitness function and, after that, selected for reproduction.

### **3.2. Schedule encoding and decoding**

GAs were derived from examining biological systems. In biological systems evolution takes place on chromosomes which are organic devices for programming the structure of living beings. In this sense, a living being is a decoded structure of all chromosomes. Natural selection is the link between chromosomes and the performance of the decoded structure. When implementing the GA, the variables that characterize an individual are represented in arrays (by index ordered lists). Each variable corresponds to a gene and the array is equivalent to a chromosome in a biological system.

We have decided to use the encoding schema introduced by BEAN (1994) to build the chromosomes termed "Operation Based Representation". Encoding commences with the enumeration of jobs and corresponding activities in a list. Each activity in a job is encoded with the numerical id of the job in which it resides. All jobs and activities are encoded following that description in one potential schedule for the problem. The result is a chromosome which represents a potential schedule.

The GA requires a few additional operators to function. These operators are methods necessary when working with encoded information:

### **3.3. Genetic Operators**

#### **3.3.1. Crossover**

The GA uses crossover, where mating chromosomes are cut. Crossover is the most delicate operation of GA because it may produce unwanted irregular activity sequences within a job. A corrected 2-point crossover to avoid non-regular activity sequences of orders was used, which GOLDBERG (1989) refers to as a PMX-crossover operator (e.g. STEININGER 2007, p. 146 f.).

#### **3.3.2. Mutation**

Mutation describes the process of randomly changing values of a gene resulting in a variant form. This occurs with small probability. It is therefore not a primary operator, but it ensures that the possibility of searching any section in the problem space is never zero and prevents complete loss of genetic material through reproduction and crossover. We execute the mutation operator as a permutation by first picking (and deleting) a gene before reinserting it at a randomly chosen position of the permutation.

#### **3.3.3. Fitness**

The fitness function is used to evaluate the fitness of each individual in the population and depends on the specific application. Generally a GA proceeds towards creating healthier individuals. If the fitness value is the only information available to the GA, the performance of the algorithm will be highly sensitive to the fitness

function. Therefore when creating streamlined routines, fitness is the value of the objective function to be optimized.

#### **3.3.4. Selection**

To selectively reproduce the population and to determine the next generation we use a hit and miss selection procedure based on the fitness function. This could be implemented using a roulette wheel method. An imaginary roulette wheel is constructed with a segment for each individual in the population. An individual's section size is based on their particular fitness value, with a fitter individual occupying a larger slice of the roulette wheel than a weaker one. Selection is made by rotating the roulette wheel a number of times equal to the population size. When the roulette wheel stops, the individual it points to is selected. In all fitter individuals will have a propensity to be selected more frequently than weaker ones.

### **3.4. Genetic Algorithm Working Set Parameters**

The GA needs a few additional parameters to work. These parameters specify the size of the population, the use of operators and so fourth.

#### **3.4.1. Population size**

The population size depends on the nature of the problem. Typically, it contains several hundreds or thousands of possible solutions. The population is generated randomly, making it possible to cover the entire range of possible solutions. Here a population size of 500 individuals will represent 500 feasible schedules.

#### **3.4.2. Probability of crossover**

The parameter probability of crossover affects the rate at which the crossover operator is applied. A higher crossover rate introduces new chromosomes more quickly into the population. If the crossover rate is too high, good individuals are eliminated faster than selection can produce improvements. A low crossover rate may cause stagnation due to the lower exploration rate.

#### **3.4.3. Probability of mutation**

Probability of mutation describes the likelihood that every gene of each individual in the new population will undergo a random change after a selection process. A low mutation rate helps to prevent any gene positions from getting stuck to single values, whereas a high mutation rate results in essentially a random search.

### **3.5. Final result**

It is a characteristic of the GA that once fairly good solutions have been found their features will be carried forward into even better results, which will ultimately lead to a near-optimal solution. Therefore, GAs are particularly attractive for scheduling.

Compared with other optimization methods, GAs are suitable for traversing large search spaces since they can do this relatively rapidly and because the mutation operator diverts the method away from low scale optima. Being suitable for large search spaces is a useful advantage when dealing with schedules of increasing size since the solution space will grow very rapidly. It is important that this large search space is scanned as fast as possible to enable the practical and useful implementation of schedule optimization.

#### 4. SIMULATION STUDY AND COMPARISON OF RESULTS

The simulation study was implemented according to the organization-oriented simulation method *FEMOS* (German acronym for production and assembly simulator) developed at the Institute of Human and Industrial Engineering of the University of Karlsruhe (Germany). *FEMOS* was used to analyze the entire production-logistical process chain (cf. ZÜLCH, GREINKE 2004; ZÜLCH, WARRISCH 2004, ZÜLCH 2008).

##### 4.1. Test structure

The objective was to detect and define the relationships arising between the logistical sequence strategies and chosen logistical target parameters. In this analysis the results were compared to those of *REIMOS*. The test structure was based on empirical values which have proved viable during previous practical projects. The influencing factors, which were to change in the course of the simulation study, included the planned starting date (point of entry into the system), the planned finish date, the sequence of initialization of production orders or orders of a product group, respectively, plus different priority rules. The priority rules comprised a further selection of rules such as: first come first served, shortest operation time, longest operation time, slack time, and the earliest due date.

Furthermore, eight policies based on varying order initialization were examined. Policy S1 described the sequence of order initialization as specified by the industrial partner. S2 choose initialization according to product groups with increasing work content (the later the point of inclusion, led to the higher the work content). Here individual goods from the product group were combined equidistantly. S3 utilized the incorporation of individual products in equidistant intervals and the product groups were mixed. For S4 distinct incorporations in equidistant intervals occurred. These were arranged according to decreasing product groups with declining work content (the later the point of incorporation, led to the lower the work content). Individual products from the group were incorporated equidistantly. S5 described product groups with increasing work content (the later the point of incorporation led to the higher the work content). They were added one after the other whilst the individual products of the group are were combined at the same point in time. In S6 the order initialization of product groups occurred in equal intervals and the product groups were mixed. S7 looked at product groups with decreasing work content (the later the point of incorporation, led to the lower the work content). They were combined one after the other, with the individual products of a group incorporated at the same point in time. Finally S8 described block initialization, where all products were incorporated at point zero.

According to this test plan, individual order initialization policies were simulated first, followed by the simulation of order initialization policies in combination with priority rules. The second test plan used the mean value analysis to identify those resources consti-

tuting bottleneck resources due to their high capacity utilization. They were then examined by means of several specific order initialization policies by choosing processing- and arrival time-based priority rules for them, as opposed to the date-based priority rules used for all other resources. In the third test plan, the most important operations of the conclusive assembly processes were given higher priorities than the preceding assembly operations. This meant that the operations rated with high priority were those at the end of the production process which would – according to the hypothesis – prove most effective with regards to the reduction of makespan and an increase in reliability of delivery.

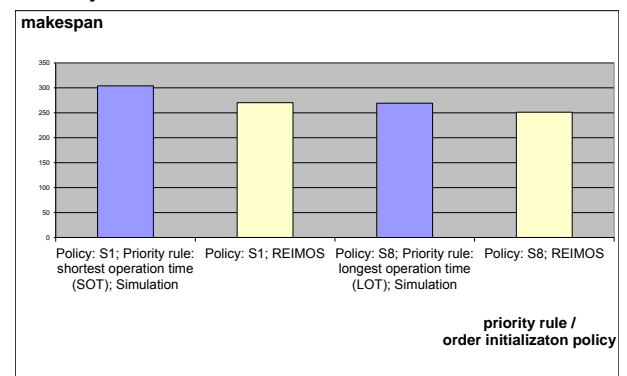


Figure 4. Comparison of makespans identified by means of *FEMOS* using order initialization policies and priority rules and those generated with the *REIMOS* planning method.

##### 4.2. Results and comparison of results with *REIMOS*

The simulation runs were evaluated with *REIMOS* based on the target parameters makespan, delivery reliability and resource utilization. The results were then compared to those of the *REIMOS* planning method. The benchmark parameters employed were the lowest makespan of *REIMOS* and the simulation study achieved with order initialization policies 1 and 8. The reasons for choosing these policies for further analyses were that policy 1 was the best policy identified by the industrial partner up to that point and that policy 8 allowed for the highest possible degree of flexibility as all orders are initialized at point  $t=0$ . This ensured a solid basis for the comparison of results from both methods. Out of the two policies, those with the best result in combination with the priority rules were selected. In addition, the simulation methods were compared with regards to the respective makespan of the product groups (block makespan). Figure 4 sets the makespan of the sequences identified by means of the *REIMOS* planning method alongside those generated by the *FEMOS* method. The strategies used were the priority rule shortest operation time in combination with order initialization policy 1 and longest operating time in combination with order initialization policy 8 respectively.

The comparison shows that *REIMOS* provided better results for both order initialization policies. The makespan achieved with *REIMOS* is approx. 11% lower for order initialization policy 1 and almost 7% lower for policy 8 compared with sequence planning on the basis

of priority rules. This proves that the more complex Genetic Algorithm is superior to the priority rules which are less challenging in terms of application.

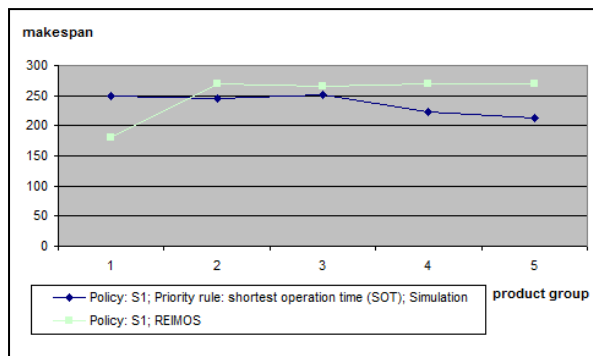


Figure 5. Comparison of makespans identified by means of FEMOS using order initialization policies and priority rules and those generated with the REIMOS planning method.

The diagram illustrated in Figure 5 is the product of a comparison of makespan for production orders of the 5 product groups generated with REIMOS and the simulated application of priority rules respectively. Within the simulation study, the comparison of makespan gave the best result for order initialization policy 1 in combination with the shortest operation rule. However order initialization policy 8 proved to be most effective when combined with the longest operation rule. Figure 4 shows a comparison of these simulation results with the results of REIMOS.

When the makespan of individual product groups for order initialization policy 1 was compared, it turned out that the values were 6% higher with REIMOS than when the priority rules were applied. The utilization of resources and the prioritization of orders were simulated. The best results of the simulation method were compared to REIMOS. In addition, the values were more evenly distributed with the exception of product group 1.

Comparing the makespan of individual production orders for order initialization policy 8 delivered the same results. The values of REIMOS were approx. 21% higher than those generated with the simulated use of priority rules. The individual makespan within one product group did not vary strongly from the group mean. This gives rise to the assumption that a correlation between an equidistant or almost equidistant spacing of interim arrival times within a product group and an improvement of makespan exists.

All of the above proves that the makespan results delivered by the use of priority rules fall short of those delivered by REIMOS.

## 5. CONCLUSION AND OUTLOOK

A computer algorithm based on the evolution of living beings may be surprising, but the extent to which this approach is applied is even more astonishing. Genetic algorithms have already proven their efficiency in many fields of application such as commercial, educational and scientific. Their usefulness in solving various kinds of problems have made them a preferable choice

compared to other, more traditional approaches for a multi-criterion approach to target selection for modeling and evaluations.

The adaptation of a GA to schedule jobs in manufacturing workshops with time, resource and precedence constraints has been demonstrated here (STEININGER 2007). When using such GAs, it is often crucial to implement goal criteria. A new idea is to combine simulation and optimization processes. One concrete scenario would be the use of the simulation tool as a fitness function of the GA in order to allow for more complex weighting functions to be taken into account.

The simplicity of the methods used supports the assumption that GA can provide a highly flexible and user-friendly solution to the JSP. The use of standard software and an implemented "add-in" for Microsoft Project 2003 to realize the GA has shown that this approach can be profitable for solving real world scheduling problems (STEININGER 2007).

## REFERENCES

- Bagchi, T. P., 1999. *Multiobjective Scheduling by Genetic Algorithms*. Boston, Dordrecht, London: Kluwer Academic Publishers.
- Bean, J. C., 1994. Genetic Algorithms and Random Keys for Sequencing and Optimizations. *ORSA Journal of Computing*. Hanover. 6 (2), 154-160.
- Brucker, P., 2004. *Scheduling Algorithms*. Berlin, Heidelberg, New York et al.: Springer Verlag.
- Conway, R. W., Maxwell, W. L., Miller, L. W., 2003. *Theory of Scheduling*. Mineola (NY): Dover Publications.
- Davis, L., 1996. *Handbook of Genetic Algorithms*. Florence (KY): International Thomson Computer Press.
- Gallo, G., Pallottino, S., 1982. A new Algorithm to find the Shortest Paths between all Pairs of Nodes. *Discrete Applied Mathematics*. Amsterdam, 3 (4), 23-25.
- Goldberg, D. E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston (MA), München: Addison Wesley.
- Graham, R. L., Lawler, E. L., Lenstra, J. K. et al., 1979. Optimization and approximation in deterministic sequencing and scheduling. *Annals of Discrete Mathematics*. Amsterdam, 16 (5), 287-326.
- Steininger, P., 2007. *Eine Methode zur Reihenfolgeplanung bei Mehrprodukt-Fertigungssystemen*. Dissertation, Universität Karlsruhe. Aachen (D): Shaker Verlag.
- Taillard, É. D., 2006. *Scheduling instances*. <http://mistic.heig-vd.ch/taillard/problemes.dir/ordonnancement.dir/ordonnancement.html>. Status: 11.04.2011
- Zülch, G.; Greinke, J. 2004. *Simulation-aided Reconfiguration of an Industrial Service System for the Repair of Electrical Tools*. Espoo: Sim-Serv.
- Zülch, G.; Warrisch, W. 2004. *Simulation-aided Segmentation of a Mechanical Parts Manufacturing*. Espoo: Sim-Serv.

# IMPROVING PATIENT SAFETY: A MODEL OF DRIFT IN HEALTH CARE TEAMS

R. Severinghaus, M.S., CMSP<sup>(a)</sup>, C. Donald Combs, Ph.D.<sup>(b)</sup>

<sup>(a)</sup>The Aegis Technologies Group, Huntsville, AL USA

<sup>(b)</sup>Eastern Virginia Medical School, Norfolk, VA USA

<sup>(a)</sup>[rseveringhaus@aegistg.com](mailto:rseveringhaus@aegistg.com), <sup>(b)</sup>[combscd@evms.edu](mailto:combscd@evms.edu),

## ABSTRACT

Rasmussen & Cook (2005) provide a framework for developing a model of risk management to improve patient safety by applying systems dynamics modeling to understand team performance and how it drifts from best practices. Drifting into failure is not so much about individual deficits as it is about organizations not adapting effectively to the complexity of their structure and environment (Dekker 2005). Expanding the conceptual model using the taxonomy of the medical domain to more fully define the interconnections of individual and team performance, risk assessment, management of resources, and recognition of the limits of team workload and maintenance of high performance is an essential next step. Lamb, et. al. (2010) indicates the usefulness of modeling drift in the naval domain. This paper summarizes research underway to develop a health care team specific model of drift. This model characterizes the sets of factors that influence team performance boundaries, identifies and describes team performance and decision making behaviors that determine the size and positioning of the team operating space, a concept developed within the model. A dynamic system will then be described in a manner that can serve as the basis for the development of a variety of simulations intended to improve team performance, maintain high performance over time, and thus improve patient safety.

Keywords: operating room, human error, risk management, team performance, medical errors

## 1. INTRODUCTION

Human error is often cited as a cause of errors occurring in the operating room. As an explanation, it is a time honored practice to place blame on *human* error, and it is strikingly similar to practices in other professions, most prominent among them civilian aircraft accident investigations, but also within many military organizations, both U.S. and international. Such assignment of cause to human mistakes provides both a concise and remediable explanation of what went wrong. However, root causes are rarely, in truth, so readily compartmentalized into categories – others include material failure, unexpected patient complications, etc. While the human component of

failure is certainly present, there are others factors at work. Sidney Dekker [Dekker, 2005], a long experienced aviation analyst, introduced, in his book, *“Ten Questions About Human Error: A New View of Human Factors and System Safety”*, the concept of “drift into failure” to explain aviation accidents. His assertion is that the root cause of human performance failure is frequently an unintentional “drifting” away from best safety practices over time due to pressures of mission, schedule, etc. And this drift is often subtle, not noticed by members of the performance team experiencing such drift.

Recent studies of naval mishaps lend credence to Dekker’s concept, when applied to large team performance. Lamb [Lamb, 2010] and others, based on study of recent naval mishaps, have hypothesized the characteristics of drift for performance teams on long maritime missions. Based on this study and concepts from Rasmussen (1997), they developed a conceptual model of the forces acting to increase the likelihood of a mishap. Their model “provides a framework for ship managers and higher levels of command to better evaluate the current dangers and take actions to reduce them before crew drift meets a golden opportunity for a mishap.”

This characterization of human performance and human error applies as well to medical practice.

## 2. THE PROBLEM

### 2.1. Making the Analogy to Medical Practice

A hospital surgery operating room (OR) is an exercise in highly precise, orchestrated, and monitored execution of complicated and complex procedures, often of high risk, directed by a designated leader guiding the activities of a specialized team, each functioning as a member of a hierarchical organization. This team’s focus is on execution of one major objective (and at times simultaneous sub-objectives), in a time stressed and decision stressed environment. At times, the lead surgeon must make important decisions affecting risk to the patient in real-time, with partial and/or ambiguous information. Successful surgical outcomes depend on a combination of surgeon’s skills, detailed understanding

of the patient's condition, proper execution of all support operations and procedures, and a sequence of correct decisions by all members of the team. Missteps, poor decisions, decisions not made, and loss of situation awareness can each contribute to "drift" from best-possible-practice, and can occur without notice by either the lead surgeon or individual subordinate team members.

## **2.2. Risk Management – Seeing Risk in Real Time**

In many fields, Operational Risk Management (ORM) is supposed to address issues like those found in analyzing naval operations mishaps. However, for humans performing in complex and ambiguous situations, traditional ORM procedures do not provide an effective solution. What is lacking is some way, some methodology, or some mental approach that serves to get all team members on the same page, at the same time, and conscious of existing and potential risks arising in execution of intended tasks or operations. It is our contention this can best be accomplished by creation and implementation of a simple model, analogous to that proposed by Lamb et al, that sees through the complexity and details of high risk patient procedures, and provides a tool to help performers – whether surgeons, naval commanders, or disaster response team leaders - "see" the overall operating situation and correct for drift.

Cook and Rasmussen [Cook, 2005, Cook and Rasmussen, 2005] have described a theory of patient safety in terms of the coupling of technologies and staff in the context of high stakes medical practice, and related loss of safety to the loss of margins, or buffers, to safety limits. The essence of the argument is that when people, technologies, and schedules are tightly coupled, with scant margin for variation, errors and mistakes result from 'going solid', a condition analogous to operation of pressurized water nuclear power plants. The condition is one in which very small perturbations (in water pressure for a nuclear plant; in sequencing patient care action in an ICU) can cause unintentionally large, and often not anticipated, excursions in important parameters.

Naval operations are often an equivalent exercise in highly precise, orchestrated, and monitored execution of complicated and complex procedures. The organization is hierarchical, the procedures difficult, and time and decision pressures intense. In submarine operations especially, in which the hierarchical organization is in many ways a mirror image of a surgical team, effective team interaction and mutual understanding of risks present are essential to sustainment of error-free operations. Where drift occurs from best-possible-practice, opportunities for mistakes multiply, creating opportunity for significant mishaps – the equivalent of medical errors - to occur.

The difference between such team operations in the naval domain and team operations in the surgical domain is team practice, or team rehearsal.

A recent spate of naval mishaps, especially collisions and groundings, has been perplexing and extremely frustrating for those involved. Examples include the underwater collision of two submarines, one French (Le Triomphant) and one British (HMS Vanguard), in the middle of the Atlantic; the grounding of a United States surface warship ship (USS Port Royal) on a reef near Honolulu; and the collision of a United States submarine (USS Philadelphia) with a Turkish merchant ship off the coast of Bahrain.

While these specific mishaps seem to show no discernible pattern in terms of location, circumstance, or ship type, a relevant question is, "could they actually have something in common?" Rasmussen's concepts of 'drift' would argue that it is NOT the case that each naval mishap is a one-in-a-million event that could never have been predicted and can never happen again.

There have been at least 10 major mishaps involving submarines (U.S., UK, and France) since 2000. A review of mishaps involving U.S. submarines in the past 10 years revealed that in all cases there was no single, simple underlying cause. In fact, many errors were found to be subtle and cumulative and, thus, very hard to see. The crews were performing normal tasks in a normal situation. In many reports, it was apparent that the crew's decisions made sense to them in real time and that the mistakes made were not recognized at the time. It seemed to be the case that the crews "drifted" into their eventual troubles. After the fact, during mishap reviews, with the advantage of hindsight, investigators found mistakes, procedural violations, and deviation from best-possible-practice.

Similarly, in review of medical errors, investigators can find mistakes and violations of procedure not seen by OR teams during the course of surgery.

## **2.3. Drift into Failure**

The notion of drift, or drift into failure, was originally developed by Dekker to explain aviation accidents. Dekker [Dekker 2005] defines drift into failure as the "slow, incremental movement of systems operations towards the edge of their safety envelope." This concept can be leveraged to help us understand what could be working to potentially undermine a normally safe, highly complex, sociotechnical system – such as exemplified by surgical OR teams or the command team of a nuclear submarine.

In terms of risk to safety, 'drift' is this accumulation of unintentional, typically minor, human errors and/or procedural violations. As Dekker has pointed out, drift is the "greatest residual risk in today's safe sociotechnical systems." The naval mishap research



revealed that crews were often surprised by their bad outcomes; they never saw it coming. Those same crews also had ample time to invoke a process like ORM, yet, in trying to deal with the complexity and ambiguities of situations they found themselves in, they typically did not (or could not) see the situation for what it really was.

#### 2.4. The Missing Piece

While Dekker's model has much to offer, it does *not* address what is really needed: the development of a practical tool that can be used *proactively* to prevent drift and improve safety. A limitation of the "drift into failure" model is the 'failure' part. Evaluating human performance based on after-the-fact outcomes is too late; recognizing trends – drift – during process/procedure execution is what's missing. A reasonable hypothesis is that performance drift is always present to some degree; at minimum, it must always be expected. An important question is thus: How can health care workers recognize drift in time to correct mistakes before they lead to serious error?

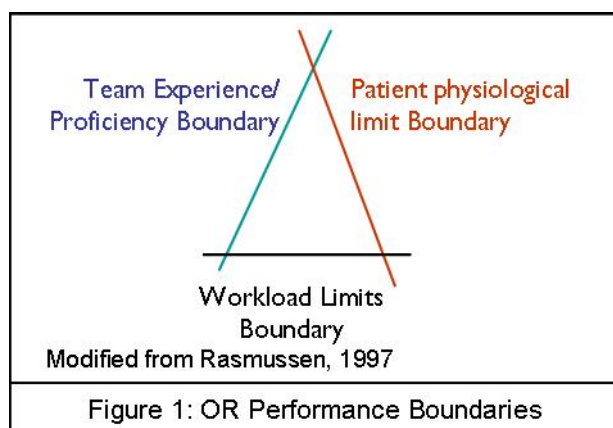
Rasmussen [Rasmussen, 2000] describes a proactive strategy to risk management that is based on:

- Identification of the boundaries for safe performance,
- Efforts to make these boundaries visible to decision makers, and
- Efforts to counteract pressures that drive decision makers toward the boundaries.

His dynamic model of risk and safety can be used to formulate a model of OR safety.

The model is not singly focused on traditional definitions of safety, but rather incorporates a complete system of safety and the various pressures that can contribute to 'un-safety', and thus mishaps.

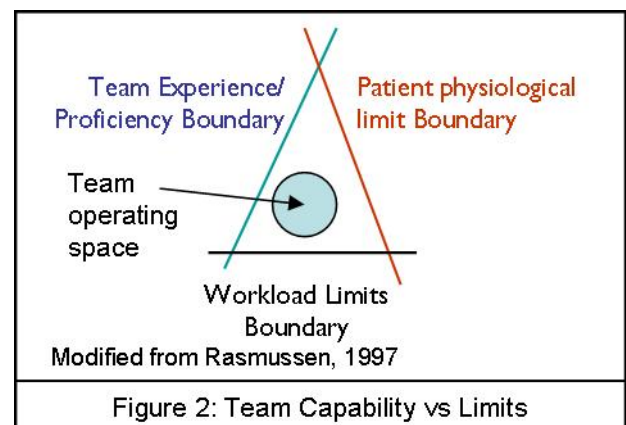
Based on some initial work, we've identified three boundaries for safe OR performance: Team proficiency/experience, patient physiological condition, and team workload. The extent to which a team can stay within these boundaries determines how much drift the environment (OR) and its team can tolerate without failure (*i.e.*, 'drift limits'). Figure 1 depicts this concept.



The triangular area represents an 'operating envelope'; its size is situation-specific and is, therefore, a function of how constraining each of the three boundaries is at the time. The boundaries become more or less constraining depending on the specific details of the surgery /procedure.

Each failure boundary position is set by a number of factors. The workload failure boundary includes rework sleep characterization, fatigue, competing tasks, distractions, etc.

If the boundaries represent the limits of safe operation, then the model must provide a way to determine whether or not the team is inside or outside those limits. Determining this requires an understanding of the current team's *expected* performance level. The term 'functional capability' is used in this context to represent the variability of expected, and allowed, surgical team performance. In terms of the diagram of Figure 1, this 'functional capability' is illustrated by use of a circle placed within the triangle of limits. The capability circle is shown in Figure 2.



The smaller the circle, the less performance is expected to vary; the circle's distance from boundaries is a qualitative indicator of existing buffer, or margin, to the limits shown. The functional capability – the circle's size and positioning – is driven primarily by individual and team skill, experience, and proficiency in execution of the procedures for which the performance boundary diagram applies.

In terms of risk in the Operating Room, "drift" is an unintentional accumulation of human errors and/or procedural violations, each of which are typically minor. In the naval domain research of Lamb et al, it was clear that skilled, experienced watch team were almost always surprised by their bad outcomes. And the mishaps occurred to crews who practiced ORM, who trained and practiced together, and who had all the technology at hand to help avoid errors, mistakes, and bad outcomes. What seemed common to many of the mishaps was that watch teams, in trying to deal with the



complexity and ambiguities of the situations they found themselves in, could not see the situation for what it really was.

## **2.5. Reducing Error in the Operating Room**

This research suggests that what may be of benefit to surgical teams in the OR is development of a surgical team specific drift model - one that characterizes the sets of factors that influence the OR Performance Boundaries of Figure 1, and research to identify and describe the team performance and decision making activities and behaviors that determine the size and positioning of the Team Operating Space as illustrated in Figure 2. Just this research in a naval domain has proven useful as a way to improve team members' self awareness of the performance relationships and pressures influencing real time behaviors and decision making. It is suggested here that focused research to characterize and develop a surgery team specific OR Performance model, with parallel (and related) research understand and characterize individual surgical team member behaviors and activities in terms of the Team Operating Space and its approach to Performance Boundaries would contribute significantly to initiatives focused on reducing errors in the OR, and in improving patient outcomes.

## **REFERENCES**

- Dekker, S. 2005. *Ten Questions About Human Error: A New View of Human Factors and System Safety*. New York: Lawrence Erlbaum Associates.
- Lamb, C., Severinghaus, R., Steed, R., Marvin, K., and Lamb, J. 2010. *A Golden Opportunity; Human Performance in Maritime Mishaps. Proceedings of the Human Performance at Sea conference, Glasgow, UK, June, 2010.*
- Cook, RI, 2005. *Toward a Theory of Patient Safety – Lessons Learned from the First Decade.* In Tartaglia, R, Bagnara S Bellandi T & Albolino S (Eds). *Healthcare Systems Ergonomics and Patient Safety*.
- Cook, RI. and Rasmussen, J. 2005. "Going Solid": a model of system dynamics and consequences for patient safety." *Qual Saf Health Care* 2005; 14:130–134. doi: 10.1136/qshc.2003.009530.
- Rasmussen, J. 1997. Risk Management in a Dynamic Society: A Modelling Problem, *Safety Science*, 27, 183-213.
- Rasmussen, J. and Svedung, I. 2000. *Proactive Risk Management in a Dynamic Society*. Sweden: Swedish Rescue Services Agency.

# MODELING OF GESTURES WITH DIFFERING EXECUTION SPEEDS: ARE HIDDEN NON-MARKOVIAN MODELS APPLICABLE FOR GESTURE RECOGNITION

Sascha Bosse<sup>(a)</sup>, Claudia Krull<sup>(b)</sup>, Graham Horton<sup>(c)</sup>

<sup>(a)(b)(c)</sup> Otto-von-Guericke-University Magdeburg  
P.O. Box 4120  
39016 Magdeburg, Germany

<sup>(a)</sup>[sbosse@st.ovgu.de](mailto:sbosse@st.ovgu.de), <sup>(b)</sup>[claudia.krull@ovgu.de](mailto:claudia.krull@ovgu.de), <sup>(c)</sup>[graham.horton@ovgu.de](mailto:graham.horton@ovgu.de)

## ABSTRACT

Gesture recognition is an important subtask of systems implementing human-machine-interaction. Hidden Markov Models achieve good results for gesture recognition in real-time supporting a low error rate. However, the distinction of gestures with different execution speeds is difficult. Hidden non-Markovian Models provide an approach to model time dependent state transitions to eliminate these problems. In this paper, a basic non-Markovian model structure for gesture recognition is developed. The experiments show that Hidden non-Markovian Models are not only applicable in the field of gesture recognition, but that they can also distinguish gestures with different execution speeds.

Keywords: Gesture recognition, Hidden non-Markovian Models

## 1. MOTIVATION

Automatic gesture recognition plays a key role in human-machine-interaction within virtual environments and multi-modal feedback systems. To comply with the requirements of such systems, the recognition must be performed in real-time and with a small number of errors (Rigol, Kosmola, and Eickeler 1997). Hidden Markov Models (HMMs), successfully applied in the field of pattern recognition, provide satisfactory results that meet these conditions (Frolov, Deml, and Hanning 2008; Rigol, Kosmola, and Eickeler 1997; Chen, Fu, and Huang 2003).

However, by definition, time dependent processes cannot be described easily in these models. A differentiation of gestures with varying speeds of execution is hardly realizable. This could be useful to reduce the number of relevant gestures or to classify gestures by their execution speed to optimize the performance of movements, for example in a virtual environment.

Hidden non-Markovian Models (HnMMs) were developed to support the easy modeling and fast execution of hidden models without the restriction of memoryless state transitions. For this purpose, to each

state transition an arbitrary distribution function can be assigned. The paper will therefore investigate whether HnMM can distinguish gestures with different execution speeds.

## 2. GOAL AND TASKS

The first goal of this paper is to determine whether HnMMs are applicable in the area of gesture recognition in real-time with a similar error rate as provided by HMMs. For this purpose a non-Markovian model will be tested on a standard-PC and should recognize 90% of the executed gestures.

If this goal is achieved, the question remains whether the non-Markovian approach is able to distinguish gestures with differing execution speeds. The second goal of this paper is to determine that and to compare the results with those of the Markovian approach.

The following tasks were identified to reach the goals:

1. Development of an HnMM structure that models gestures
2. Selection of a gesture catalog to test these models
3. Comparison of the HnMMs' result with those of HMMs

## 3. APPROACH

Existing approaches to gesture recognition using HMMs may be divided into two categories: Recognition from image data and recognition from data gloves. The latter approach, in a simplified form, will be pursued here. For this purpose a so-called Wiimote (Figure 1), a remote control of the 2006 published Nintendo Wii, will be used. It can gather motion data in a triple-axis space.



Figure 1: Wii Remote

A great advantage of HMMs is the ability to automatically train models to improve the performance successively. Unfortunately this cannot be done yet with HnMMs, so the developed models must be trained manually (Krull and Horton 2009). Due to the fact that this procedure is very time-consuming, only a small number of gestures will be selected.

#### 4. RELATED WORK

Many papers deal with the application of HMMs in the field of gesture recognition. In the following, some examples of those papers and their results will be highlighted.

Frolov, Deml, and Hanning (2008) investigate the ability of HMMs to improve multi-modal haptic feedback. It is shown that this approach can predict what physical attributes a user wants to know, recognized from his gestures. For that purpose, eight gestures were chosen, which differ in their muscular movement (recorded by a P5-Glove), but not in their execution speed.

In (Rigol, Kosmola, and Eickeler 1997), a real-time system for gesture recognition from image data is presented. It uses HMMs for 24 gestures with a recognition rate of 92.9%. Every state in the model represents a picture frame of the input data. This provides the ability to execute a gesture at any speed, but not the distinction of different speeds.

Chen, Fu, and Huang (2003) present an approach where image data is processed to a spatial and temporal feature vector using the Fourier descriptor and motion analysis. The temporal features are used to enable time-varying hand shapes while recognizing a gesture. The variances in speed are not considered.

Lee and Kim (1999) focus on the recognition of non-gesture hand motion to determine when an important gesture starts and when it stops. An artificial threshold model is used, so that any possible gesture can be described by it. Thus, the likelihood of the dedicated model for this gesture will be highest. The tested gestures are characterized by different spatial motions while the differences in execution speeds are not integrated into the results.

The selected works show that Hidden Markov Models are successfully used to recognize gestures in various different contexts. But the ability to distinguish gestures at different execution speeds was never considered. The question remains whether Markovian

models are not applicable in this context or whether it simply has not been tested yet.

In the following, two approaches (one Markovian and one non-Markovian) are modeled to examine the ability of gesture recognition and distinction.

#### 5. MODELING

The first step to achieve the goals is to bring the data stream of the Wii Remote in a format that can be processed further. Data from the Wii Remote is received nearly 100 times per second, so the values of the three dimensions of the acceleration sensor can be retrieved. Each value stands for the amount of gravity the Wiimote is exposed to in the three directions, normalized to a range from -1 to 1. Because of earth's own gravity also the current orientation of the remote can be accessed. To remove noise from the signal, some kind of windowing is necessary. For this purpose, every ten measurements a mean value is computed, which will be processed for the different models.

##### 5.1. Signal Outputs

In the HMM case, the so formed continuous data stream will be interpreted as the signal output trace of the HMM. Therefore, the values are categorized in specific ranges representing the discrete signal values. These ranges are: *large negative acceleration* [-1,-0.5], *small negative acceleration* [-0.5,-0.1], *no acceleration* [-0.1,0.1], *small positive acceleration* [0.1,0.5] and *large positive acceleration* [0.5,1].

A similar procedure is applied to create traces for the non-Markovian case, but, only significant changes in this data stream are used as trace signals of the model. The next step is to develop basic model classes for the two approaches.

##### 5.2. Markovian Model Structure

To perform a gesture, three phases are passed. The first is to bring the hand in the initial position of the gesture, while the second phase is to execute the different atomic motions. In the last phase, the hand moves from the end to a neutral position. According to this process, the HMMs consist of four states: *Neutral*, *Start*, *Execution* and *Stop*. Figure 2 shows a simplified HMM (hiding the reflexive DTMC arcs) with possible output signals in each state. Each circle represents a state while the solid arrows represent the possible state changes. The arrows with the dashed line represent the possible output symbols of each state.

In the neutral state, only a specified orientation of the remote should be emitted. Every notable movement of the Wiimote should cause the model to change to the start state. If the next motions could be matched as parts of the gesture, the model will change to the execution state. In the other case, the model goes back to the neutral state or it remains in the start state. This procedure should guarantee that random motions do not influence the recognition of a gesture.

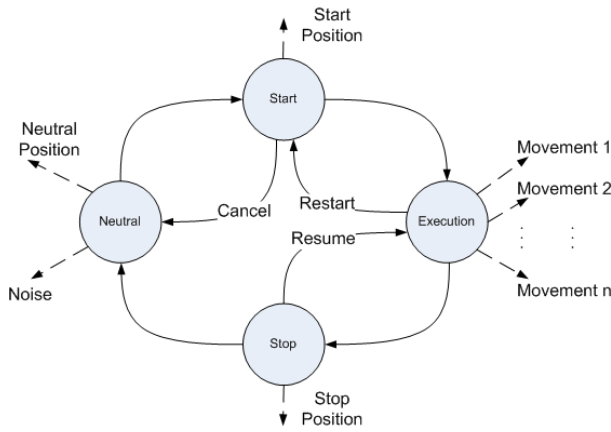


Figure 2: Simple HMM Model for Gesture Recognition

The execution of a gesture consists of different atomic movements, which should be emitted in the execution state. If the execution fails, the user could restart the gesture, so the model returns to the start state. When the remote's movement stops, the model changes to the stop state. In this state the execution can be resumed and further movements can be executed. After performing the gesture the Wiimote should return to its neutral position so that the model changes to the neutral state.

### 5.3. Non-Markovian Model Structure

For the non-Markovian models, a similar approach is developed. The basic idea of the gesture phases can be transferred, but because the output signals are emitted by transitions instead of states, only changes of the movement generate signal outputs. These changes are characterized by the variation of the acceleration values. This variation  $v = acc(t) - acc(t-1)$  has a value range from -2 to 2, so the following classification is done: *large negative variation* [-2,-1], *medium negative variation* [-1,-0.5], *small negative variation* [-0.5,-0.1], *no variation* [-0.1,0.1], etc. Alternatively, a mixture of variation and current acceleration could be used.

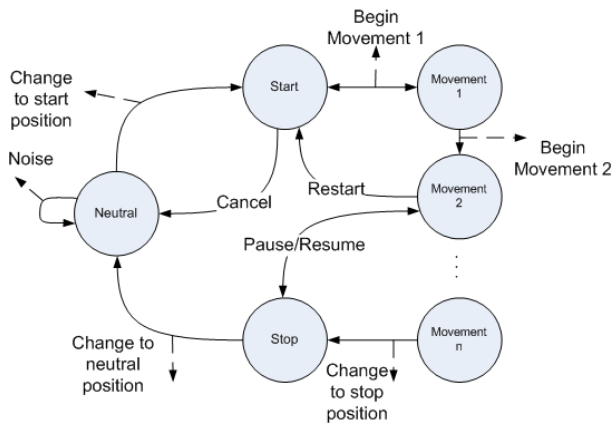


Figure 3: Abstract HnMM Model for Gesture Recognition

In contrast to the HMM, there are no discrete time steps in this model. Time progresses continuously and to each state transition a stochastic time distribution is

assigned. In Figure 3 an abstract state space of the model is presented. Every state transition generates an output. Each circle represents a state and the solid arrows represent the possible state changes. The arrows with the dashed line represent the possible output symbols of each state change.

The model's state space is similar to the HMM, only the execution state is split into  $n$  movements. This is done for reasons of clarity. In the final model there will be only one execution state with  $n$  reflexive transitions. Only these transitions are not uniformly distributed. This should guarantee that the model likelihood depends only on the execution time for movements, not on time for irrelevant movement changes.

## 6. IMPLEMENTATION

After developing the conceptual models, the necessary functions can be implemented. These are:

1. Receive data from the Wii Remote,
2. Compute signal outputs,
3. Train the models and
4. Compute likelihood of Markovian and non-Markovian models.

To communicate with the Wiimote, the open source *Managed Library for Nintendo's Wiimote* (WiimoteLib) for .NET in version 1.7 is used (<http://wiimotelib.codeplex.com/>).

It provides an event which is fired every time the status of the Wiimote changes. In this event the data from the acceleration sensors can be processed. After smoothing the input (see last section) the different signal outputs for both model classes are generated. Depending on the program mode, they are discarded (normal mode), exported to a csv-file (train mode) or saved in lists (recognize mode).

### 6.1. Training and Evaluation of the HMM

To train the HMMs, firstly a movement is chosen. After this, the train mode must be activated and the movement must be carried out. The thus obtained data is used to train the Markovian models automatically, using the Baum-Welch Algorithm as described in (Fink 2008). The following open source implementation was used: *Hidden Markov Models in C#* project by César de Souza (<http://www.codeproject.com/Articles/69647/Hidden-Markov-Models-in-Csharp.aspx>). In addition to the training function this code also provides the evaluation of signal outputs with respect to the models (Fink 2008). So for every model the likelihood of the gesture having been carried out can be computed.

Because there are three dimensions of the acceleration data, five possible output values per dimension in the Markovian and seven output values in the non-Markovian case, the three dimensions are treated independently. That means that for every gesture there are three models trained only with the respective

dimension data. If not, there would be 125 output symbols for the HMM (343 for the HnMM), which would slow down the computation and complicate the training of the non-Markovian models unnecessarily. However, to prevent that the assumption of full independence distorts the results too much, the three models are connected in a way that a state transition must be executed simultaneously in every model. The probability of emitting a special symbol tuple is computed by multiplying the three single output probabilities.

## 6.2. Training and Evaluation in the HnMM

To obtain the model parameters in the non-Markovian case, a manual training is carried out. For that purpose the created csv-files are checked. Every time a significant change of the acceleration sensors was recognized, the time stamp and the symbol outputs were recorded. With respect to the gesture carried out, the characteristic movement changes are identified as state transitions. The parameters of the transitions non-Markovian distributions are estimated from the time span between these movement changes. The symbol output probabilities are derived from the recorded symbols by their frequency in the particular csv-files.

After training the HnMMs the likelihood of a gesture must be computed. To reach the goal of real-time computation, an adaption of the established HMM algorithms is required. Since the here developed models fulfill some properties (every transition omits a symbol, only one transition can be fired between two states and no race age transitions are allowed), the original formula of the Forward-Algorithm (Fink 2008) to evaluate a given symbol sequence (see Equation (1)) can be adapted to Equation (2), where the probability of the state transition  $a_{ij}$  is replaced by the integral of the state change rate over the time elapsed since the last state change (Krull and Horton 2009). This rate corresponds in this case to the instantaneous rate function (IRF) shown in Equation (3), where  $f(t)$  is the probability density function and  $F(t)$  is the cumulative distribution function of the state change distribution from  $i$  to  $j$  over time (Horton 2002).

$$\alpha_{k+1}(j) = \sum_{i=1}^N \alpha_k(i) * a_{ij} * b_j(o_{k+1}) \quad (1)$$

$$\alpha_{t_{k+1}} = \sum_{i=1}^N \alpha_{t_k}(i) * \int_0^{t_{k+1}-t_k} a_{ij}(x) dx * b_{ij}(o_{k+1}) \quad (2)$$

$$a_{ij}(t) = \mu_{ij}(t) = \frac{f(t)}{1 - F(t)} \quad (3)$$

To compute an approximation of the integral, the trapezoidal rule (see Equation (4)) is applied.

$$\int_a^b f(x) dx \approx (b-a) \frac{f(a) + f(b)}{2} \quad (4)$$

With the initialization of  $\alpha_0 = \pi_i$ , a recursion can be implemented which finally provides the sum of the probability of all possible paths, which can be interpreted as the likelihood of the model to generate the given symbol sequence.

When all likelihood values are computed, the results are compared and the best fitting model corresponds to the most likely gesture. So now the evaluation results of the Markovian and the non-Markovian models can be compared for the chosen gestures.

## 7. EXPERIMENTS AND RESULTS

### 7.1. Experiment Description and Expectations

To compare the two different approaches, some experiments will be carried out. To this end, gestures are chosen and the corresponding models are created. While performing these gestures, the output will be saved, so the models can be trained. Afterwards both models should recognize the gesture when it is executed in real-time. This whole procedure is done with one gesture in two execution speeds. If the non-Markovian model can distinguish these two speeds while the Markovian cannot, the first model is a better approach for this purpose.

This comparison is planned for one simple and one more complex gesture. The simple one is a movement upwards. Executing this at different speeds could be used to distinguish a scrolling move from a “to-the-top-move”. The more complex gesture is a sequence of the atomic movements left, right, up.

For the four gestures up fast ( $Uf$ ), up slow ( $Us$ ), left-right-up fast ( $LRUf$ ) and left-right-up slow ( $LRUs$ ), The Hidden Markov Models are trained with ten csv-files until the difference between the likelihood of two consecutive iterations is smaller than  $1 \cdot 10^{-5}$ . As described in the above section, the non-Markovian models are trained manually.

### 7.2. Initial Experiment

To test the trained models, every gesture was executed ten times and the most likely gesture was recorded. Table 1 shows the results of the Markovian and Table 2 the results of the non-Markovian models. The bold numbers mark a correct recognition. In this initial experiment, the HMMs provide a recognition rate of only 50%, the HnMMs a rate of 75% over all performed gestures.

Table 1: HMM Results of Initial Experiment

	Exec.	Uf	Us	LRUf	LRUs
Reco.		0	0	20	20
Uf	10	<b>0</b>	0	10	0
Us	10	0	<b>0</b>	0	10
LRUf	10	0	0	<b>10</b>	0
LRUs	10	0	0	0	<b>10</b>

Table 2: HnMM Results of Initial Experiment

Reco.	Exec.	Uf	Us	LRUf	LRUs
		10	17	9	4
Uf	10	<b>8</b>	2	0	0
Us	10	0	<b>10</b>	0	0
LRUf	10	2	0	<b>8</b>	0
LRUs	10	0	6	0	<b>4</b>

While the results of the non-Markovian models are acceptable, the classical models failed abnormally. Table 1 indicates that the atomic gesture *Up* cannot be distinguished from the more complex gesture *Left-Right-Up*.

After closer examination, it was found that a property of the Wiimote's acceleration sensor is responsible: Because of earth's own gravity, the sensor's z-axis is much more sensible to changes than the other dimensions. Since the left-right-movements left the Wiimote in its initial alignment with respect to the ground these movements were identified as a kind of noise, thus the distinction of the two gestures failed.

### 7.3. Experiment with Adapted Gestures

To make the different gestures more easily distinguishable, the two gestures LRUf and LRUs are changed, so that the left and right movements are supported by a rotation. Therefore, the effect of the gravity should be recognized in other dimensions as well.

The result of the experiment with the adapted gestures are shown in Table 3 for the HMM and Table 4 for the HnMM. They show a recognition rate of 85% in the Markovian case and 88% in the non-Markovian case.

Table 3: HMM Results with Adjusted Gestures

Reco.	Exec.	Uf	Us	LRUf	LRUs
		5	9	15	11
Uf	10	<b>5</b>	0	5	0
Us	10	0	<b>9</b>	0	1
LRUf	10	0	0	<b>10</b>	0
LRUs	10	0	0	0	<b>10</b>

Table 4: HnMM Results with Adjusted Gestures

Reco.	Exec.	Uf	Us	LRUf	LRUs
		8	14	11	7
Uf	10	<b>8</b>	2	0	0
Us	10	0	<b>10</b>	0	0
LRUf	10	0	0	<b>10</b>	0
LRUs	10	0	2	1	<b>7</b>

The results are now in line with the expectations. The non-Markovian model shows slightly better results, because the Markovian one still cannot completely distinguish the *Up* from the *Left-Right-Up* gesture. The errors in the non-Markovian case can be explained by the lack of robustness induced by the manual training.

However, contrary to our expectations, the distinction of slowly and fast executed gestures can be

done by both models. A possible cause for this behavior could be that the chosen symbol sequences, the acceleration sensor values, are indicators of the execution speed, because their actual values strongly depend on the intensity of a movement. To eliminate this behavior, the symbol sequences are adapted so that they only specify whether a value is positive, neutral or negative.

### 7.4. Experiment with Reduced Output Symbol Set

To reduce the number of possible output signals (and also their information value), the discretisation intervals are adjusted as follows: In the HMM case, the signals are *negative acceleration* [-1,-0.1], *no acceleration* [-0.1,0.1] and *positive acceleration* in [0.1,1]. In the HnMM case, they are *negative variation* [-2,-0.1], *no variation* [-0.1,0.1] and *positive variation* [0.1,2].

After this reduction a new training is performed. The results of the following experiments are presented in Table 5 and in Table 6 (non-Markovian). In this sample the HMMs recognize 55% and the HnMMs 80% of the executed gestures. After a closer look on the results it becomes obvious that the non-Markovian model's behavior is similar to the last test series. The Markovian model however is no longer capable of distinguishing fast and slow gestures correctly.

Table 5: Final HMM Results

Reco.	Exec.	Uf	Us	LRUf	LRUs
		6	3	19	12
Uf	10	<b>6</b>	0	3	1
Us	10	0	<b>3</b>	5	2
LRUf	10	0	0	<b>7</b>	3
LRUs	10	0	0	4	<b>6</b>

Table 6: Final HnMM Results

Reco.	Exec.	Uf	Us	LRUf	LRUs
		7	12	16	5
Uf	10	<b>7</b>	0	3	0
Us	10	0	<b>10</b>	0	0
LRUf	10	0	0	<b>10</b>	0
LRUs	10	0	2	3	<b>5</b>

To illustrate this conclusion, the normalized likelihoods of two exemplary gestures are shown. Figure 4 shows the likelihoods of all eight models when a *Left-Right-Up Fast* movement is executed, which was recognized correctly by both models. Figure 5 shows the likelihoods of all eight models when an *Up Slow* movement is executed, which was only recognized by the HnMM. These diagrams show that the recognition is clear in the non-Markovian case but ambiguous in the Markovian case.

So in the end the Hidden non-Markovian Models provide better results in the differentiation of gestures' execution speed if the decoded signal outputs are not an indicator of the speed of the gesture.



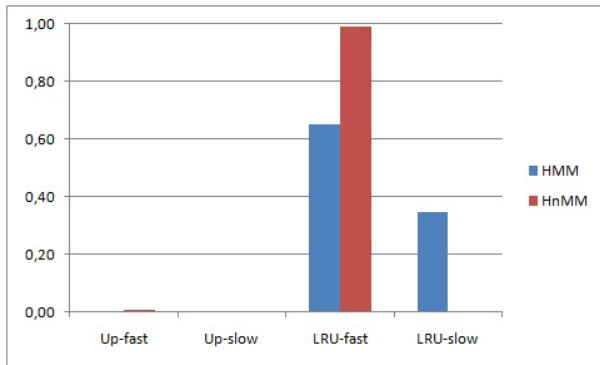


Figure 4: Normalized Likelihood of LRUF

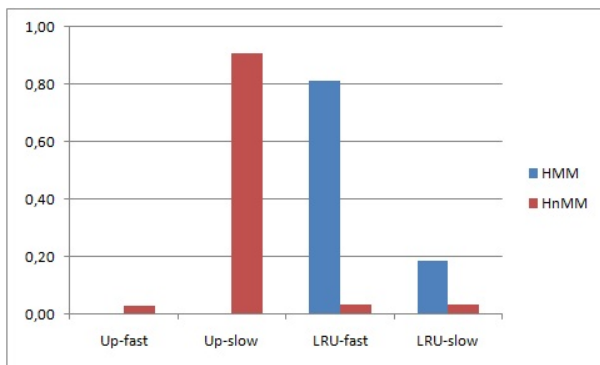


Figure 5: Normalized Likelihood of Us

## 8. CONCLUSION

In this paper it was shown that Hidden non-Markovian Models are applicable in the field of gesture recognition. We could also demonstrate that, under certain circumstances, non-Markovian models provide a better distinction of gestures with different execution speeds than Markovian models do. So the goals of this paper are reached. Further it was shown that it can be necessary to discard a Markovian model because the ability of gesture distinction is not sufficient. Hidden non-Markovian Models represent an alternative approach to solve this problem.

So it is feasible to introduce time dependent state transitions to distinguish similar signal sequences by their output speed. This opens up opportunities to model transitions with arbitrary stochastic distributions in applications with hidden states. Adding meaning to the execution speed of a gesture can reduce the gesture catalogue, which a user has to learn. Looking ahead it can be postulated, that HnMMs are able not only to decide whether an arbitrary pattern was executed correctly, but also whether it was executed at the right speed.

The main disadvantage of HnMMs is the manual training process. This includes the development of the basic model and the computation of state change distributions and output probabilities. In addition to the large time effort, this method also produces less robust models than automatic learning algorithms do. This is still an area of future work and must be automated somewhat to make the approach feasible in real applications.

Apart from the modeling approach it is also shown that the usage of the Wii Remote in the field of gesture recognition has disadvantages. The built-in acceleration sensor is very sensitive to rotation movements but not that sensitive to translation movements. Maybe a combination of acceleration sensor and position tracking with the infrared camera could eliminate this problem.

## REFERENCES

- Chen F.; Fu C.; and Huang C., 2003. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing* 21. 745-758.
- Fink G.A., 2008. Markov Models for Pattern Recognition. Springer Verlag. ISB 978-3-540-71766-9.
- Frolov V.; Deml B.; and Hanning G., 2008. Gesture Recognition with Hidden Markov Models to Enable Multi-modal Haptic Feedback. In *EuroHaptics 2008*. 786-795.
- Horton G., 2002. A new paradigm for the numerical simulation of stochastic petri nets with general firing times. In *Proceedings of the European Simulation Symposium 2002*. 129-136.
- Krull C. and Horton G., 2009. HIDDEN NON-MARKOVIAN MODELS: FORMALIZATION AND SOLUTION APPROACHES. In *Proceedings of 6<sup>th</sup> Vienna International Conference on Mathematical Modelling*. 682-693.
- Lee H. and Kim J., 1999. An HMM-Based Threshold Model Approach for Gesture Recognition. In *IEEE Transactions on Pattern Analysis and Machine Intel- ligence* 21. 961-973.
- Rigol G.; Kosmola A.; and Eickeler S., 1997. High Performance Real-Time Gesture Recognition Using Hidden Markov Models. In *Lecture Notes in Computer Science Vol. 1371*. 69-80.



# MODELLING AND INTERPOLATION OF SPATIAL TEMPERATURE DURING FOOD TRANSPORTATION AND STORAGE BY THE VARIOGRAM

Reiner Jedermann<sup>(a)</sup>, Javier Palafox- Albarrán<sup>(b)</sup>, Pilar Barreiro<sup>(c)</sup>, Luis Ruiz-García<sup>(d)</sup>, Jose Ignacio Robla<sup>(e)</sup>,  
Walter Lang<sup>(f)</sup>

<sup>(a),(b),(f)</sup> IMSAS, University of Bremen Germany

<sup>(c),(d)</sup> Universidad Politécnica de Madrid, Spain

<sup>(e)</sup> National Center for Metallurgical Research, Spanish Council for Scientific Research (CENIM-CSIC), Spain

<sup>(a)</sup> [rjedermann@imsas.unibremen.de](mailto:rjedermann@imsas.unibremen.de), <sup>(b)</sup> [jpalafox@imsas.unibremen.de](mailto:jpalafox@imsas.unibremen.de), <sup>(c)</sup> [pilar.barreiro@upm.es](mailto:pilar.barreiro@upm.es),  
<sup>(d)</sup> [luis.ruiz@upm.es](mailto:luis.ruiz@upm.es), <sup>(e)</sup> [jrobla@cenim.csic.es](mailto:jrobla@cenim.csic.es), <sup>(f)</sup> [wlang@imsas.unibremen.de](mailto:wlang@imsas.unibremen.de)

## ABSTRACT

A better control of food transport, storage and processing is achieved if single-point temperature measurements are replaced by spatial monitoring. The statistical temperature distribution inside a room is described by a Variogram model. The temperature for any point in space can be interpolated by the Variogram based Kriging method. The accuracy of the Kriging method was tested on 14 experimental data sets recorded in cold-storage rooms, delivery trucks, and containers. The interpolation error can be reduced by up to 68% compared to an average measurement. Kriging showed also a clear advantage over the Inverse-Distance-Weighting interpolation method, with a higher reduction of the error by 20% except for 4 data sets with an insufficient spatial sensor density. Different approaches for automated estimation of the Variogram model parameters were compared. The influence range of temperature deviations was evaluated to values between 1.1 and 4.7 meter, depending on the air-permeability of the food packing.

Keywords: Variogram modelling, Kriging, spatial interpolation, temperature supervision

## 1. INTRODUCTION

Local temperature deviations during the processing, storage and transport of foods can degrade their quality. Our measurements showed that temperature cannot be considered as constant over space; the temperature profile shows a number of cold and hot spots, which influence their neighbourhood within a certain range.

Because the number of available sensors is limited in practical applications, the temperature for points in between the sensors has to be calculated by interpolation. The Kriging method provides a statistically correct interpolation. Aim of this paper is to test the performance and accuracy of this method for typical spatial temperature supervision tasks in food logistics on recorded data sets of 14 experiments, carried out in cold storage rooms, delivery trucks and containers.

The first step of the Kriging method comprises of a statistical analysis of the given measurements. A so-called theoretical Variogram model is derived from the experimental data, describing the relation between expected temperature deviation and the distance between the probe points. There is only a small set of mathematical functions, which are typically used as Variogram models, sharing the same parameters: nugget, sill, and range.

The nugget gives the deviation that should be expected even for small distances caused by noise, sensor tolerances, and border effects. The expected deviation increases within the Variogram range but stays almost constant afterwards at the sill value. If the probe area is large enough, the variance of all measurements can be taken as a raw estimation for the sill value.

The following section will briefly introduce the Variogram and the Kriging method. It also defines the test method to evaluate the interpolation error. Section 3 will introduce different approaches for an automated estimation of the Variogram parameters.

Our experimental database is described in Section 4, consisting of 8 sets recorded in a cold storage room either empty or loaded with 2 tons of water bottles to simulate food storage. Six further sets were recorded during regular food transports.

The size of the influence range of temperature deviations depends very much on the air permeability of the cargo. Range of values between 1.1 and 4.7 meter were found depending on the type of cargo (Section 5).

The evaluation of the interpolation accuracy in Section 6 showed that the Kriging interpolation reduces the prediction error by 35% to 68% for the tests in cold storage rooms and trucks, compared to a Null-model that ignores spatial dependency and takes the average sensor measurements as temperature prediction. But the interpolation of the container data sets hardly showed any improvement. An index value to evaluate whether a given sensor density is sufficient for reliable interpolation was derived from the data sets.

The spatial interpolation shall be integrated into an intelligent container for fully automated supervision of

food transportation. The required telemetric system and an automated shelf life evaluation is currently developed as a separate focus of our project. Our evaluation of the required CPU recourses in section 7 showed that the Kriging method is fully capable of running on embedded systems that could be integrated into a truck or a container.

## 2. BACKGROUND OF THE KRIGING METHOD

Kriging (Wackernagel 2003; Schabenberger and Gotway 2005) applies a linear interpolation to predict the temperature in one destination point by multiplying the available measurements with a set of weighting factors. There are some heuristic methods to set the weighting factors such as the Inverse-Distance-Weighting (IDW). The application of the Variogram to set the Kriging weights provides a statistically correct estimator for the weighting factors, and therefore, is the best linear estimator under the condition that the expected value for the difference between two points depends only on their distance vector and not on their absolute position. The Kriging method originates from Geology, but it has been adapted to sensor measurements in food transports in our previous publication (Jedermann and Lang 2009).

An experimental Variogram is calculated from the measurements and then fitted with a theoretical model. The Spherical model and the Gauss model gave the best results for our tests. The model parameters are selected in a way to minimize the error between experimental and theoretical Variogram  $\Phi$ .

### 2.1. Method for evaluation of the interpolation error

The experimental data sets were split into a set of source points as input for the interpolation and a set of destination or reference points. The measurement of the latter set was compared with the result of the interpolation.

The error of the interpolation  $\varepsilon$  was calculated as Root-Mean-Square difference between the interpolated prediction and the actual measurement over all sampling intervals and destination points.

Because of sensor tolerances and the nugget effect, it is not possible to reduce the interpolation error to zero. The interpolation error of the Kriging method was compared with two other prediction methods as reference.

The Null-model assumes that there is no dependency between location and measurement. For this case, the only possible way is to take the average of all available measurement locations as predictions for all destination points. The mathematically simpler Inverse-Distance-Weighting was used as the second reference method.

The Kriging method does not only provide a prediction for the destination points, but it also estimates the standard deviation  $\sigma_k(i)$  of the prediction for any destination point  $i$ .

Furthermore, the general accuracy of the Variogram and the Kriging method can be evaluated by comparing the average relation  $\theta$  between the predicted Kriging standard deviation  $\sigma_k(i)$  and the actual error  $\varepsilon(i)$  for each destination point  $i$  as suggested by Wackernagel (2003).

The relation should be about 1; otherwise, the Variogram model could be inaccurate, the temperature distribution could contain an anisotropic dependency that has not been considered, or the temperature differences could depend on absolute positions and not only on the distance vector.

## 3. METHODS FOR AUTOMATED VARIOGRAM ESTIMATION

A fully automated system for transport supervision has to adapt the Variogram parameters range, nugget and sill to the experimental data without human interference. Three unsupervised methods were tested and compared according to their stability, computational efficiency, resulting interpolation error  $\varepsilon$ , and test of the relation  $\theta$ .

### 3.1. Grid search

As the first attempt, a *brute-force* grid search was implemented. The grid search tests all combinations of range, nugget and sill for the lowest fitting error  $\Phi$  between a given set of boundaries. This method turned out to be very stable. For an adequate setup of the boundaries, the variance of all measurements  $\sigma_M^2$  is taken as raw estimation of the sill. The grid search tested for values between 60% and 180% of the measurement variance  $\sigma_M^2$ .

The nugget must be smaller than the sill; otherwise, the Variogram model does not increase monotonically. But, we found that the nugget values of zero as well as too-high values can lead to a poor relation  $\theta$ . Therefore, the lower and upper grid boundaries for the nugget value were set to 2% and 20% of the measurement variance  $\sigma_M^2$ , respectively. The boundaries for the value of the Variogram range were rather uncritical; therefore, the grid search tested for values between 1 and 10 meters.

### 3.2. Nelder-Mead algorithm

A more efficient Variogram fitting algorithm was found in (Schwaghart 2009). He provides a Matlab script to minimize the fitting error  $\Phi$  to an experimental, isotropic Variogram. However, the minimum is not found by the least squares method, instead, it uses the algorithm from Nelder and Mead (1965), which is a heuristic, well-known method to be effective and computationally compact as it does not need any matrix inversion.

Furthermore, it provides additional advantages that may improve the good fitting of the function: it allows weighting the least squares if the number of observation per experimental lag is provided. Two weighting schemes are selected from Geostatistics literature. The first one is based on Cressie (1985). It automatically gives most weight to early lags and down-weights to

those with small number of lags; the second one, which makes use of the Akaike information criterion as a measure of the goodness of the fit, is based on McBratney and Webster (1986).

### 3.3. Fixed Variogram parameters

The sill depends very much on the experimental conditions such as initial temperature differences at loading, whereas the range and nugget have only little variations for transports with similar loading schemes. Based on this observation, a further method for parameter estimation was created that adapts only the sill value to the actual experiment. Our database was divided in to 5 groups according to the type of cooling and loading scheme. The experiments of each group shared the same nugget and range values.

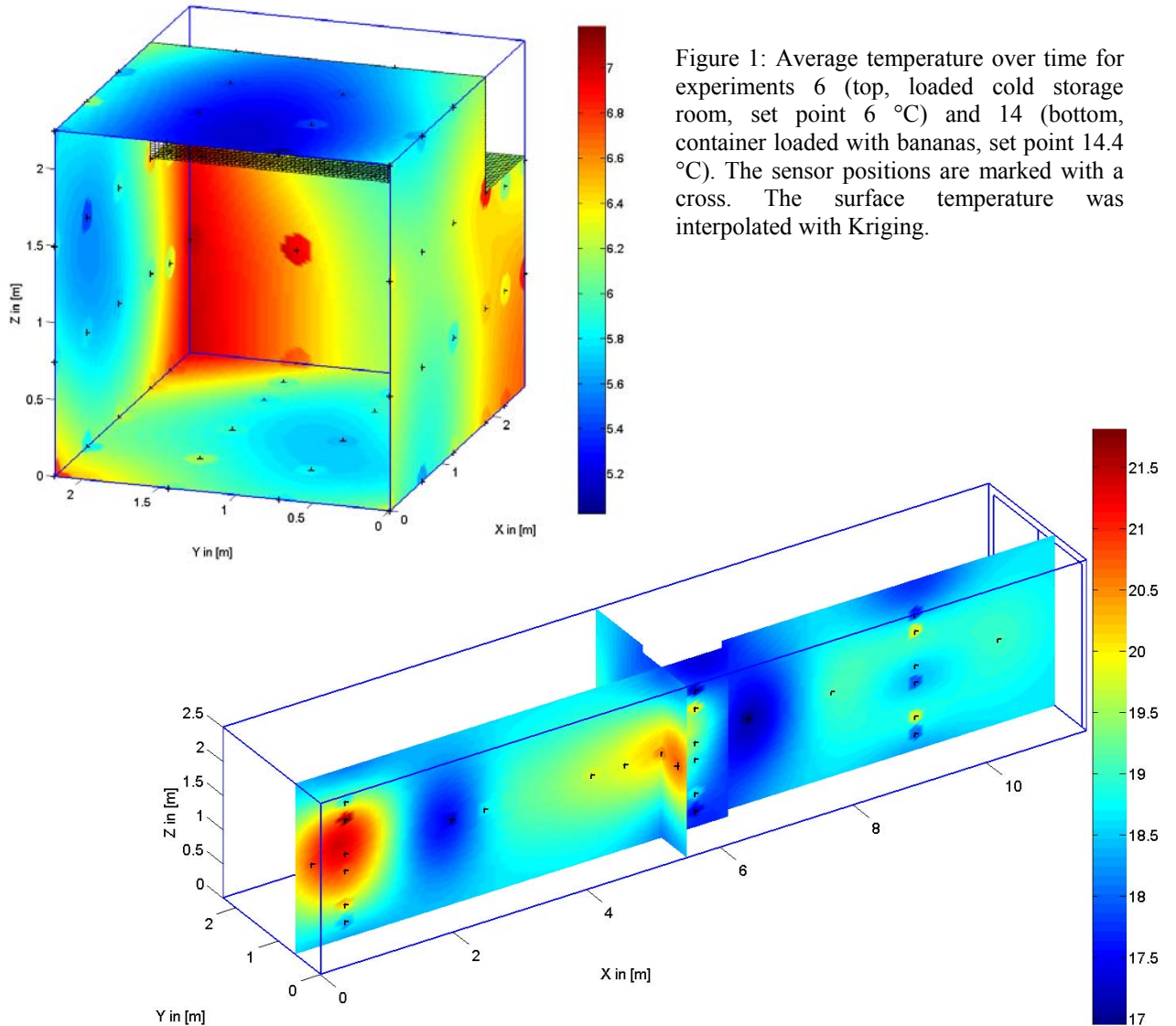
The range was set to the average value of the previous experiments belonging to the same group. Because of the problems defining the boundaries for the nugget value, we tried to set the nugget directly by physical considerations. The nugget depends, among other factors, on the tolerance of the sensors. The

nugget was directly set to the square of the measured sensor tolerances.

The sill was calculated to fit the average value of the theoretical model to the experimental Variogram for large distances. The number of required mathematical operations was thereby largely reduced. Furthermore, the accuracy of the experimental Variogram is less critical because only the average is required.

## 4. EXPERIMENTAL DATABASE

Eight data sets were recorded by three authors of this paper at the wall-sides of a cold storage room (Rodríguez-Bermejo, Barreiro, Robla and Ruiz-García 2007). The measurements were taken in one compartment with the size of  $2.6 \times 2.2 \times 2.3$  meters with 54 or 68 PT100 probes. Figure 1 (top) shows the typical temperature distribution at the walls. Three dual-state conditions were combined to make up the 8 experiments. The conditions are loading state (empty/full), set point ( $0^\circ\text{C}/6^\circ\text{C}$ ) and type of cooling (on-off /modulated).



Further data sets were recorded in delivery trucks by the first author of this paper (Jedermann and Lang 2009) in cooperation with Rungis Express, which is a German food supplier for hotels and restaurants. The trucks had 3 separate chambers or temperature zones. Two sets (Experiment 9 and 10) were recorded in the deep freezer chamber with a size of  $2.9 \times 2.5 \times 2.35$  meters at a set point of  $-29^{\circ}\text{C}$ . The compartment was partly filled with frozen meat in air-permeable boxes. In 2009 a first test of the Kriging method was carried out with these two data sets, but the results are not directly comparable because of the now refined method for evaluation of the interpolation error and the modified error criterion.

In further tests, the data loggers were not placed at the walls but inside the freight. Two data sets from a sea transport of bananas with 45 sensor positions were provided by Maersk (Experiment 11 and 12). Two sets were recorded by the first author of this paper in cooperation with Dole on banana transports from Costa Rica to Europe in 2010 and 2011. Twenty-seven iButton data loggers were packed in the centre of the banana boxes for the experiment 13 and 31 loggers for the experiment 14. Because the bananas inside the boxes were packed in a plastic bag to prevent humidity loss, the air could only flow through small gaps between the pallets.

## 5. RESULTING VARIOGRAM MODELS

Table 1 shows the average range of the grid search Variogram models sorted by groups of experiments. The average range values were used as parameters for the fixed models. The Table indicates that the range depends on the packing of the loaded cargo. If the air can circulate freely in the room, temperature distortions can spread over a wider range by convection. The highest range of 4.7 metre was measured in empty cold storage rooms. Partially filled cold storage rooms (Figure 2) and trucks (Figure 3) showed almost the same range of 3.25 to 4 meter. For densely packed cargo such as bananas in sea containers (Figure 4) the range dropped to 1.65 or 1.125 meter.

The experimental Variograms for the container tests turned out to be very noisy which might be caused by anisotropic dependencies of the variance from the direction of the distance vector between the pair of points. But the number of available probe points in the container tests was clearly insufficient for an analysis of anisotropic effects.

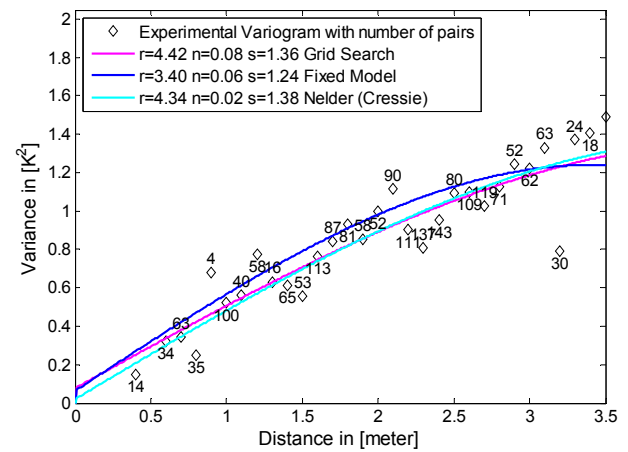


Figure 2: Experimental Variogram for experiment 8 (loaded cold storage room) and Spherical models resulting from different estimator algorithms.

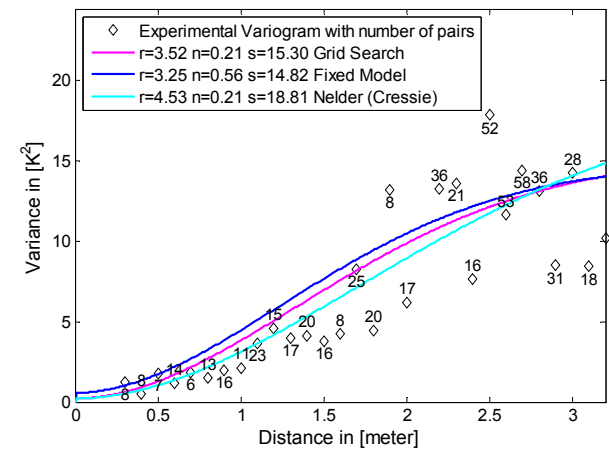


Figure 3: Gauss Variogram for partly filled truck (experiment 10)

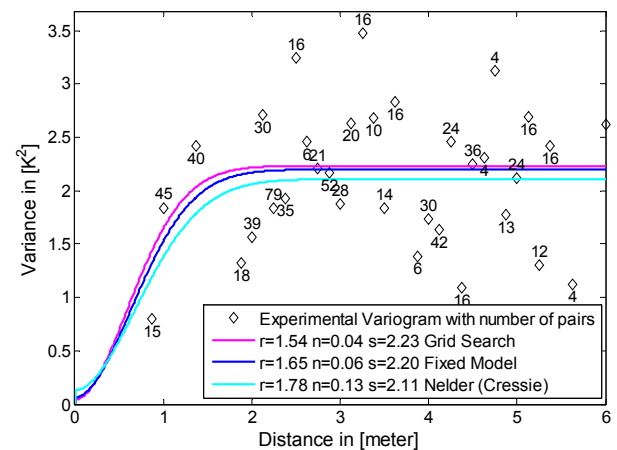


Figure 4: Gauss Variogram for sea container loaded with bananas (experiment 11)

## 6. EVALUATION OF INTERPOLATION ACCURACY

The Kriging interpolation was calculated based on the resulting Variogram parameter for the grid search, the fixed model and for the Nelder/Cressie search algorithm. The prediction error for the Kriging interpolation was compared with the Null-model and the Inverse-Distance-Weighting model as reference in Figure 5.

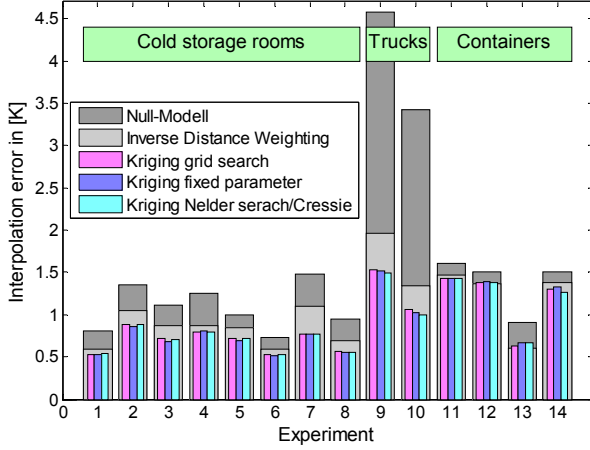


Figure 5: Error between prediction and measurements for different methods for Variogram estimation. Compared with the Null- and the IDW-model.

The highest reduction of the interpolation error was achieved for the truck tests by Kriging. The tests in the cold storage room showed a smaller but still remarkable improvement compared with the two reference models. But the Kriging method brought only very little advantage compared to the Null-model and was sometimes even worse than the Inverse-Distance-Weighting for the container tests. The selection of the method for estimation of the Variogram parameters had only a negligible effect on the interpolation error.

Therefore, the best method was selected based on the relation between Kriging Variance and actual interpolation error  $\theta$ , given in Figure 6. The relation for the tests in the cold storage room almost arrived at the

target value of 1. But for the last two container experiments at Dole the relation increased to values of up to 3.5. This is a further indication that the interpolation is inaccurate for the containers with the current sensor setup. But this does not mean that the interpolation is not possible for densely packed containers, or that there are no spatial dependencies of temperature, but merely that a higher number of sensors are required.

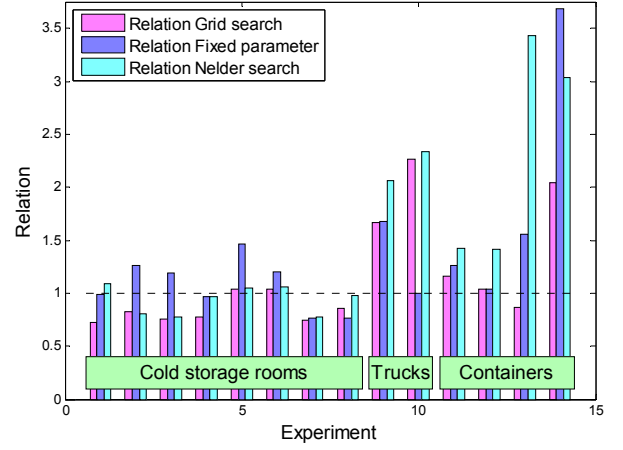


Figure 6: Relation between Kriging Variance and actual interpolation error  $\theta$  for different methods for Variogram estimation.

Depending on the type of experiment, different methods for Variogram estimation gave the best result for the relation  $\theta$ . The selected methods are summarized in Table 2. The interpolation error can be reduced by up to 68.5 % if Kriging is used for interpolation compared with the Null-model. Compared with the Inverse-Distance-Weighting, the Kriging interpolation is in average 20% better except for the container tests.

The weighting according to McBratney and Webster (1986) was rejected because it gave only inaccurate Variogram parameter for two tests in the cold storage room. Furthermore, the relation was worse than by all other methods.

Table1: Fixed range parameter for groups of experiments. For the number of neighbours in range see section 6.1.

Group	Range	Model type	Neighbours in range
Empty cold storage room	4.7 metre	Spherical	29.9
Loaded cold storage room	3.4 metre	Spherical	24.7
Truck, partly filled	3.25 metre	Gauss	24.5
Container Maersk, inside bananas	1.65 metre	Gauss	3.8
Container Dole, inside bananas	1.125 metre	Gauss	4.0

Table 2: List of methods that gave the best relation  $\theta$  for the different types of experiments.

Type	Best method	Model type	Improvement over Null-model	Improvement over IDW-model
Cold storage room	Nelder/Cressie	Spherical	35.8 %	16.0 %
Truck	Fixed parameters	Gauss	68.5 %	23.4 %
Container	Grid search	Gauss	16.1 %	1.4 %



### 6.1. Index for sensor density

The question arises why the interpolation gave such a poor result for the container experiments although a similar number of probe points were used. But, the difference between the types of experiments becomes obvious if for each destination point the number of neighbouring source points within the Variogram range is counted. This number is important because the source points with a distance lower than the Variogram range have the most influence on the interpolation result.

The cold storage rooms and trucks had an average number of *neighbours in range* between 24 and 30, whereas the sensor setup for the containers had only 3.8 or 4 neighbours in range as listed in Table 1.

The number of neighbours in range can be taken as an index value to evaluate whether a proposed sensor setup is sufficient for interpolation.

In order to estimate a threshold for the index value, the number of source points was incremented step by step in Figure 7. The horizontal axis was recalculated in order to directly display the number of neighbours in range instead of the total number of source points.

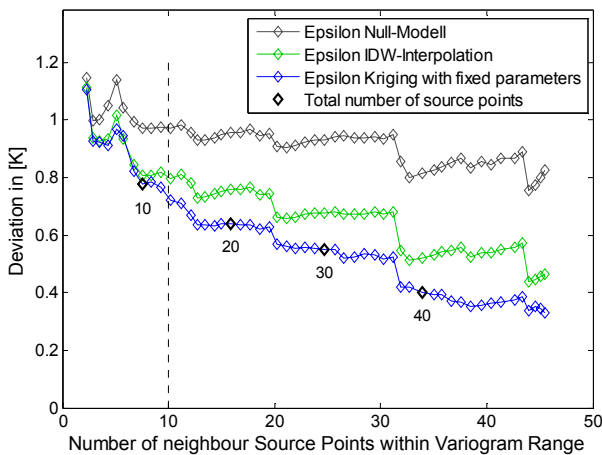


Figure 7: Interpolation error as function of the number of neighbours in range for experiment 8 (loaded cold storage room). Total number of source points marked by diamonds.

If the number of neighbours in range is higher than 10 (dotted line in Figure 7), the Kriging interpolation results in a lower error than the Inverse-Distance-Weighting. At this index value, Kriging has also a clear benefit compared to the Null-model.

## 7. EMBEDDED KRIGING ON INTELLIGENT SENSORS AND CONTAINERS

So far, Kriging was considered as an offline tool to analyse recorded data sets, but the paramount aim of our project is to obtain real-time information during transport. Measurement data should be directly processed within the means of transportation or even by the sensors themselves.

Besides the Kriging interpolation several other approaches for local data processing were evaluated within our project, including calculating the effects of

temperature deviations on the product quality by a shelf life model and predicting the future temperature development of a cool-down process by a system model (Palafox et al. 2011). We consider wireless sensor nodes not only as a tool for remote data acquisition, but as a platform for decentralised decision making. But, it has to be questioned whether the CPU resources of wireless sensor platforms are sufficient to execute complex algorithms.

The Variogram estimator and the Kriging interpolation were implemented as Java software bundles and tested on the iMote2 sensor with a PXA 271 ARM X-Scale processor (Crossbow 2007), used as a pilot platform. The estimation of the theoretical Variogram turned out as the most CPU-resource consuming task. The measurements showed clear advantages of the Nelder-Mead algorithm, which required only 32 seconds of CPU time, whereas the Grid-Search required more than 4 minutes. The calculation of a Kriging weighting matrix for 20 destination points was much faster with only 2 seconds of required CPU time.

Although the processor is occupied for a longer period of time during the initialisation of the Kriging interpolation, this has only a little effect on the average load because the initialisation is required only once per transport. Tasks that need to be executed once for each sampling interval can be more critical. But, the application of the weighting matrix took only 17.5 ms per sampling interval for a set of 20 sensors. The Kriging method is therefore fully suitable to run on embedded systems such as wireless sensor nodes.

The Kriging method can provide an estimation of the temperature in any point of space under the precondition that the spatial sensor density is high enough for interpolation (see section 6.1). The higher mathematical complexity of the Kriging method compared with simpler methods such as the Inverse Distance Weighting is rewarded with an improved accuracy, which is 20% better in average. Furthermore, the accuracy of the Kriging method depends on correct estimation of the Variogram model parameters. Best results were achieved either by the Nelder-Mead algorithm combined with the weighting according to Cressie or by setting the nugget by known sensor tolerances and the range by an average value of previous experiments.

## ACKNOWLEDGMENT

This research was supported by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 'Autonomous Cooperating Logistic Processes' and by the Federal Ministry of Education and Research, Germany, under reference number 01IA10001 ('The Intelligent Container'). We especially thank Rungis Express Germany, Maersk Copenhagen and Dole in Costa Rica and Germany for support in field tests and provision of recorded data. Further information about the project can be found at [www.intelligentcontainer.com](http://www.intelligentcontainer.com).



## REFERENCES

- Cressie, N., 1985. Fitting variogram models by weighted least squares. *Mathematical Geology* 17(5), 563-586. doi:10.1007/bf01032109
- Crossbow, 2007. IMote2 - High-performance Wireless Sensor Network Node. Product data sheet available at [http://wsn.cse.wustl.edu/images/e/e3/Imote2\\_Datasheet.pdf](http://wsn.cse.wustl.edu/images/e/e3/Imote2_Datasheet.pdf)
- D'Errico, J., 2005. Description of the 'fminsearchbnd' function. *Matlab Central - File Exchange*, <http://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd>.
- Jedermann, R. and Lang, W., 2009. The minimum number of sensors - Interpolation of spatial temperature profiles. *Wireless Sensor Networks*, 6th European Conference, EWSN 2009, *Lecture Notes in Computer Science* (LNCS), Springer, Berlin/Heidelberg. Doi: 10.1007/978-3-642-00224-3\_15
- Jedermann, R., Palafox-Albarrán, J., Jabbari, A. and Lang, W., 2011. Embedded intelligent objects in food logistics - Technical limits of local decision making. In: Hülsmann, M., Scholz-Reiter, B., Windt, K. (eds.) *Autonomous Cooperation and Control in Logistics*. 207-228, Springer, Berlin. Doi: 10.1007/978-3-642-19469-6\_16
- McBratney, A.B. and Webster, R., 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science* 37(4), 617-639.
- Nelder, J.A. and Mead, R., 1965. A Simplex Method for Function Minimization. *The Computer Journal* 7(4), 308-313. Doi:10.1093/comjnl/7.4.308
- Palafox-Albarrán, J., Jedermann, R. and Lang, W., 2011. Energy-Efficient Parameter Adaptation and Prediction Algorithms for the Estimation of Temperature Development inside a Food Container. In: Cetto, A.J., Ferrier, J.-L., Filipe, J. (eds.) *Lecture Notes in Electrical Engineering - Informatics in Control, Automation and Robotics*. 77-90. Springer, Berlin. Doi: 10.1007/978-3-642-19539-6\_5
- Rodríguez-Bermejo, J., Barreiro, P., Robla, J.I. and Ruiz-García, L., 2007. Thermal study of a transport container. *Journal of Food Engineering* 80(2), 517-527. doi:10.1016/j.jfoodeng.2006.06.010
- Schabenberger, O. and Gotway, C.A., 2005. Statistical methods for spatial data analysis. *Texts in statistical science series*. Chapman & Hall/CRC.
- Schwanghart, W., 2009. Description of the 'variogramfit' function. *Matlab Central - File Exchange* <http://www.mathworks.com/matlabcentral/fileexchange/25948>.
- Wackernagel, H., 2003. *Multivariate geostatistics: an introduction with applications*, 3rd, completely revised ed. Springer.

## AUTHORS BIOGRAPHIES

**Reiner Jedermann** finished his Diploma in Electrical Engineering 1990 at the University of Bremen. After two employments on embedded processing of audio signals, he became in 2004 a research associate in the Department of Electrical Engineering at the University of Bremen. He finished his Ph.D. thesis on automated systems for freight supervision in 2009. His current tasks comprise of the development of the sensor and data processing system for the intelligent container.

**Javier Palafox-Albarrán** received the B.Sc. degree in Electronic Systems from the "Toluca Institute of Technology", Mexico (ITT), in 2000. He finished his M.Sc. in Information and Automation in 2009 at the University of Bremen. Since October 2010 he is a PhD Student at the International Graduate School for Dynamics in Logistics in Bremen where he is researching the "Analysis and prediction of sensor and quality data in food transport supervision".

**Pilar Barreiro**, PHD Agricultural Engineer since 1995, University Lecturer from 2003 and full-Professor since 2010 in the Department of Rural Engineering ETSIA, Technical University of Madrid (UPM). She began her career in the field of post harvest sensing technologies (on-line and at-line). Actually, the research topics focus on the development and implementation of smart sensors, mechatronics and robotics in the framework of agricultural mechanization, an activity that combines with various active projects on educational innovation.

**Luis Ruiz-Garcia** studied Agriculture Engineering at the Polytechnic University of Madrid and received is Diploma in 2003. In 2004 he was one the winners of the UNACOMA vision award, a prestigious price about the future of agriculture. In 2008 he finished is PhD about developing monitoring systems for refrigerated fruit transports. Since that time his research is focused in RFID and WSN applied to agro-systems and cold chain control.

**José Ignacio Robla** studied Material Science at the Chemistry Faculty of the Complutense University of Madrid (1981) He finished his Ph.D. thesis on Occluded Gases in Die Castings in 1986. In 1990 he joined the Spanish Council for Scientific Research (CSIC). Actually is the head of the Sensors Laboratory of the National Center for Metallurgical Research (CENIM-CSIC).

**Walter Lang** studied physics at Munich University and received his Diploma in 1982 on Raman spectroscopy of crystals with low symmetry. His Ph.D. in engineering at Munich Technical University was on flame-induced vibrations. Since 2003 he is the head of the Institute for Microsensors, -actors and -systems at the University of Bremen. His research focus includes the manufacturing of miniaturized sensor components and the automated processing of sensor data.

# MODELING AND SIMULATION OF AN ASSEMBLY LINE: A NEW APPROACH FOR ASSIGNMENT AND OPTIMIZATION OF ACTIVITIES OF OPERATORS

Domenico Falcone, Alessandro Silvestri, Antonio Forcina, Antonio Pacitto

Department of Mechanics, Structures, Environment, University of Cassino, Via G. Di Biasio, 43, 03043, Cassino – Italy

[falcone@unicas.it](mailto:falcone@unicas.it), [silvestr@unicas.it](mailto:silvestr@unicas.it), [a.forcina@unicas.it](mailto:a.forcina@unicas.it), [a.pacitto@unicas.it](mailto:a.pacitto@unicas.it)

## ABSTRACT

In the paper we will suggest an approach based on modeling and simulation for the assignment of activities to the operators of a flexible assembly line. Thanks to the above approach, it will be possible to optimize the workload of each operator, in particular to reduce non-value and increase saturation. Through the adoption of a specific modeling logic, we can simulate lines with high flexibility of manufacturing and assembly, in various industrial sectors. Starting from the analysis of a real case and following an incremental approach, a simulation model has been realized, verified and validated, which allows to obtain useful information about: production mix; percentages of saturation; lead time and optimization of the assembly line.

Keywords: non-value added, saturation, lead time, productive capacity

## 1. INTRODUCTION

Thanks to the simulation of reality, it is possible to predict events, under specific conditions of using. Simulation is a powerful experimental tool that uses the possibilities offered by computer calculation. The "conceptual model" of the real system, is translated in a logical-mathematical procedure, permitting to understand how the system works (Law, and Kelton 1999).

The present paper proposes the development of a virtual model simulating an assembly line. A simulation model is created, verified and validated, in order to obtain useful information about:

- productive capacity;
- partial and total lead time;
- percentage of saturation, Value Added (VA) and Non-Value Added (NVA), for each worker and station;
- line optimization.

The virtual model is the evolution of a study carried out previously, considering few stations of the same line (Falcone and others 2010) (Falcone and De Felice 1998). During the development phase, an important goal has been pursued: the possibility to use and change

the model without knowing the creation logic. It is possible to change input parameters through a database, which contains details about operations. Furthermore, a timetable was created with the times of the various operations along the line, which feeds the simulation model. Through the connection between the database and the physical model and the introduction of constraints assigned to the various operations, it is possible the construction of a simulation model able to reproduce the real productive line. The goals are the identification and optimization of the line in terms of minimization of NVA and no-saturation of operators.

## 2. STUDY OF THE OPERATIVE TIMES

The production mix consists of two models of the same product, realized along an assembly line, with many workstations and a big number of manual operations, performed on the front, back, the right and the left side of the product. If the operations are performed on one or both sides, in the stations we will find, respectively, one or two operators; if we provide also the operations on the front or back of the product, there will be another operator. Each product is labeled with information about the components required and the correct assembly sequence. For both models there is a number of optional components that can be installed on the product, so we have more configurations of the same model. For each operation the total time assigned is characterized by five aliquots: two times for the activities adding value to the product (product transformation and quality control) and three aliquots that do not add value to the product (supplying or picking, walking and similar activities).

The above subdivision allows evaluating VA and NVA for each operator and workstation. During the fixed advancing time of the line, linked to the takt time, the operators must perform all the operations assigned to him in his workstation. The difference between the total cycle time and the whole working time, allows evaluating no-saturation of operators.

## 3. VALUE ADDED (VA) AND NON-VALUE ADDED (NVA)

In economics, the value added (also abbreviated VA), or capital gains, is the measure of the increased value resulting in the production and distribution of goods and

services, thanks to factors of production: capital and labor. The difference between the value of goods and services produced and the value of goods and services purchased for the production process is the value added.

Each company has adopted strategies to maximize VA and minimize non-value added (also abbreviated NVA), which represents the decrease in value due to the time spent to perform all the activities that the customer does not see and therefore is not willing to pay.

In order to reduce NVA, the process manager has to ensure saturation of the operators, splitting operations to the operators. In practice, he allocates the operations in a single workstation, taking into account the total number of activities of the work shift, according to the production mix (Falcone and De Felice 2007). The aim is to combine tasks, in order to achieve the total number of minutes of work in one shift. This aim is simple, but not obvious, in fact almost no one can ever saturate a worker for the exact duration of an entire shift, but only for a shorter time.

The difference between the work shift time and the result reached is the no-saturation: the minutes paid to the operator, while he doesn't "work" because he has nothing to do and obviously represents a loss.

The manufacturing cycles include information about operation allocated in different workstations, with an indication on how much of the total time is made of real assembly or manufacturing (value-added activities) and how much of supplying, moving, picking and so on (non-value added activities) (Baracchino 1989) (Ciappei 1988). Using these data it is possible to analyze NVA activities, searching for possible improvements. The approach adopted so far is mainly related to the experience of the process manager and requires to start over according to any change of production. For these reasons it's important to look for tools and procedures able to accelerate and objectify all the actions of optimization described.

#### 4. VIRTUAL MODEL

The present work takes into consideration a previous work of the same authors. The model was previously realized for only two workstations, subsequently extended to the whole production line, with the same programming logic. The old simulation model is shown in the figure n. 1. The model developed initially, allowed a manual optimization, useful in the case of a little number of stations. Instead, in the case of a big number of stages, an improvement of the model is required.



Figure 1: Old simulation model view

An important aspect of the programming logic is the separation between the simulation model and the input data-base, more complex the first and very simple the second. In fact it is possible to change input data simply acting on an Excel table, shown in the figure n. 2, even without any simulation knowledge.

CADENZA 1.06		TEMPI CICLO				
	OPERAZIONE	TRASFORMAZIONE [min]	QUALITA' [min]	VARIE [min]	RIFORNIMENTO [min]	WALK [min]
STAZIONE 7 LATO DX	292	0.09	0.00	0.00	0.02	0.01
	26010	0.04	0.00	0.00	0.01	0.00
	7144	0.00	0.08	0.00	0.02	0.03
	7187	0.09	0.00	0.02	0.04	0.01
	7407	0.06	0.00	0.00	0.00	0.00
	7466	0.11	0.00	0.01	0.00	0.01
	7488	0.09	0.00	0.00	0.00	0.00
	7490	0.18	0.00	0.00	0.00	0.01
	7160	0.00	0.13	0.04	0.03	0.01
	7172	0.15	0.00	0.00	0.00	0.01
STAZIONE 7 LATO SX	294	0.04	0.00	0.00	0.01	0.00
	520	0.10	0.00	0.00	0.02	0.01
	26020	0.04	0.00	0.00	0.01	0.00
	7122	0.09	0.00	0.00	0.02	0.01
	7124	0.03	0.00	0.00	0.02	0.00
	7146	0.00	0.08	0.00	0.02	0.03
	7204	0.06	0.00	0.00	0.02	0.00
	28020	0.05	0.00	0.00	0.00	0.00
	7444	0.03	0.00	0.05	0.00	0.01
	7470	0.11	0.00	0.01	0.00	0.01
ST. 8	7494	0.09	0.00	0.00	0.00	0.01
	7162	0.00	0.13	0.04	0.04	0.01
	21564	0.00	0.00	0.00	0.05	0.04
	21738	0.00	0.00	0.00	0.04	0.17
	5032	0.05	0.00	0.00	0.03	0.02
	7092	0.35	0.00	0.00	0.12	0.06

Figure 2: Time Table of Operations

But, if we need to modify operations position along the line, in order to consider different constraints, a simulation model redefinition is necessary. Therefore, the model improvement requires the creation of a single modular workstation, in which a flexible number of activities is included, using many "labor machine" icons, at the beginning not initialized. Subsequently, the whole assembly line can be modeled through the duplication of standard workstations, opportunely linked together (Benjaafar 1992). In each workstation, the particular operation icon will be activated or not by the dynamic input database, including all the operations of all product models, characterized by different combinations of optional components (Fig. 3).

A particular codified label related to the specific product under production, allows understanding which records of the dynamic database have to be considered, determining the path of the product in the simulation model.



Figure 3: Extended model

Therefore the programming logic is the following: the product label defines which operations are required, subsequently some records of the database are identified and finally some labor machines of the model activated, able to read input data directly from the data base (Fig. 4).

In this way, when we have a new product, we don't need to modify the virtual model but we can act outside the model and inside the database, in which there are all production times.

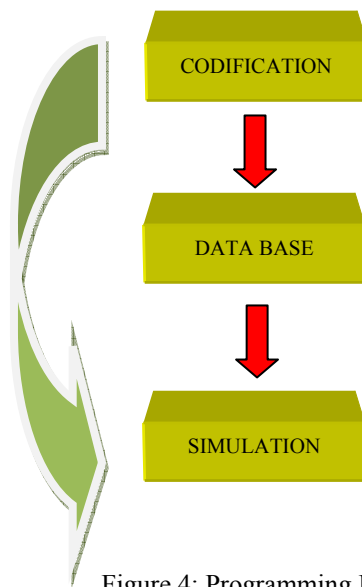


Figure 4: Programming Logic

Each record of the database includes the code of a particular operation and the relative five time aliquots.

## 5. OPTIMIZATION PHASE

At this point the optimization phase starts. In order to increase the saturation of each operator and station, the process manager has to assign operations to the available operators in different stations, according to the production order. Thanks to the improvement of the model, it is possible an automatic assignment using software tools acting directly on the simulation model rather than on the database.

The optimization step has to consider many variable constraints, in particular:

- Technological constraints: they are related to the nature of the technological cycle, conditioning the correct sequence of operations.

- Ergonomic constraints: they are related to the ergonomics of each workstation. Some operations can be performed only in particular stations.
- Equipment constraints: they are related to the resources of each station. Some operations may require special equipments.
- Constraints of space: Some tasks require the physical space for operators to perform their duties, according to ergonomic constraints too.

In order to respect the above constraints, some records in the database, are linked to "fixed" positions in the model. Subsequently the optimizer can't move the corresponding operations, in order to balance the assembly line and reach particular goals. Figure 5 shows the general structure of the model.

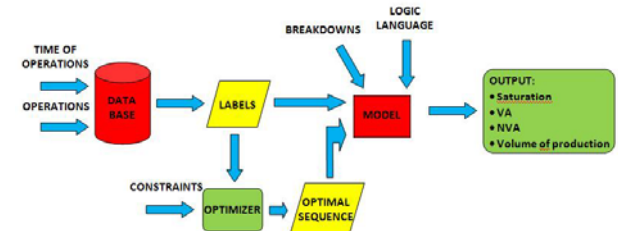


Figure 5: General structure of the model

Another important change of the model, has been the introduction of downtimes, always present in the case of manual or automatic assembly lines. Basing on the study of the real failure rates, it is possible to enter configuration data in each labor machine of the simulation model (figure 6).

General   Setup   Breakdowns   Shift   Actions   Costing   Reporting   Notes									
	Description	Check Only At Start Of Cycle	Breakdown Mode			Breakdown Duration			
			Mode	No. of Operations	Time Between Failures	Actions on Down	Labor Rule	Repair Time	Act Rt
1	Fault belts	<input checked="" type="checkbox"/>	No. of Op	Undefined		N	N	0.02	N

Figure 6: Table of downtimes

The last extension of the model previously made, has been the introduction of probabilistic times. In fact, an operator is not able to perform the same task always in the same time (Zhou and Venkatesh 1999). So, deterministic times have been replaced by probabilistic

ones, characterized by a particular probability distribution.

All the modifications proposed, permits to obtain a more realistic model of the studied flexible line (Raouf and Ben Daya 1995).

## 6. MODEL VERIFICATION AND VALIDATION

Many verification and validation activities occurred (Tocher 1967). First of all, we verified the codification labeling procedure, in order to identify in the whole timesheet, all assembly operations of a particular product.

Then, the production quantity both in the case of deterministic and probabilistic times, with and without breakdowns, has been compared with real data. During 450 production minutes for each working shift, with a production frequency of 1.06 minutes, 425 products are realized, according to the real production and its standard deviation.

We have also verified the impact of changes of times on the operator saturation. See the following example.



Figure 7: Saturation data before change



Figure 8: Saturation data after change

Finally, after the production sequence optimization, we have verified that the allocated activities were realized correctly on the product in the right station by the right operator, testing the model flexibility.

## 7. CONCLUSIONS

This work concerns the development of a simulation model of an assembly line. The programming logic adopted shows characteristics of modularity and generality which could be extended to different production systems (Rumbaugh and others 2005). By using commercial simulation software, a virtual model has been created and gradually improved, making it closer to reality, as well as more complex. The end result reached is a flexible virtual model that can simulate and optimize the operating conditions of the production line examined. The model allows seeing the

number and types of products made in a specific period of time, the percentages of VA, NVA and the saturation rate of each operator or workstation, rather than the whole line. An external data base was realized, usable by anyone, even inexperienced against the simulation software. It is possible to modify the processing times and consequently the location of each activity along the line, simply clicking on the value and changing the variable initialization. This function allows you to obtain an increase in saturation of the operators, taking into account the process constraints, technological, ergonomic and others. So, you can see the impact of any change introduced, through an extensive reporting on working and crossing times, VA and NVA activities, saturation of operators and so on. Finally, the introduction of an external database, collecting data of all operations, permits to manage the production flow easily. Through the use of an optimizer, the operations are moved from one station to another, in order to achieve the fixed goals. For example, you could optimize the production flow in order to saturate the operators as much as possible; or you could minimize any time spent in activities that do not add value to the product. The general approach followed and simulation model realized could be friendly applied in different sectors and industrial contexts (Dallari and Marchet 2003).

## REFERENCES

- Baracchino, N., 1989. *Approvvigionamenti. Quali-Come-Perché*. Milano: Franco Angeli.
- Benjaafar, S., 1992, *Intelligent simulation for flexible manufacturing system*. School of Industrial Engineering, Purdue University, West Lafayette (USA).
- Ciappei, C., 1988. *La funzione logistica nell'impresa*. Torino: Giapichelli.
- Dallari, F., Marchet, G., 2003. *Rinnovare la Supply Chain*. "Il sole 24 ore".
- Falcone, D., De Felice, F., 2007. *Progettazione e gestione degli impianti industriali*. Milano: Hoepli.
- Falcone, D., De Felice, F., 1998. Operating planning of production in a mechanical industry through simulation techniques. *Proceedings of ESM98-12th European Simulation Multiconference*. Manchester (UK).
- Falcone, D., Silvestri, A., Di Bona, G., Pacitto, A., Forcina, A., 2010. Study and modelling of very flexible lines through simulation. *Spring Simulation Multiconference. Proceedings of EAI2010*. Orlando (USA).
- Ferrozzi, C., Shapiro, R.D., Heskett, J.L., 1987. *Logistica & Strategia*. Torino: ISEDI.
- Law, M., Kelton, D., 1999. *Simulation Modeling and Analysis*. McGraw-Hill.
- Raouf, A., Ben Daya, M., 1995. *Flexible manufacturing system*. Amsterdam: Elsevier.
- Rumbaugh, J., Jacobson, I., Booch, G., 2005. *Unified modeling language reference manual*. USA: Addison-Wesley Professional.



Tocher, K.D., 1967. *The art of simulation*. England: English Universities press.

Zhou, M., Venkatesh, K., 1999. *Modeling, simulation and Control of Flexible Manufacturing Systems*. Singapore: World Scientific Publishing Co Pte Ltd.

#### **AUTHORS' BIOGRAPHIES**

**Domenico Falcone** received his degree in Mechanical Engineering from the University of Naples, Italy, in 1984. Currently, he is Professor of Industrial Plants at University of Cassino, Italy. His research interest concerns mainly industrial plants analysis and economics, quality management and industrial safety. He is member of AICQ.

**Alessandro Silvestri** received his degree in Mechanical Engineering from the University of Cassino, Italy, in 1998. Currently, he conducts research activities at Department of Mechanics, Structures, Environment as Assistant Professor and researcher. His research interest concerns mainly supply chain, industrial plants analysis and economics.

**Antonio Pacitto**, PHD Student at University of Cassino. He received his degree in Mechanical Engineering from University of Cassino, in 2009. Currently, he conducts research activities at Department of Mechanics, Structures, Environment mainly on simulation of production processes and balancing of production line.

**Antonio Forcina** received his degree in Mechanical Engineering from the University of Cassino, Italy, in 2006. He received his Ph.D. in mechanical engineering in the same University in 2010. Actually, he conducts research activities at Department of Mechanics, Structures, on numerous issues, including logistics, industrial plants analysis and RAMS analysis.



# DYNAMIC SPECTRUM MANAGEMENT WITH MINIMIZING POWER ALLOCATION

Tawiwat Veeraklaew<sup>(a)</sup>, Settapong Malisuwan<sup>(b)</sup>

<sup>(a)</sup>Laboratory of Mechanical and Aeronautic Engineerings  
Defence Technology Institute, THAILAND

<sup>(b)</sup>Dynamic Spectrum Management Project  
National Broadcasting and Telecommunications Commission, THAILAND

<sup>(a)</sup>tawiwat.v@dti.or.th, <sup>(b)</sup>settapong\_m@hotmail.com

## ABSTRACT

This paper deals with an algorithm used to manage Dynamic Spectrum (Dynamic Spectrum Management, DSM) called nonlinear programming. Now a day, DSM problem is an open problem that many scientists and researchers try to develop such an algorithm to manage the frequency ranging from 3 KHz to 300 GHz called Radio Spectrum. Radio Spectrum is quite an important range of frequencies as this range may be used for wireless communication. In this paper, the problem is proposed as minimizing power allocation since a power allocation is quite important in the competitive market. Problem is set as a nonlinear programming with example to demonstrate the algorithm as well.

Keywords: dynamic spectrum, power allocation, minimizing

## 1. INTRODUCTION

Dynamic spectrum management (DSM) is a very famous technology powerfully shares the spectrum among the users in the communication system. The cross-talk interference can be reduced in the digital subscriber line (DSL) while using DSM [3]-[5]. One of the promising candidates for multiple accesses in cognitive radio is also DSM [6]. There are many users coexist in a channel. This causes co-channel interference and the goal of DSM is to manage the power allocations in all channels in order to maximize the sum of the data rates of all users subject to any constraint such as power [3]. It is quite obvious that this problem is non-convex and can not be solved efficiently in polynomial time [5]. The game-theoretic formulation has been used in variety of contexts including wireless and DSL [7] since in this formulation, each user maximizes her data rate, the Shannon utility function, while given the knowledge of the other user's power allocations. Under this condition the Nash equilibrium exists and is also unique; therefore, the beauty of this formulation that the problem can be solved efficiently since it is convex. In

Nash equilibrium, users tend to compete for good channels regardless the interference that might cause to the others and it is known as tragedy of the common from economics [8]. A competitive equilibrium (CE) of a market model is a set of prices and the corresponding power allocations which maximizes all user's utility and clear the market by making the total power allocated meet the spectral mask. Although the CE has become a famous problem in computer science, the application to resource communication management system is still rare. The continuation of CE for DSM was proven in [1] and also [2] show that CE achieves better total transmission rate than the Nash equilibrium. This is solved with properly assigned budgets in order to guarantee equality among all users. However, the efficient CE prices is an open problem.

[1] and [2] show that the market competitive equilibrium (CE) has become better system performance in term of data transmission rate (by reducing cross talk) compare to the Nash equilibrium (NE). [9] has shown that the CE is the solution of a linear complementarity problem (LCP) and can be computed efficiently. The reason behind this is the condition that is posted to the problem that users of a channel experience the same noise levels and the cross-talk effects between users are low-rank and weak.

This paper focuses on determining the user's budget by assuming it as power function in each channel. We show the algorithm for using as a simple model in order to demonstrate the idea on how to minimize the user's budget by using nonlinear programming.

## 2. PROBLEM STATEMENT

Generally, a communication system consists of  $n$  users and  $m$  channels. At the same time, multiple users may use the same channel. This will cause the interference to each other. Assuming the power that allocated by user  $i$  to channel  $j$  is  $x_{ij} \geq 0$  and the total

power allocated by all users in the  $j^{th}$  channel is bounded by the spectrum mask  $c_j$ ,  $\sum_{i=1}^n x_{ij} \leq c_j$ .

In cognitive radio, the interference experienced by the primary user due to transmissions from secondary user must be limited. For this reason, the power allocations should be scaled so that  $x_{ij}$  represents the power received by the primary user on channel  $j$  from user  $i$ . In order to achieve an efficient allocation of spectrum, we associate a price  $p_j > 0$  with each channel  $j$ . Therefore, for a given vector of prices,  $\bar{p} = [p_1, p_2, \dots, p_m]^T$  associated with user's powers, the user's budgets will also be associated with their power as  $w_j > 0$ .

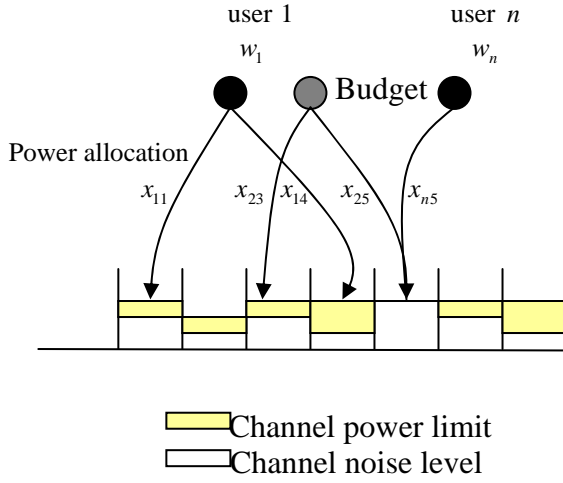


Fig. 1 Competitive spectrum market model

From above mention, the problem statement can be set as finding  $x_i \geq 0$  such that

$$\min J = \sum_{i=1}^n w_i^2 \quad (1)$$

subject to

$$p_j + n_j \sum_{i=1}^{n_j} x_i \leq w_j, \quad j = 1, \dots, m \quad (2)$$

$$x_i > 0, \quad i = 1, \dots, n \quad (3)$$

where  $\bar{n} = [n_1, n_2, \dots, n_m]^T$  and  $\sum_{j=1}^m n_j = n$ .

Equations (1), (2), and (3) can be solved together as set of algebraic equation and inequality equations,

respectively. This is known as the form of nonlinear programming problem with sets of inequality constraints.

### 3. NONLINEAR PROGRAMMING

As an introduction to the problem statement, this section will focus on the use of nonlinear programming (NLP) problem. The NLP problem requires finding a finite number of unknown variables such that an objective function or performance index is optimized without violating any set of equality and inequality constraints. The NLP problem is quite often referred to as parameter optimization. It is quite known that special cases of the NLP problem include linear programming (LP), quadratic programming (QP), and least squares problems [10].

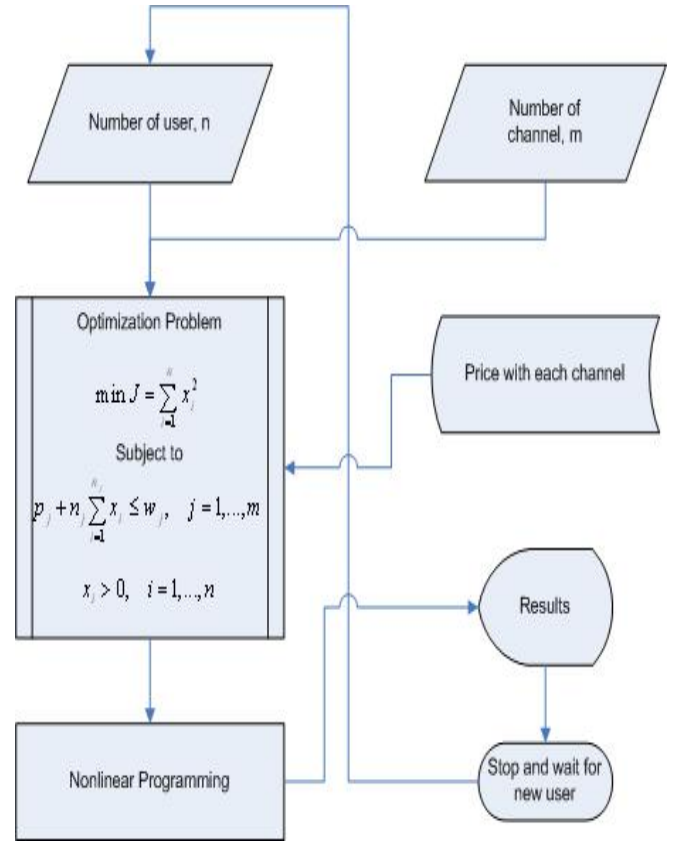


Fig. 2 A flowchart for DSM with minimizing user's budget

The general inequality constrained nonlinear programming problem can be stated as finding the  $n$ -vector  $x^T = (x_1, x_2, \dots, x_n)$  to minimize the scalar objective function or performance index

$$F(x) \quad (4)$$

subject to the  $m$  constraints

$$c_L \leq c(x) \leq c_U \quad (5)$$

and the simple bounds

$$x_L \leq x \leq x_U. \quad (6)$$

However, the equality constraints can be included by setting  $c_L = c_U$ .

The necessary conditions for optimal values  $x^*$  require that

$x^*$  must be feasible and satisfied by Eq.(5) and (6);

the Lagrange multipliers corresponding to Eq.(5) and (6) satisfy

$$g = G^T \lambda + \nu. \quad (7)$$

the Lagrange multipliers for the inequality constraints be

nonpositive for active upper bounds,  
zero for strictly satisfied constraints,  
nonnegative for active lower bounds.

Therefore, the optimization problem statement and nonlinear programming can be put together as a flowchart shown in Fig. 2.

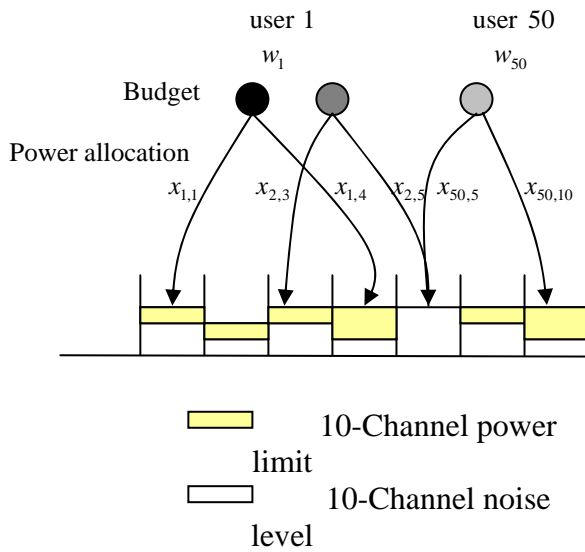


Fig. 3 Competitive spectrum market model

#### 4. EXAMPLE

The procedure outlined in this paper for DSM with minimizing user's budget can be demonstrated by this example. Assuming that the number of users  $n=50$  and the number of channels  $m=10$ . Also, in this paper, we set the price of each channel as

$$p_j = 2^{n_j}, \quad j = 1, \dots, 10 \quad (7)$$

for regulatory reason.

According to the Fig. 3 and for simple demonstration, the power that allocated by users has 5 levels as 5, 10, 15, 20 and 25. According to 5 levels of the power that allocated above, the corresponding budgets are 32, 64, 128, 256, and 512, respectively. Therefore, Eq. (1), (2), and (3) can be rewritten as

$$\min J = \sum_{i=1}^{50} w_i^2 \quad (8)$$

subject to

$$p_j + n_j \sum_{i=1}^{n_j} x_i \leq w_j, \quad j = 1, \dots, 10 \quad (9)$$

$$x_i > 0, \quad i = 1, \dots, 50 \quad (10)$$

and choosing number user in each channel equally for simplicity.

Due to the inequality constraints in Eq. (9), the results from nonlinear programming show that the total user's budget from 50 users is equal to 11,360.

For the next step when some of new users is connected to these channels, all step of the computation must be recomputed again as shown in Fig. 2.

#### 5. CONCLUSION

From the solution of the example above, the new algorithm in this paper is can be implement to the Dynamic Spectrum Management in order to minimize power allocation for the competitive market model. As known that optimization theory can be applied to many subjects and applications, there will be many goals to be achieved along this model for the future work.

#### ACKNOWLEDGMENTS

This work was supported in part by the National Broadcasting and Telecommunications Commission, Bangkok, Thailand. The financial support is gratefully acknowledged.

#### REFERENCES

- Ye, Y., 2008. Competitive communication spectrum economy and equilibrium. *Working paper*. Available: <http://www.stanford.edu/~yyye/spectrumpricing2.pdf>
- Ling, M., Tsai, J., and Ye, Y., 2009. Budget allocation in a competitive communication spectrum economy. *EURASIP Journal on Advances in Signal Processing*, Available: [http://www.stanford.edu/~yyye/spectrumbudget7\\_eurasip1.pdf](http://www.stanford.edu/~yyye/spectrumbudget7_eurasip1.pdf)
- Yamashita, N. and Luo, Z.-Q., 2004. A Nonlinear Complementarity Approach to Multiuser Control for Digital Subscriber Lines. *Optimization Methods and Software*, no. 19, pp. 633-652.
- Luo, Z.-Q. and Shang, J., 2006. Analysis of Iterative Waterfilling algorithm for Multiuser Power Control in Digital Subscriber Lines. *EURASIP*

*Journal on Applied Signal Processing*, pp. 1-10, Article ID: 24012.

Luo, Z.-Q. and Zhang, S., 2008. Dynamic Spectrum Management: Complexity and Duality *IEEE Journal of Selected Topics in Signal Processing*, Vol. 2, no. 1, pp. 57-73.

Niyato, D. and Hossian, E., 2008. Microeconomic Models for Dynamic Spectrum Management in Cognitive Radio Networks. *Cognitive Wireless Communication Networks*. US: Springer.

Chiang, M., Han, P., Lan, T., and Tan, C., 2008. Power Control in Wireless Cellular Networks, ser. *Foundation and Trends in Networking Sample*. Now Publishers Inc.

Samuelson, P., 1983. Foundations of Economic Analysis. *Harvard University Press*.

Xie, Y., Armbruster, B., and Ye, Y., 2009. Dynamic Spectrum Management with the Competitive Market Model. *Working paper*. Available: <http://www.stanford.edu/~yyye/spectrumpricing2.pdf>

Betts, J. T. Betts, 2001. Practical Methods for Optimal Control Using Nonlinear Programming. *SIAM*.

## AUTHORS BIOGRAPHY



### **Colonel Tawiwat Veeraklaew**

received the Ph.D. degree in mechanical engineering from University of Delaware, Newark, DE, USA in 2000. He is a Platform and Material Senior Researcher at Defence

Technology Institute (Public Organization), Bangkok, Thailand and supervise the Platform and Material Laboratory. He has published more than 30 both in conference and journal articles. His current research interests are in the area of controlled mechanical systems, dynamic optimization and special software hardware design.



### **Colonel Settapong Malisuwan**

received the Ph.D. degree in Electrical Engineering from Florida Atlantic University, Boca Raton, FL, USA in 2000. He is a member of Spectrum Refarming Committee, National

Telecommunications Commission, Thailand. He has published more than 70 both in conference and journal articles. His current research interests are in the area of Mobile Wireless Communication Systems, Electromagnetic Interference and Compatibility (EMI/EMC) and Spectrum Management.

# LARGE TARGET-WEAPON ASSIGNMENT PROBLEMS WITH EVOLUTIONARY PROGRAMMING AS APPLIED TO THE ROYAL THAI ARMED FORCES GROUND-TO-GROUND ROCKET SIMULATION SYSTEM

Settapong Malisuwan<sup>1,2</sup>, Teeranan Nandhakwang<sup>2</sup> and Tawiwat Veeraklaew<sup>1</sup>

<sup>1</sup>Laboratory of Mechanical and Aeronautic Engineerings

Defence Technology Institute, THAILAND

<sup>2</sup>Dynamic Spectrum Management Project

National Broadcasting and Telecommunications Commission, THAILAND

settapong\_m@hotmail.com

teeranan@rtarf.mi.th

Tawiwat.v@dti.or.th

## ABSTRACT

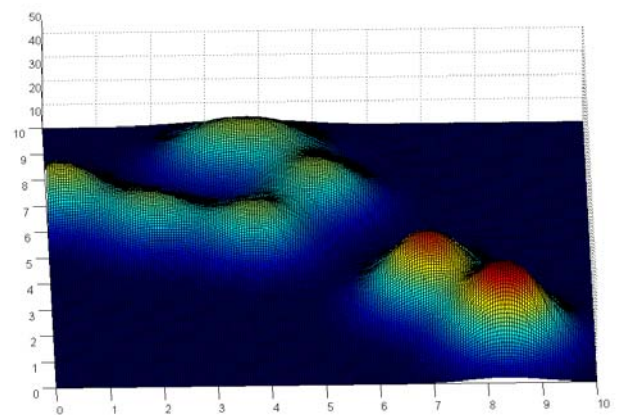
An algorithm based on evolutionary programming (EP) is developed and presented for large numbers of target-weapon assignment. An optimal assignment scheduling in one, which allocates target to weapon such that the total expected of target surviving the defense, is minimized. The proposed method improves EP with reordered mutation operator to handle a large-scale assignment problem. The main advantage of this approach is that the computation time can be controlled via tradeoff performance between the computation time and target surviving value.

**Keywords:** Assignment problem, Large-scale problems, Evolutionary computation, Evolutionary programming, Reordered mutation, Target-weapon assignment problem

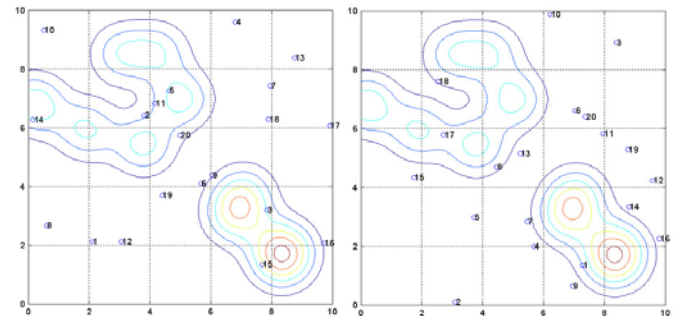
## 1. INTRODUCTION

Considering a battlefield scenario, Figure 1.1 illustrates terrain of the area of operation (AO). Figure 1.2, which maps the 3D terrain into 2D, illustrates location of threats that need to be attacked in AO. Each threat can move to different locations at any time, and thus places different demands weapon allocations. Therefore, to maximize the probability of threat or target surviving value, it is desirable to assign weapons at optimum engagement opportunities.

In military operations, problems in planning and scheduling often require feasible and near to optimal solutions with limited computing resources and within very short time periods. To overcome this time dilemma, fast-execution weapon-target (WT) algorithms are needed. This paper proposes evolutionary programming (EP) based algorithm to solve large scale weapon-target assignment problems with reordered mutation operator. The main advantage of this approach is that computation time can be controlled via tradeoff performance between computation time and target surviving value.



**Figure 1.1** An example of terrain of the area of operation (AO).



**Figure 1.2** Threats location in the area of operation (AO) for 2 time intervals.

The remainder of this paper is organized as follow: Section 2 presents a brief description of WT problems in general; section 3 describes an evolutionary programming overview; section 4 explains the proposed algorithm in greater detail; section 5 provides numerical results; and section 6 summarizes the experiment.

## 2. The WT Problem

The general weapon-target assignment problem deals with allocation of weapons to offensive targets or threats. One objective is to minimize the surviving value of targeted assets [1, 2]. An asset is an any entity (or collection of entities) of the important military objects, e.g. tactical command posts, bridges, missiles, battleships, etc. Each asset has a value assigned to it. A threat can be anything potentially causing damage to defensive assets. Any threats may be come from different position, speed, distance or all off them. A weapon is an object to eliminate a threat, e.g. missile, artillery, close air support, which is assumed to have an inventory of shots [2]. Difference weapons may require different amounts of time to engage the same threat.

The number of threats  $N_T$ , the number of weapons  $N_W$ , and the number of assets  $N_A$  may be dependent upon time  $t$ . Any weapon can kill any threat in range with a probability of kill and fire time.

The problem is to minimize the expected threat by determining the best firing schedule over time, given  $N_w(t)$ , weapons to defend  $N_A(t)$  assets against  $N_T(t)$  threat, where  $N_w, N_A, N_T \in \mathbb{Z}^+$  [2]. The size of this problem may be very large.

To formulate the problem, let  $a_{ijk}$  represent the assignment of weapon  $i$  to threat  $j$  on the time period engagement interval  $k$ . The  $a_{ijk}$  are the decision variables, and have strictly nonnegative integer value:  $a_{ijk} \in \mathbb{Z}^+$ . Let  $p_{ijk} \in [0,1]$  denote the probability that weapon  $i$  fired during the engagement interval  $k$  and kills threats  $j$ . For scheduling purposes the time between some initial firing time and a time horizon  $T$  may be divided into a number,  $P$ , of discrete temporal intervals  $t_1, t_2, \dots, t_P$  of equal length  $t$ , with  $T = P \cdot t$ . The objective is to find the maximization of the total expected surviving value of the asset under attack, or alternatively the minimization of the total threat [1]:

$$S(A^*) = \max_{ijk} \sum_{k=1}^P \sum_{j=1}^{N_T} \sum_{i=1}^{N_W} p_{ijk} a_{ijk} \quad (2.1)$$

subject to  $\sum_{k=1}^P \sum_{j=1}^{N_T} \sum_{i=1}^{N_W} a_{ijk} \leq M$  upper bound on total number of shorts fired,

$\sum_{i=1}^{N_W} a_{ijk} \leq 1$  at most one shot at each threat  $j$  in any interval

$\sum_{i=1}^{N_T} a_{ijk} \leq 1$  at most one shot by each weapon in any interval,

$a_{ijk} \in \{0,1\}$  a weapon is either active or not during interval, where  $S(A^*)$  represents the surviving value, corresponding to firing plan  $A = \sum_{k=1}^P A_k$ , and  $A^*$  denote the optimal firing schedule.

## 3. Evolutionary Programming [3, 4, 5]

Evolutionary programming (EP) is a stochastic optimization strategy, which is similar to genetic algorithms. In 1961, L. J. Fogel proposed applying the concepts of natural evolution, selection, and stochastic mutation to achieve intelligent behavior through simulated evolutions. Originally, EP was developed to evolve finite-state machines. The basic EP algorithms use normally distributed mutations to modify real-valued vectors and emphasize mutation as essential operators for searching in the search space.

EP starts its search with a set of randomly initialized population in a given bound space. Thereafter, each individual of a created population is altered through application with a mutation operator, and produce the new individual. Then, each individual is evaluated to calculate its fitness function. Frequently, the fitness function can also be the same as the objective function. An EP algorithm is formulated as follows [3]:

### Algorithm 1 (EP)

```

00  $\Phi := \text{Problem Formulation}(\text{weapons, targets, costs})$ 
01  $t := 0$ 
02 initialize  $P(0) := \{a'_1(0), a'_2(0), \dots, a'_\mu(0)\}$ 
03 evaluate  $P(0) := \{\Phi(a'_1(0)), \Phi(a'_2(0)), \dots, \Phi(a'_\mu(0))\}$ 
04 while not terminate do
05 {
06   mutate  $P(t) := m \odot_m (P(t))$ 
07   evaluate  $P'(t) := \{\Phi(a'_1(t)), \Phi(a'_2(t)), \dots, \Phi(a'_\mu(t))\}$ 
08   select  $P(t) := \max\{s \odot_s (P'(t) \cup Q)\}$ 
09    $t := t + 1$ 
10 }
11 return  $\max\{P(t)\}$ 

```



where  $a'$  is an individual member in the population,  $\mu \geq 1$  is the size of the parent population,  $P(0) := \{a'_1(t), a'_2(t), \dots, a'_\mu(t)\}$  is the population at time  $t$ ,  $\Phi: I \rightarrow \mathbb{R}$  is the fitness mapping,  $m\Theta_m$  is the mutation operator with parameters  $\Theta_m$ ,  $s\Theta_s$  is the selection operator, and  $Q \in \{0, P(t)\}$  is a set of individuals additionally accounted for the selection step.

#### 4. Proposed Method

We present an Evolutionary Programming approach to solve target-weapon assignment problems that uses reordered technique to mutate solutions. The use of reordered mutation and is useful for problems that require permutations of the integers and for which traditional mutation presents feasibility problems.

The objective is to find a maximum total expected surviving value or a minimum total threat, which is the summation of the maximum surviving to weapon  $i$  to threat  $j$  to on the time period engagement interval  $k$ . A number of weapon and threat is randomly generated and established in term of Equation (2.1).

##### 4.1. Representation

Although traditional EAs use a bitstring representation, EAs are not restricted to bitstring [3, 4, 5, 6]. Real-valued vectors are also used as a representation. In permutation problems such as the assignment problem, the feasible solutions are usually an order-based number representing a permutation. For example,  $\{4, 1, 3, 2, 3, 4, 2, 1\}$  is a possible solution, which is represented weapon 4 is assigned to threat 1 at engagement period 1, weapon 1 is assigned to threat 2 at engagement period 1, weapon 3 is assigned to threat 3 at engagement period 1, weapon 2 is assigned to threat 4 at engagement period 1, weapon 3 is assigned to threat 1 at engagement period 2, weapon 4 is assigned to threat 2 at engagement period 2, weapon 2 is assigned to threat 3 at engagement period 2, and weapon 1 is assigned to threat 4 at engagement period 2.

##### 4.2. Initial Population

The initialization process is created using a population of individuals, where each individual is set of an ordered number from  $1, \dots, n$ . All individuals are generated randomly with Uniformly Distribution.

##### 4.3. Selection Operator

A selection is a major operator used in evolutionary algorithms which emphasizes the better solution in a population. Indeed, after it selects good solutions to be parents, the remainder solutions are

deleted from the population. Afterward, the best individuals from the combination pool of parents will create offspring by mutation operator.

##### 4.3. Objective and Fitness Function

A modification of the Equation (2.1) is chosen as the fitness function. The fitness function  $\Xi$  is expressed as

$$\Xi(a_m) = \sum_k^P \sum_i^{N_W} x_{k,i,per(i)} \quad (4.1)$$

where  $N_W$  is the number of weapons,  $P$  is a total time interval,  $a_m$  is an individual  $m$ th in a population,  $x_{kij}$  is a survival value of weapon  $i$  assigned to threat  $j$  at engagement period  $k$ ,  $k$  and  $i$  is a set of ordered integer form  $[1, n]$ , and  $per(i)$  is a set of permutation  $i$ . For example, let payoff array represent in Table 4.1 as follow:

**Table 4.1** Payoff array for survival values of each weapon engages to each threat

Survival Value	Engagement Period 1				Engagement Period 2			
	Threat 1	Threat 2	Threat 3	Threat 4	Threat 1	Threat 2	Threat 3	Threat 4
Weapon 1	0.35	0.48	0.72	0.66	0.85	0.19	0.56	0.15
Weapon 2	0.82	0.25	0.64	0.44	0.66	0.56	0.42	0.53
Weapon 3	0.35	0.83	0.92	0.16	0.77	0.90	0.35	0.22
Weapon 4	0.88	0.52	0.77	0.34	0.35	0.75	0.84	0.75

For example, individual  $a_1$  has parameters  $k = 1$ ,  $i = \{1, 2, 3, 4\}$ ,  $j = \{4, 2, 1, 3\}$ , and  $k = 2$ ,  $i = \{1, 2, 3, 4\}$ ,  $j = \{2, 1, 4, 3\}$ . Therefore,  $\Xi(a_1)$  can be evaluated by  $x_{114} + x_{122} + x_{131} + x_{143} + x_{212} + x_{221} + x_{234} + x_{243}$  or  $0.66 + 0.25 + 0.35 + 0.77 + 0.19 + 0.66 + 0.22 + 0.84$ , which equals to 3.94.

##### 4.3. Mutation Operator

The assignment problems are naturally represented as permutations. The major problem of permutations is that it cannot be processed using the mutation operator in the traditional way. Reordered mutation has been applied to mutation operator [7, 8]. The primary difference of this operator and traditional mutation operator is the use of permutation technique on the selected values in represent solutions instead of change  $0 \rightarrow 1$  or  $1 \rightarrow 0$ . In this case, a feasible

solution and the reordered mutation operator, denote  $\Theta$ , as follows:

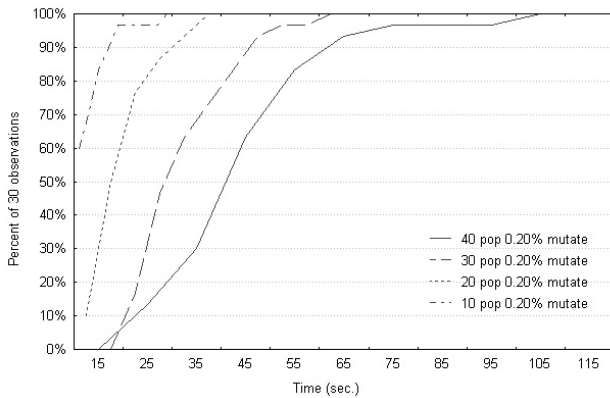
$$\Theta_{\{8,10,2\}}\{5,8,2,3,12,6,11,1,4,7,9,10\} = \{5,1,2,3,12,6,11,7,4,8,9,10\}.$$

Where a feasible solution is  $\{5,8,2,3,12,6,11,1,4,7,9,10\}$  with rate 0.25% (3-value need to be reordered), a selected index value is  $\{2,8,10\}$ , and reordered index is  $\{8,10,2\}$ , therefore, an offspring from this feasible solution is  $\{5,1,2,3,12,6,7,1,4,8,9,10\}$ . In reordered mutation, we generate index and reordered index with uniformly distribution.

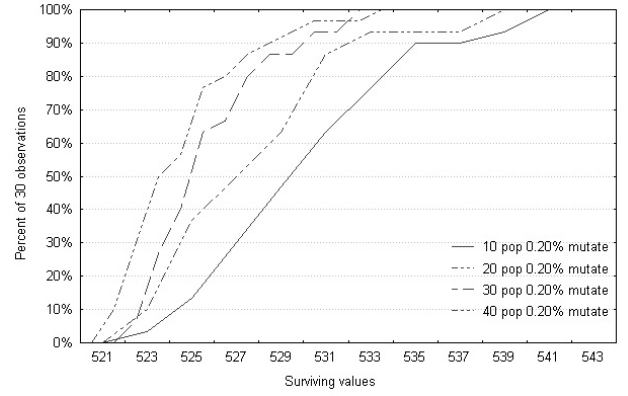
## 5. Numerical Results

In the experiment, the payoff array is uniformly generated in a unit cost from  $[0, 1)$ . The initial population is created by generating  $\mu$  feasible solutions. This experiment uses 10 – 40 population sizes. A mutation rate of 0.20% is used in this experiment. The problem size is  $10000^{10}$  variables. This algorithm was implemented in MATLAB and tested with a 1 GHz AMD Athlon processor and 512 MB memory. The algorithm terminates when 100 iterations have been executed, or when 20 consecutive iterations have failed to improve the best-known solution.

The experiment compares the proposed evolutionary programming with the different population size of the evolution in MATLAB. Results form the experiment using  $100^{10}$  decision variables are shown in Figure 5.1. – 5.2.



**Figure 5.1** Cumulative of computation time with  $10000^{10}$  decision variables in 30 observations using different population sizes.



**Figure 5.2** Cumulative of survival value with  $10000^{10}$  decision variables in 30 observations using different population sizes.

Table 5.1 shows computation time statistics for this experiment. The average of computation time of EP with 40 as the population size using 0.20% mutation rate is 48.31467 seconds when the problem size is  $10000^{10}$  decision variables. Parameters can be altered to reduce computation time, such as using 20 as the population size with mutation rate 0.20% will reduce the computation time to 21.98577 second. Moreover, computation can be reduced to 11.80433 second when a 10 population size with mutation rate 0.20% is used.

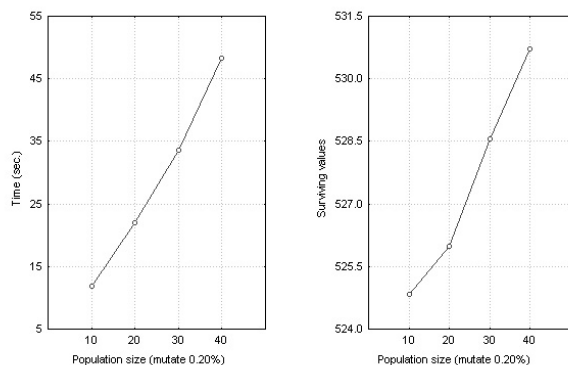
Table 5.2 shows total cost statistics. Average of total surviving value of EP with 10-population size using 0.20% mutation rate is 524.8358. While increase population size, the proposed EP can find the larger average total surviving, e.g. the average total cost of EP with 40 population size using mutate rate 0.20% is 530.7279, which is better than the average total cost of EP with 10 population size using mutate rate 0.20%, as shown in Table 5.2.

**Table 5.1** Computation time statistics using  $10000^{10}$  decision variables

Computation Time (sec.)	Mean	Confid. - 95.0%	Confid. +95.0%	Minimum	Maximum	Variance	Std.Dev.
EP 10 pop 0.20% mut.	11.80433	9.95462	13.65405	6.59900	28.3310	24.5385	4.95363
EP 20 pop 0.20% mut.	21.98577	19.44620	24.52533	13.25900	39.7370	46.2548	6.80109
EP 30 pop 0.20% mut.	33.61610	29.76347	37.46873	20.30900	62.0190	106.4516	10.31754
EP 40 pop 0.20% mut.	48.31467	42.31826	54.31108	26.30800	103.2790	257.8812	16.05868

**Table 5.2** Total survival value statistics using 10000<sup>10</sup> decision variables

Surviving Values	Mean	Confid. - 95.0%	Confid. +95.0%	Minimum	Maximum	Variance	Std.Dev.
EP 10 pop 0.20% mut.	524.8328	523.7413	525.9244	521.2625	533.6613	8.54549	2.923266
EP 20 pop 0.20% mut.	525.9841	524.9654	527.0027	522.7080	532.4460	7.44203	2.728008
EP 30 pop 0.20% mut.	528.5580	527.0562	530.0599	522.9000	538.7854	16.17687	4.022048
EP 40 pop 0.20% mut.	530.7279	529.0559	532.4000	523.5912	540.7522	20.05048	4.477776



**Figure 5.3** Compare and contrast computation time and survival values with 10000<sup>10</sup> decision variables using mutation rate 0.20%

Figure 5.3 illustrates compares and contrasts the parameters of the proposed EP. It can be seen that the proposed EP can find the better solution; however, computation will increase. Therefore, tradeoff analysis needs to be performed to decide optimal parameters.

## 6. Conclusion

This paper presents a robust heuristic to solve target-weapon assignment problems. The proposed method is developed by applying evolutionary programming with a reordered mutation initial experiment. Simulation results in this research reveal that this method is very effective. In fact, the evolutionary algorithm (e.g. genetic algorithm, evolutionary strategy, and evolutionary programming) works very well in the combinatorial optimization problems when the number of decision variables is large.

This paper makes no claim that the proposed algorithm is the most efficient for large-scale assignment problems. The contribution here is that evolutionary algorithms can effectively solve large and complex problems. This paper does not compare the proposed approach with other evolutionary algorithms, especially genetic algorithm, because crossover operator will increase the computation time in which we need to reduce. Work is continuing on building an algorithm to have further validation for the obtained results.

## Reference

- [1] Toet, A., Waard H. (1995). The Weapon-Target Assignment Problem. CALMA Report CALMA.TNO.WP31.AT.95c.
- [2] Wacholder, E. (1989). a neural network-based optimization algorithm for the static weapon-target assignment problem. *INFORMS Journal on Computing* (Vol.1 No. 4), pp. 232-246 .
- [3] Back, T. (Editor), Fogel, D. B. (Editor), Michalewicz, Z. (Editor) (1999). *Evolutionary Computation 1: Basic Algorithms and Operators*: Iop Pub.
- [4] Back, T. (Editor), Fogel, D. B. (Editor), Michalewicz, Z. (Editor) (1999). *Evolutionary Computation 2: Advance Algorithms and Operators*: Iop Pub.
- [5] Michalewicz, Z. (1999). *Genetic Algorithms + Data Structures = Evolution Programs* 3rd: Springer-Verlag New York.
- [6] Bean, J. C. (1994). Genetic Algorithms and Random Keys for Sequencing and Optimization. *ORSA Journal on Computing* (Vol. 6, No. 2), pp. 154-160.
- [7] T. Nandhakwang, M. A. Shaikh, An Efficient Algorithm for Large-Scale Assignment Problems, Hawaii International Conference on Statistics and Related Fields, Honolulu, Hawaii; 2003.
- [8] T. Nandhakwang, M. A. Shaikh, "Reordered Mutation Operator in Evolutionary Programming: An Assignment Problem Application", 4<sup>th</sup> International Conference on Dynamic Systems and Applications, Atlanta, Georgia; 2003.

## AUTHORS BIOGRAPHY



### Colonel Tawiwat Veeraklaew

received the Ph.D. degree in mechanical engineering from University of Delaware, Newark, DE, USA in 2000. He is a Platform and Material Senior Researcher at Defence Technology Institute (Public Organization), Bangkok, Thailand and supervise the Platform and Material Laboratory. He has published more than 30 both in conference and journal articles. His current research interests are in the area of controlled mechanical systems, dynamic optimization and special software hardware design.



### Colonel Settapong Malisuwan

received the Ph.D. degree in Electrical Engineering from Florida Atlantic University, Boca Raton, FL, USA in 2000. He is a member of Spectrum Refarming Committee, National Telecommuni- Cations Commission, Thailand. He has published more than 90 both in conference and journal articles.

# DATA PRESENTATION AND TRANSFORMATION IN TRAIN SCHEDULE INFORMATION SYSTEMS

Eugene Kopytov<sup>(a)</sup>, Vasilij Demidovs<sup>(b)</sup>, Natalia Petukhova<sup>(c)</sup>

<sup>(a)</sup> Transport and Telecommunication Institute, Riga, Latvia,

<sup>(b)(c)</sup> State Joint-Stock Company "Latvian Railway", Riga, Latvia,

<sup>(a)</sup> [kopitov@tsi.lv](mailto:kopitov@tsi.lv), <sup>(b)</sup> [dem@ldz.lv](mailto:dem@ldz.lv), <sup>(c)</sup> [natalia@ldz.lv](mailto:natalia@ldz.lv)

## ABSTRACT

In the given paper the authors have considered two approaches of the multi-version data modeling in train schedule information systems, based on the different forms of data presentation: point and interval forms correspondingly. The method of the interval data transformation in the point form is proposed. It is realized using SQL tools. The received results are applied in the information system of the inland train traffic schedule on the Latvian Railway.

Keywords: train schedule, database, data presentation, point form, interval form, data transformation

## 1. INTRODUCTION

The principal issues of the train traffic schedule storage in railway information systems (IS) are connected with the existence of the multitude of its versions conditioned by the seasonal cycles of schedule changes, the days of week, by the systematic and unplanned repair operations, the transfers of the working and festive days, and other factors (Kopytov, Demidovs, and Petukhova 2008).

The efficiency of manipulating with the schedule greatly depends on the organization of temporal data in the database. In practices there exists a variety of the data presentation models determined by a set of different IS using the trains' schedule and characterized by its functionality, the requirements to the resources and the traditions of development.

Research of different models of presenting temporal data is considered in the paper (Jensen, Soo, and Snodgrass 1994). The peculiarities connected with the management of the events' periodicity in the schedule tasks and with the development of a correspondent class of temporal models have been studied in the works (Terenziani 2003, Luca 2005, Terenziani and Luca 2006). The existing models of presenting temporal data have been studied by the authors in more details with the account of the specific passenger trains' schedule (Kopytov, Demidovs, and

Petukhova 2008; Kopytov, Petukhova and Demidovs 2010).

In the given paper the authors have considered two approaches of the multi-version data modeling in train schedule IS, based on the different forms of data presentation: point and interval forms correspondingly.

*Point (daily) form* is convenient for processing operative and analytical queries, but requires huge memory resources and great labour consumption for maintaining trains' traffic schedule in actual state. Though, the main problem of the above approach lies not in the volume of the stored data but in the complexity of providing the reliability of data and the immediacy of the schedule changes. Under the condition of frequent changes, these tasks become urgent.

Taking into account the fact that the schedule is usually made for a lengthy period of time (for example, for a season), it is more convenient to use the *interval form* of presenting data. This form allows substantial reduction of the stored data volume and provides flexibility of the schedule management. The interval form allows to use the data of the schedule's operative temporal database and to account all changes in the schedule as soon as they appear in the system. But along with this, the interval form is inferior to the point form in processing schedule queries and processing analytical queries in particular.

Thus, at different stages of processing the trains' schedule, there is reasonability of using different forms of presenting data, in one case the point form is preferable, in another case priority is given to the interval form. In the suggested research the authors use both models of the trains' schedule in one IS: the interval form – for storing data and the point form – for performing analytical calculations. For the transition of the data presented in the interval form to their point presentation, the authors have developed a special

method of data transformation which is considered in the given article.

## 2. PECULIARITIES OF TRAIN SCHEDULE INFORMATION SYSTEM

*The train schedule system* considered by authors is the reference system, presenting the information on the trains schedule for any period of time on the web site. Its principal task is the storage of the information on train traffic as at the current moment, so in the future and in the past, as well as provision of the access to this information for the railway customers on the purpose of travel planning. Note, that this information is used by decision-makers on the purpose of carrying out the versatile analysis of the train traffic too.

The system of the train schedule possesses the following peculiarities, determining the rather significant level of the complexity of its implementation considered below.

- The active schedule of the trains traffic is arranged on the basis of core fundamental (seasonal) timetable and the number of other schedules, occurring as a result of such procedures as the trains removal, changes in the traffic schedule due to the accidents, pre-planned or unplanned repair operations, public events (the City Days, concerts, etc.), short-run including / excluding the stations in / from the route. Further these schedules are called as *the schedule versions*.

- In the course of time the various versions rotate. One and the same version can come into operation repeatedly: the fundamental timetable exchanges the altered one, and vice versa.

- One and the same train might have the different schedules on the different week days.

- On holidays and other special days (for example, holding the NATO summit in Riga in November 2006) all the trains run in accordance with the day off schedule. There is an exception for the certain trains when the special timetable is assigned for these days. But from practical side it is known that only 1% of all trains are assigned with the special schedule.

- The trains run on the days of week or month the schedule is assigned for (for example, on Fridays only, or twice a week on Tuesdays and on Thursdays, on the working days, on the week-ends, on the even days of the month, daily).

- There is a break in the local commuter traffic in Latvia at night. In this connection the following peculiar features can be mentioned: the passengers and the system percept the opening and the closing of the day in different ways: the system day spans from 00:00

till 23:59, but the passengers' day can span for more than 24 hours, and for convenience are determined to span from 04:00 of one day and till 03:59 of the following day. All the trains, departing the night of the following day, the passengers refer to the current day. There is a special term – the night trains. For the passengers' convenience the system of the schedule delivery should take into account this peculiarity of understanding the day notion.

- The system of scheduling the long-distance trains must take into consideration the time zones, changing in the course of the train movement along the route. There are also peculiarities connected with the schedule data management (schedule registration).

- First of all the system obtains the core fundamental schedule, then with the time running the operative changes are contributed to this schedule. Consequently, several versions of the schedule appear. Quite often the period of operation of such altered schedule is so prolonged, that it also becomes subjected to certain corrections. The operational time of these or those corrections can be contracted or prolonged.

- The data on the train schedule version comprise the concise information about the order, in accordance with which the new schedule is introduced, the date of this schedule version coming into operation, the date of going this schedule version out of operation, the type and the number of carriages in the train, the set of stops at the stations with information on the time of arrival or departure, the type of the transport at the stations (in case of certain repair works the train between the definite stations can be exchanged by the bus).

- The system of schedule is utilized by the customers by means of the Internet very intensively. As the statistics states, the number of addresses to the system reaches 50 times during one minute period. The search for the corresponding train takes place under different conditions and criteria consideration: the station of departure and arrival, the date and the time range of the departure, the possibility of travel in transit, and so on. As all the queries to the system occur in the on-line mode, it becomes very important to provide the velocity of the result delivery. This task is complicated by the great volume of the processed data. In the relation database one version of the timetable for one train takes about 1 KB. The minimum of data is taken into account, without considering the reference information (such as the stations names, trains types, etc.). The integrated volume of all data of the passenger trains commuters schedule of the Latvian Railway in

interval form takes more than 100 MB (about 250 thousand records).

### 3. PERIODICITY IN THE TRAIN SCHEDULE

One of the most important properties of the schedule is its periodicity. It is the property that makes the basis of the effective schedule presentation in the databases. Periodicity means that the continuity of the object version's lifespan is broken cyclically and is valid only at the moments, which match the specified periodicity.

The periodicity of event is mathematically represents by the pair:

$$\langle [t_s, t_e], p \rangle,$$

where  $[t_s, t_e]$  is time interval;  $t_s$  and  $t_e$  are time of event beginning and time of event termination accordingly;

$p$  is periodicity; the examples of  $p$  are the following: "on Tuesday and Wednesday", "on even days", "at weekends", "every first Monday of the month".

Recently, some commercial databases started to support user-defined periodicity in queries in order to provide "a human-friendly way of time handling" (see, e.g., TimeSeries in Oracle 8). On the other hand, only few relational data models support user-defined periodicity in the data, mostly using "mathematical" expressions to represent periodicity. In this area Paolo Terenziani and his followers' research is worth attention (Terenziani 2003, Terenziani and Luca 2006, Luca 2005). This group of researches offer a high-level "symbolic" language for representing user-defined periodicity, which is more human-oriented than mathematical ones. In the above-mentioned works researchers consider temporal relational model which supports user-defined "symbolic" periodicity (e.g., to express "on the second Monday of each month") in the validity time of tuples and cover time ranges (e.g., "from 01.01.2009 to 28.02.2009"). Temporal counterpart of the standard operators of the relational algebra is defined, new temporal operators and functions are introduced. Temporal algebra, which is consistent extension of the classical (atemporal) one, is offered.

Periodicity  $p$  can be represented in different ways, ranging from binary or numeric representation and ending with a high-level symbolic language, for example (Terenziani 2003). The first way is not visual and is limited to a specific set of recognizable types of periodicity. That is, when the periodicity is a very complex expression, then you may encounter with a limitation or ambiguity. For example, the periodicity of "the second Monday of every even month, as well as the first working day of every odd month except

February" is a complex periodicity based on the basis of more simple types of periodicity connected by the certain rules. This approach allows to take into account such a complex periodicity. To do this, it computes the set of time moments of activity of each simple component of the periodicity and performs the subsequent operations with the computed item sets. Among these operations, there are all operations that can be performed with item sets: *merge*, *intersection*, etc.

However, you can look at the periodicity from another angle. It also represents a logical expression built on the basis of simpler types of periodicity, connected by the logical connectives in a certain sequence. Such a view gives quite different possibilities for computation. However, this approach has never been investigated so far.

### 4. EMPLOYING THE PERIODICITY PRINCIPLE IN THE RAILWAY SCHEDULE

As a matter of practice the train can have several active schedules simultaneously during the long period of time. Furthermore every train schedule is organized inclusive of days of running with the different *periodicity properties*, for example: on working days, on weekends, on the definite days of week, the first working day of week, the last day off of the week (notably it can be even Monday), on even or odd days etc.

As an example the multi-variant train schedule is presented in

Figure 1. This figure performs that the train has two basic schedules: one for the working days *wd*, and one for the weekends *rd*. Both timetables are valid on the time interval  $[t_1, t_4]$ , but on different week days. Then on Tuesdays and Thursdays (*Tue, Thu*) the amendment of this train schedule in the time period from  $[t_1, t_4]$  is introduced for the systematic repair operations providing. Therefore in the time interval  $[t_2, t_3]$  there are three timely schedule versions, activated on the days corresponding to the characteristics of *wd*, *rd* and (*Tue, Thu*), moreover, only one schedule version *wd* or *rd*, or (*Tue, Thu*) can be active in one day.

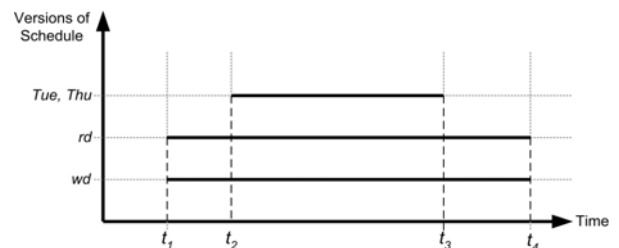


Figure 1: Multitude of the valid versions of the train schedule of different periodicity: *wd*, *rd* and (*Tue, Thu*)



This assertion is illustrated with help of the diagram in Fig.2. The axis “Day of week” shows the days of week, respectively 1 is Monday, 2 is Tuesday and so on. Time in this system is of the discrete nature, the level of the schedule granularity equals one day period. As the mentioned figure presents, one of three versions is the active train timetable in different days of week, and the other two versions are in the “shadow” zone at this moment. Moreover, the train schedule on Tuesdays and Thursdays (characteristic *Tue,Thu*) overlaps the working day schedule with the characteristic *wd*.

However besides the recurring versions of the train schedule there are also one-time changes in the timetable, connected with the special (altered) calendar. The following changes can be listed in this category: the festive days, the additionally appointed days-off, or the cancelled days-off. There is also such special occasion schedule alteration when the calendar working day exchanges the places with the calendar day-off. For example, in Latvia in 2007 the 30<sup>th</sup> of April, Monday (a working day) was transferred on the 14<sup>th</sup> of April, Saturday (a day-off). As a result on the 30<sup>th</sup> of April the trains ran according to the schedule of the weekends, and on the 14<sup>th</sup> of April the trains ran according to the working days timetable – instead of the 30<sup>th</sup> of April.

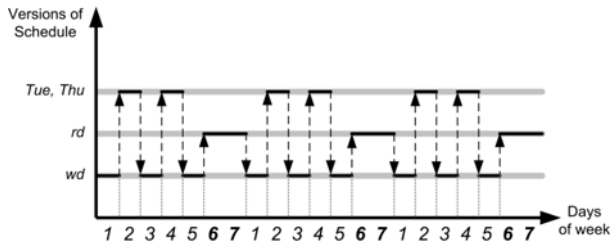


Figure 2: Periodicity in the train schedule

Such types of changes in the timetable or assignments provoke the anomalies (paradoxes). For example, the 19<sup>th</sup> of November, Monday, in 2007 was announced as an additional day-off, and all the trains moved this day according to the schedule of the weekends and days-off (like on the 18<sup>th</sup> of November), and this event arouse the malfunction of the periodicity, as it is shown in Fig.3. It is important to add that several trains with periodicity (Sundays) were cancelled on the 18<sup>th</sup> of November, because the trivial Sunday is not only the day-off, but also the last day of weekend, and the additional Sundays trains are considered for the passengers transportation from the traditional locations of the resting facilities on the days-off.

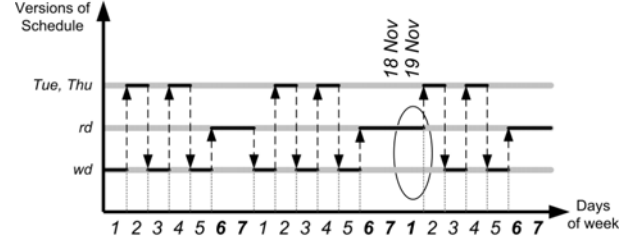


Figure 3: Periodicity malfunction in the train schedule

For correct functioning of the scheduling system, which is to take into account the above considered non-trivial rules, the well-defined formalization of the task assignment and the description of the scheduling temporal model are demanded. For this purpose the group of sets, characterizing the rules of arranging the trains running periodicity, the railway objects, and the entire system of scheduling itself, are introduced.

## 5. TRAIN SCHEDULE PRESENTATION MODELS

Two basic sets are introduced for the train schedule description:

$S = \{s_1, s_2, \dots, s_k\}$  is the set of all stations of the railways,  $|S| = k$ ;

$N = \{n_1, n_2, \dots, n_m\}$  is the set of the trains,  $|N| = m$ .

The schedule for the train with the number  $n \in N$  will be determined by the following vector:

$$v = \{n, \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_\alpha^{(n)}, T_\alpha \rangle\}, \quad (1)$$

where the pair  $\langle s_j^{(n)}, T_j \rangle$  determines the  $j$ -th train stop, the station  $s_j^{(n)} \in S$  exactly, and the train departure time  $T_j$ ;  $\alpha$  is the number of the stops of the train running on the timetable  $v$ .

In railway information system the train schedule is stored for the given period of time – the period of observation. In connection with possible changes of the train traffic schedule there appears the task of adjusting the train schedule to the particular dates. Here we can use two kinds of models of presenting the schedule in the point and the interval forms considered below.

**Point form.** The data model of the *daily schedule presentation* presents data in the point form. This model was very popular in the first generation of train schedule IS. We can find examples of this approach in some up-to-date IS, for example in the International System of the Tickets E-sales and E-reservation “Express-3” (Russia).

The given model is based on making a calendar of each train traffic for every day. Then  $i$ -th schedule

version  $v_i^{(n)}$  for the train with number  $n \in N$  will be determined by the tuple:

$$v_i^{(n)} = \{n, \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle, t_i\}, \quad (2)$$

$$i = 1, 2, \dots, m,$$

where the pair  $\langle s_j^{(n)}, T_j \rangle$  determines the  $j$ -th stop of the train with number  $n$ ;

$h$  is the quantity of the stops of the train running on the timetable  $v_i^{(n)}$ ;  $t_i$  is the day of operational period stored in database, when the train with number  $n$  is included in daily trains' schedule;  $m$  is the quantity of days on which the schedule of the train with number  $n$  is stored in database.

It should be noted that the number of the stored records, stating on what day, at what time, at what station every train stops, will be approximately equal to the product of multiplication of the number of trains by the average number stops and by the average number of days for which the schedule is stored. So, about two million records are needed to store the schedule of the trains of the inland traffic on the Latvian Railway in the relation database for a year time period, and for the greater railway companies the volume of the corresponding database will increase by times.

**Interval form.** The model used the interval form of data presentation significantly reduces the amount of stored data and does not limit the time period of data storage (Date, Darwen, and Lorentzos 2002). The usage of the interval form in train schedule IS, as well as its modifications, are considered in the paper (Kopytov, Demidovs, and Petukhova 2008).

For the specific schedule version identification for the train with number  $n \in N$  the tuple  $\langle n, C, t_f, t_s, t_e \rangle$  is employed. The property  $C$  presents the logical expression consisting of one or several elementary characteristics of the periodicity  $p_j$  connected by the logical operations signs  $\vee, \wedge$  and  $\neg$ ;  $t_s$  and  $t_e$  are the moments of time of the beginning and ending of the operational period of the considered schedule version,  $t_f$  is the time of the fixation of the schedule version in the database.

Then  $i$ -th schedule version  $v_i^{(n)}$  for the train with number  $n \in N$  will be

$$v_i^{(n)} = \{ \langle n, C, t_f, t_s, t_e \rangle, \langle s_1^{(n)}, T_1 \rangle, \langle s_2^{(n)}, T_2 \rangle, \dots, \langle s_h^{(n)}, T_h \rangle \}, \quad (3)$$

$$i = 1, 2, \dots, q,$$

where  $q$  is the quantity of schedule versions of the train with number  $n$ , it is significant that  $m \gg q$ .

For accounting a special calendar, we'll introduce the set  $Z = \{z_1, z_2, \dots, z_\mu\}$  of the transferred days of the train schedule, where  $z_i, i = \overline{1, \mu}$  is the pair of days  $\langle d_1^{(i)}, d_2^{(i)} \rangle$ , determining, that the timetable assigned on the date  $d_1^{(i)}$ , is assigned on the date  $d_2^{(i)}$  as well;  $|Z| = \mu$ .

Due to the complexity of managing the different periodicity schedule versions set, continuously appearing operative schedule changes, taking into consideration the non-trivial calendar, subjected to the changes, the development of the efficient methods of the active version determination is highly needed.

In the work (Kopytov, Petukhova and Demidovs 2010) the authors have investigated two methods of the temporal object active version determination: the logical rules (LR) method and the temporal elements (TE) method. An approbation of the offered methods for the actual object version determination has been carried out in the train schedule IS (on the example of Latvian Railways) and the efficiency of their application for solving different practical tasks has been determined. Obtained results testify a possibility for practical application of both LR and TE methods. At the same time there is an area of effective application for each of them.

## 6. INTERVAL DATA TRANSFORMATION INTO THE POINT FORM

The majority of the employed data analysis methods and correspondingly the majority of the analytical systems operate with the point forms of the data representation (for example, with the time sequences), but not with the interval forms. In the point form the fact is associated with the set of the time moments of the previously chosen granularity (for example, with the days, months, or years). The procedures over the described sets are very convenient for the fundamental computations or for the formal characterization. However, the mathematical apparatus of the relation databases management systems does not provide the ready-to-use facilities for the transformation of the interval form into the point one, and this fact required the elaboration of the appropriate procedures to be performed by the investigation under consideration.

Within the framework of the set task, the method of the interval data transformation in the point form was designed (Kopytov and Petukhova 2010). The scheme of data transformation is demonstrated in Fig. 4.

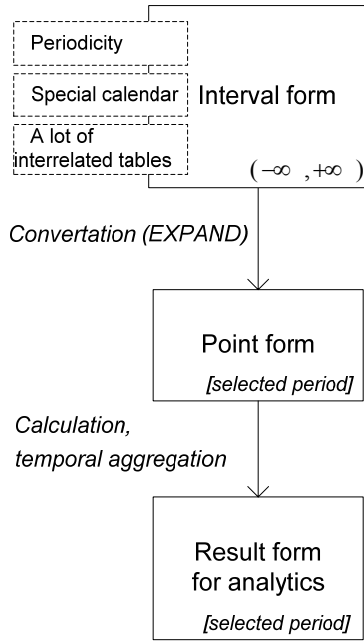


Figure 4: Scheme of the temporal data transformation

There is the explanation of the presented scheme. The classical interval form (*Interval form*) is initially inconvenient for the range of the certain calculations, but their complexity increases multifold when the following aspects exist: the periodicity, the special calendar, the necessity to use the several interconnected tables in the calculations<sup>1</sup>.

The application of the procedure EXPAND to the interval form temporal data, suggested by the authors, solves the problem of the inconvenient data presentation for the computations, by the way of turning the interval facts into the time sequences, in which every value gives characteristics of this fact existence in the definite moment of time (*Point form*). Then the data undergo the further processing, often connected with the temporal aggregation procedures. This stage might require even more serious computations connected with the temporal concepts representation. The required data form (*Result form for analytics*) is the final result of the processing procedure.

The initial sets for every stage embrace the different time periods. The set of the interval form temporal data (the first block in Fig. 4) is capable of covering the unlimited time range  $(-\infty, +\infty)$ . With the transfer to the point form (the second and the third blocks in Fig. 4) the time area of the temporal data is

<sup>1</sup> The information systems objects in the databases, according to the relation theory are represented in the form of the interconnected tables. That is why receiving the demanded result, as a rule, requires processing procedure for several tables.

narrowed to the field of analysis (selected period), and this fact considerably contracts the data bulk in the computations and substantially saves the computing resources, and solves the problem of versions existence with the indefinite values of the functioning periods beginning and ending.

There is the observation of the suggested method.

We consider two basic relations  $R1$  and  $R2$ :

$R1$  is the relation of the interval form, describing the set of the certain object versions and having the scheme  $R1(A, t_s, t_e, C)$ , where  $A = \{a_1, a_2 \dots a_k\}$  is the set of the attributes, describing the object,  $k$  is the number of attributes. Every version has the functioning period from the moment  $t_s$  to the moment  $t_e$  and possesses the feature of periodicity  $C$ . Besides, the activity moments of this or that versions depend on the specific calendar  $Z$ , describing the one day properties transfer to the others (the additional days-off or additional working days, etc.);

$R2$  is the point form temporal relation with the scheme  $R2(A, d)$ , where  $d$  is the time moment when fact  $A$  is active.

It is required to execute  $R1(A, t_s, t_e, C)$  transformation into  $R2(A, t_n)$ , and the data set of  $R2$  has to be restricted by the fixed time period (specified analysis period)  $T^{(q)} = [t_s^{(q)}, t_e^{(q)}]$ , where  $\forall d \in T^{(q)}$ .

The transformation formula has the following view:

$$R2 = \sigma_{d \in T^{(q)}} R1 \times \tau(t_s, t_e, C), \quad (4)$$

where  $\tau(t_s, t_e, C)$  is the function distributing the range  $[t_s, t_e]$  on the  $d$  dates set, responding the requirement of the specified periodicity  $C$ :

$$\tau(t_s, t_e, C) = \{d \mid d \in [t_s, t_e], p(d, C) = true\}, \quad (5)$$

where  $p(d, C)$  is the periodicity predicate, which is true in the case if date  $d$  has the property  $C$ . The predicate  $p(d, C)$  takes into consideration the specific calendar  $Z$  (for example, the transfer of the weekends schedule on first May's Tuesday in the connection with the national holidays in Latvia on 4th of May in year 2011);

$R1 \times \tau(t_s, t_e, C)$  is the construction, implementing the procedure of expansion; this procedure is the central operation of the entire transformation and provides the transformation from the interval form to the point one. The selection  $\sigma_{d \in T^{(q)}}(\dots)$  restricts the result by the analysis area  $T^{(q)}$ . The capacity of the resulting set can reach  $|R1| \cdot |\tau(T^{(q)})|$ .

After the executing the expansion procedure the following reorganizations (*grouping*, *natural joining* and *projection*) become trivial. The employment the proposed function  $\tau(t_s, t_e, C)$  permits to simplify the computations ultimately. For its implementation the SQL recursion mechanism is used. The created function TAU(dStart, dEnd, C) transforms the time interval [dStart, dEnd] into set of dates  $D = \{d\}$ , which corresponds to the periodicity and takes into account the special calendar.

The instruction of the function TAU creating has the following view:

```
CREATE FUNCTION TAU (
    dStart    DATE,
    dEnd      DATE,
    C         VARCHAR(30) )
RETURNS TABLE (D DATE )
LANGUAGE SQL
BEGIN ATOMIC
RETURN
    --start of SQL recursion
WITH D (d) AS
    (VALUES dStart
     UNION ALL
     SELECT d + 1 day as d
     FROM D
     WHERE d < dEnd
    )
    SELECT d FROM D
    WHERE P(d,C)=1;
    --end of SQL recursion
END;
```

Predicate function  $P$  implementing a predicate  $p$  determines the date InputDate compliance with the periodicity  $C$  taking into account the special calendar SPEC\_DAYS:

```
CREATE FUNCTION P (
    InputDate DATE,
    C         VARCHAR(30) )
RETURNS SMALLINT
LANGUAGE SQL
BEGIN ATOMIC
    DECLARE r    SMALLINT;
    DECLARE d1   DATE;
    DECLARE c1   VARCHAR(30);
    SET c1=c;
    IF c='wd' THEN
        SET c1='12345';
    IF c='rd' THEN
        SET c1='67';
    --spec. calendar
    SET d1=
```

```
(SELECT InsteadOfDay
    FROM SpecDays s
    WHERE s.changeDay=InputDate);
SET r=
CASE
    WHEN LOCATE(LEFT(CHAR(
        DAYOFWEEK_ISO(
            DATE(d1))), 1), c1) > 0
    THEN 1
    ELSE 0
END;
RETURN r;
END;
```

## 7. EXAMPLES OF REALIZATION OF THE TRANSFORMATION METHOD

The first example illustrates the usages of the SQL-functions described in section 5, the second example presents realization of the method of transforming the interval data into the point ones.

**Example 1.** Let the fragment of data of a special calendar  $Z$  from the table SpecDays be given in the form of the following table:

ChangeDay	InsteadOfDay
22.04.2011	23.04.2011
25.04.2011	23.04.2011
04.05.2011	30.04.2011
23.06.2011	18.06.2011
24.06.2011	18.06.2011

where the pair  $\langle \text{ChangeDay}, \text{InsteadOfDay} \rangle$  corresponds to the elements  $\langle d_1^{i1}, d_2^{i2} \rangle$  of set  $Z$ .

For selecting the days-off ( $rd$ ) in the interval [20.09.2011, 25.09.2011], the SQL-query with the use of the function TAU looks as follows:

```
SELECT d, DAYNAME(d) as weekday
FROM TABLE(
    TAU(CAST('20.06.2011' AS DATE),
        CAST('28.06.2011' AS DATE), 'rd'));
```

Having performed the given SQL-query, we'll get the following result:

D	Weekday
23.06.2011	Thursday
24.06.2011	Friday
25.06.2011	Saturday
26.06.2011	Sunday

As we see from the resulting table, the days 23.06.2011 (Thursday) and 24.06.2011 (Friday) are also related to the days-off. These dates have been described in a special calendar as the days-off.

SQL-query for selecting the week days (*wd*) in the interval [20.09.2011, 25.09.2011] looks as follows:

```
SELECT d, DAYNAME(d) as Weekday
FROM TABLE(
    TAU(CAST('20.06.2011' AS DATE),
        CAST('28.06.2011' AS DATE), 'wd'));
```

The results of this query are presented in the following table:

D	Weekday
20.06.2011	Monday
21.06.2011	Tuesday
22.06.2011	Wednesday
27.06.2011	Monday
28.06.2011	Tuesday

Note that the re-determined days 23.06.2011 (Thursday) and 24.06.2011 (Friday) are not present in the resulting table.

**Example 2.** There is examined the task of procurement of the time series, describing the changes of the trains number on the railway line Riga – Daugavpils (in both directions) for May, 2009.

The segment of the database comprising the data, necessary for calculations, is presented in Fig 5. The temporal table *Schedule* contains a lot of the trains schedule versions, each of which has its own period of operation from the moment *SchStart* to the moment *SchEnd* and possesses the periodicity characteristics *Periodicity*; this fact is reflected in the corresponding temporal attributes of the valid time *Valid Time*. The moments when this or that version is active depend on the peculiar calendar *Special calendar*, describing the transition of the properties from one days to the other ones (the additional days off or the working days, and so on) and affecting all the schedule versions. The detailed description of the schedule versions (time of the train stops at the stations) is introduced in the table *Schedule Stops*. The tables *Trains* and *Lines* are not the temporal ones and they contain the information on the trains, as well as their belonging to the railway lines.

The structure of the time series under examination is assumed in the form  $(Date, Line, TrainQty)$ , its temporal area in the form  $[01.05.2009; 31.05.2009]$  accomplishing  $T^{(q)}$ , the period of the schedule activity  $[schStart, schEnd]$  accomplishing  $T^{(S)}$ ; the periodicity attribute *Periodicity* accomplishing  $C$ .

Then the formula of transformation obtains the following form:

$$R(Date, Line, TrainQty) = \\ = \pi_{d, Line, COUNT(*)}(Lines \triangleright \triangleleft Trains \triangleright \triangleleft \\ \triangleright \triangleleft \sigma_{d \in T^{(q)}}(Schedule \times \tau(T^{(S)}, C))),$$

where construction

$\pi_{d, Line, COUNT(*)}(Lines \triangleright \triangleleft Trains \triangleright \triangleleft \dots)$  solves the standard task of grouping and the natural joining the connected relation; construction  $Schedule \times \tau(T^{(S)}, C)$  executes the expansion operation.

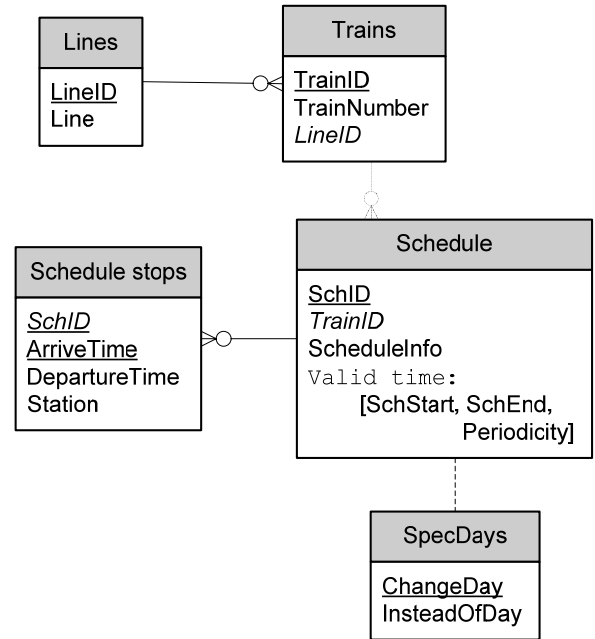


Figure 5: Segment of the database of the passenger trains schedule

As a result of its accomplishing, 80 original tuples of the interval form of the relation *Schedule* (trains schedule on the line Riga – Daugavpils) are transformed into 2170 tuples of the point form. Without taking into account other transformation this construction has the following view in the SQL language:

```
SELECT Schedule.*, d
FROM Schedule,
    TAU(schStart, schEnd, periodicity);
```

The completed SQL-statement for constructing the time series  $R(Date, Line, TrainQty)$  has the following view:

```

SELECT d as date, line,
       count(train) as TrainQty
FROM Schedule,
       TAU(schStart, schEnd, periodicity),
       Lines, Trains
WHERE Schedule.trainID = Train.trainID
       AND Trains.LinesID = Lines.LineID
       AND d between
           '01.05.2009' AND '31.05.2009'
GROUP BY line, d;

```

## CONCLUSIONS

The approach to presenting the data of the trains' schedule considered in the given paper suggests simultaneous usage of two principally different forms of storing data: point form and interval form. It allows using the advantages of both forms. Thus, the point form provides convenient processing of the schedule data in performing analytical calculations; the interval form substantially reduces the volume of the stored data and the labour consumption on the schedule maintenance in the actual state.

The two forms of presenting data and the need of their simultaneous usage have required development of a mechanism of transforming data from one form to another; for this purpose, the paper suggests development of a method of transforming the schedule data from the interval form to the point one.

The received results are applied in the information system of the inland train traffic schedule on the Latvian Railway. However, they can be implemented in the information systems of other railway companies. Moreover in the obtained results have the universal nature and can be applied in temporal databases regardless their destination.

The paper considers realization of the predicate function  $P$ , which allows employment of the simplest logical formula for the periodicity  $C$  containing the elementary attributes connected by the operators "OR". At the further stage of research, realization of the predicate function  $P$  will be elaborated in such a way that it would be possible to describe more complicated formulae of the periodicity  $C$  which can occur in practice. For this, it is supposed to develop a corresponding specialized compiler.

## REFERENCES

- Date C. J., Darwen H. and Lorentzos N. A., 2003. *Temporal Data and the Relational Model*. Morgan Kaufmann Publisher.
- Jensen C. S., Soo M. D. and Snodgrass R. T., 1994. Unifying Temporal Data Models via a Conceptual Model. *Information Systems*, 1, 513-547.
- Kopytov E., Demidovs V. and Petukhova N., 2008. Application of temporal elements in the railway

schedule systems. *Transport and Telecommunication*, 9 (2), Riga, 14-23.

Kopytov E. and Petukhova N., 2010. Application of Temporal Databases in Modeling of Railway Processes. *Proceedings of the International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO'10)*, pp. 586-594. February 8-11, SCE – Shamoon College of Engineering (Beer Sheva, Israel).

Kopytov E., Petukhova N., Demidovs V., 2010. Method for Railway Schedule Support in Temporal Databases. *Proceedings of the 9-th Join Conference on Knowledge-Based Software Engineering (JCKBSE'10)*. August 25-27, Technologija, pp. 178-191 (Kaunas, Lithuania).

Terenziani P., 2003. Symbolic User-Defined Periodicity in Temporal Relational Databases. *IEEE Transactions on Knowledge and Data Engineering*. February, Volume 15, Issue 2, pp. 489-509.

Terenziani P. and Luca A., 2006. Temporal reasoning about composite and/or periodic events. *Journal of Experimental & Theoretical Artificial Intelligence (Taylor & Francis)*. March, Volume 18, No 1, pp. 87-115.

Luca A., 2005. *Representing and reasoning with classes and instances of possibly periodic events. Theory, algorithms and applications*. Dr. techn. thesis, 118 p.

## AUTHORS BIOGRAPHY



**EUGENE A. KOPYTOV** was born in Lignica, Poland and went to the Riga Civil Aviation Engineering Institute, where he studied Computer Maintenance and obtained his engineer diploma in 1971. Candidate of Technical science degree (1984), Kiev Civil Aviation Engineering Institute. Dr.sc.ing. (1992) and Dr.habil.sc.ing. (1997), Riga Aviation University. Professor (1998). Present position: Chancellor of Transport and Telecommunication Institute, professor of Computer Science Department. Member of International Telecommunication Academy. Fields of research: statistical recognition and classification, modelling and simulation, modern database technologies. Publication: 260 scientific papers and teaching books, 1 certificate of inventions.

His e-mail address is: [kopitov@tsi.lv](mailto:kopitov@tsi.lv)



**VASILIS V. DEMIDOVVS** was born in Murmansk, Russia and went to the Riga Polytechnic Institute (Riga Technical University), where he obtained his engineer diploma in 1987. He



studied in Transport and Telecommunication Institute under the Doctor's program from 2001 to 2004 and obtained the scientific degree of Doctor of Science in Engineering (Dr.sc.ing.) in 2006 by scientific research direction "Transport and Logistics". He has got a docent since 2007. Present position: a Head of Database Management System Unit of Information Technology Centre of State Joint-Stock Company "Latvian Railway"; docent of Computer Science Department of Transport and Telecommunication Institute. Fields of research: modern database technologies, Business Intelligence, Data warehouse technologies. Publication: 31 scientific works in Belgium, Estonia, Italy, Latvia, Lithuania, Russia, USA, including 18 scientific papers.

His e-mail address is: [dem@ldz.lv](mailto:dem@ldz.lv)



**NATALIA Y. PETUKHOVA** was born in Riga, Latvia. She studied Computer Sciences in Transport and Telecommunication Institute and obtained her MS in 2001, obtained the scientific degree of Doctor of

Science in Engineering in 2011 by scientific research direction "Transport and Logistics". Present position: leader specialist and DB administrator of Database Management Systems department of Computer Centre of Latvian Railway; docent in Transport and Telecommunication Institute. Publications: 18 scientific papers. Scientific activities: databases, data warehouses, business intelligences.

Her email address is: [natalia@ldz.lv](mailto:natalia@ldz.lv)

# DYNAMIC SIMULATION BASED DESIGN OF POWERED SUPPORT IN INTEGRATED MECHANIZED COAL MINING

Zhifeng Dong<sup>(a)</sup>, Guozhu Liu<sup>(b)</sup>, Jian Wang<sup>(c)</sup>, Jiangang Qian<sup>(d)</sup>, Shouxiang Zhang<sup>(e)</sup>, Dongsheng Bao<sup>(f)</sup>

<sup>(a)</sup> Mechanical and Electronic Engineering Department of China University of Mining and Technology(Beijing)

<sup>(b)</sup> Mechanical and Electronic Engineering Department of China University of Mining and Technology(Beijing)

<sup>(b)</sup> Beijing Mining Machinery Company Limited of China Zhongmei Group

<sup>(c)</sup> Beijing Mining Machinery Company Limited of China Zhongmei Group

<sup>(d)</sup> Beijing Mining Machinery Company Limited of China Zhongmei Group

<sup>(e)</sup> Beijing Mining Machinery Company Limited of China Zhongmei Group

<sup>(f)</sup> Beijing Mining Machinery Company Limited of China Zhongmei Group

<sup>(a)</sup> [dzf@cumb.edu.cn](mailto:dzf@cumb.edu.cn), <sup>(b)</sup> [bmjlgz@163.com](mailto:bmjlgz@163.com), <sup>(c)</sup> [wjian3524@sina.com](mailto:wjian3524@sina.com), <sup>(d)</sup> [qianjiangang88@sohu.com](mailto:qianjiangang88@sohu.com),  
<sup>(e)</sup> [bmjzsx@163.com](mailto:bmjzsx@163.com), <sup>(f)</sup> [bmjbds@163.com](mailto:bmjbds@163.com)

## ABSTRACT

Powered support is one of the key equipments in the integrated mechanized coal mining. The kinematics equations of powered support are presented and the optimum model is built. The simulation and optimization improve the original design greatly. Force analysis is made and force balance equations of main parts are deduced. Computer program and software are developed to derive solutions to calculate the interact forces of the powered support. The dynamic simulations are made for powered support in different working conditions. The results show the variations of loads of the roof and the floor acting on canopy and base, and the variation forces of the front and the back linkages as the height of powered support increases. The simulation based design of powered supports have achieved great success in the application of integrated mechanized coal mining in Pingsu mines of Zhongmei Group in China.

Keywords: dynamic simulation, force analysis, powered support design, coal mining

## 1. INTRODUCTION

The coal represents a major source of energy among all natural resources, and 60-70% of the energy consumption in China depends on the coal. Underground coal mining is complex and challenging (Costa, J. F., Zingano, A. C., Koppe and J.C., 2000; Kelly, M. 1999; Singh, T. N., Singh, B. 1985). The design of supports is one of the most important for effective roof control in coal mining. Powered supports had come after a long development of wood and steel supports in longwall faces of coal mine. The earliest powered supports were manufactured by the Gullick Company of England. In the middle of the past century, the shield supports were developed to keep with easily caving faces in coal mine.

Technological progress over the past decade has allowed Chinese engineers to construct a considerable

number of modernized underground coal mines, which deeply involves mechanization, automation, information technology equipments, and which has brought the risks, toughness and disadvantages of underground coal mining to an end. Examples of such constructions are the various coal mines in Shendong of China Shenhua Group. Annual production level of these coal mines are over tens of millions of tons, and they are also home to the largest coal mine longwall face, equipped with the biggest integrated mechanized mining machines (shearer, conveyor and powered supports) in the world, machines of which the shearing height amounts to 7m. These mining machines, equipped with advanced hydraulic, controlling, sensory, communication, electronic, and computer systems, realize considerable productivity, efficiency and safety. The Pingsu coal mine of China Zhongmei Group applies integrated mechanized top coal caving mining techniques to realize annual production level of tens of millions of tons under the condition of the hard top coal and rocks, breaking the world record of top coal caving mining. The design, manufacture, assembly and operation of the integrated mechanized mining machines applied by these large longwall faces of coal mines are very demanding and complex, making simulation technology one of the best ways to achieve the desired effect (Li, J., Chen, H. and Zhang, J. 2007).

This paper deals with the dynamic simulation based design of powered support in integrated mechanized coal mining, and the analysis of the kinematics and dynamic characteristics for the optimum of powered support.

## 2. STRUCTURE OF POWERED SUPPORT

### 2.1. Integrated Mechanized Mining Equipment

The integrated mechanized mining equipments mainly consist of the shearer, the conveyor, and the powered supports. The shearer cuts the coal seam between the

top and bottom rocks by the two drums with helical blades, which is assembled with hard alloy teeth, and moves the coal on the conveyor, as shown in Fig.1. Then the conveyor transports the coal to the belt and to the ground. The powered supports support the top roof (rock or silt stone above the seam) of the longwall face, providing safety rooms and protects the top roof (rock or silt stone) from falling down and damaging the shearer and the conveyor (Dong, Z., Chang, H. and Wang, S. 2001), or wounding the workers. The shearer and the conveyor are serial, standard, products with different types, but geological conditions, such as the characteristics of rocks, silt stone and coal seams are formed by nature and therefore multifarious (Knessel, H.C.W. and Mischo, H. 1999). As the types of the shearer and the conveyor are selected according to the power, running speed and productivity of the longwall face, the design of powered supports must not only take into consideration the thickness and inclination angle of different seams and the variable pressure, hardness and strength of top roof rocks (Qan, M., He, F. and Miao, X. 1996), but also the uniformity and coordination of position and movement among the shearer and the conveyor. Therefore, the design of powered support is demanding and complex (Dong, Z., Li, H. and Sun, M. 2004), and the importance of kinematic and dynamic analysis is highlighted.



Fig. 1 Integrated Mechanized Face in Coal Mine

## 2.2 Structure of Powered Support

Powered support is mainly composed of a horizontal base plate that sits on the floor, linkages, shield, canopy, cylinders, and legs, as shown in Fig. 2. The canopy and the base are connected with the shield and the linkages by hinges. The special linkages between the base and shield maintain a constant distance between the face and canopy tip whether the hydraulic cylinders of legs move up or down. The top ends of the two legs are connected with the canopy by sphere pairs, and the bottom ends of the two legs are connected with the base also by sphere pairs. The top end of the balance cylinder is connected with the canopy by hinge, and the bottom end of it is connected with the shield by another hinge. The legs are double telescopic cylinders driven by hydraulic power, and push the canopy up to support the loads of top roof rocks, then unload and bring the canopy back down

when the powered support moves. The broke rocks and blocks from the top roof are sheltered by the shield from the workers.

Theoretically, powered support is a multi-linkages mechanism. The positions, displacements and movement paths of the different structures and parts are changing when the hydraulic cylinders drive the powered support. These geometry parameters of the powered support determine the positional relationship between the powered support and the shearer, and the relationship between the powered support and the conveyor in the working space. Kinematics parameters of the powered support determine its coordinate actions with the shearer and the conveyor. The relative positions of the shearer, the conveyor and the powered supports change in accordance with the variation of geology conditions, such as the thickness and inclination angle of the seams, therefore the alternation of the relative positions must be controlled effectively.

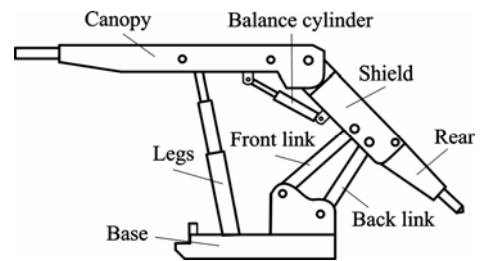


Fig. 2 Powered Support Structure

## 3. KINEMATIC SIMULATION

Kinematics equations are built to calculate and simulate the positions, displacements and movement paths of the different structures and parts when the hydraulic cylinders drive the powered support. The simulation provides fundamentals for the kinematics design of powered support, and protects the shearer, the conveyor and the powered supports from the interfering with each other or collision during operation.

### 3.1. Kinematics Equations

The coordinate system of powered support is built as shown in Fig. 3.

The point  $M$  on shield is expressed as follow:

$$U^2 + V^2 = W^2 \quad (1)$$

This is the equation of the canopy movement.

Where, there are:

$$\begin{cases} U = f[(x-d)\cos\gamma + y\sin\gamma](x^2 + y^2 + e^2 - a^2) \\ \quad - ex[(x-d)^2 + y^2 + f^2 - c^2] \\ V = f[(x-d)\sin\gamma - y\cos\gamma](x^2 + y^2 + e^2 - a^2) \\ \quad + ey[(x-d)^2 + y^2 + f^2 - c^2] \\ W = 2ef\sin\gamma[x(x-d) + y^2 - dy\cot\gamma] \end{cases} \quad (2)$$



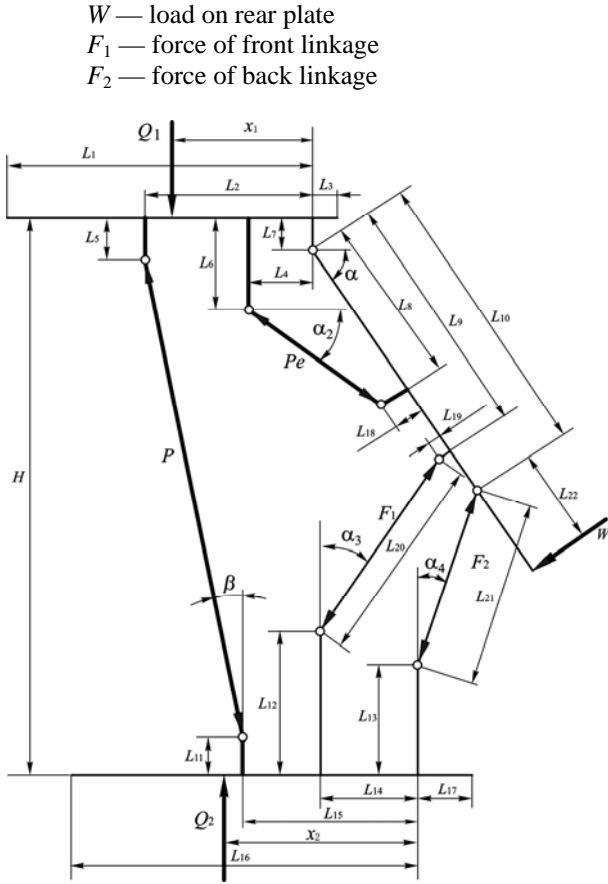


Fig.5 Loads and Forces on Powered Support

The force balance equations of shield are:

$$\begin{cases}
 P_E \cos(\alpha - \alpha_2) + R_x - F_1 \sin(\alpha - \alpha_3) - F_2 \sin(\alpha - \alpha_4) = 0 \\
 P_E \sin(\alpha - \alpha_2) + F_1 \cos(\alpha - \alpha_3) + F_2 \cos(\alpha - \alpha_4) - R_y - W = 0 \\
 P_E \cos(\alpha - \alpha_2)L_{18} + P_E \sin(\alpha - \alpha_2)L_8 + F_1 \cos(\alpha - \alpha_2)L_9 \\
 - F_1 \sin(\alpha - \alpha_3)L_{19} + F_2 \cos(\alpha - \alpha_4)L_{10} - W(L_{10} + L_{22}) = 0
 \end{cases}
 \quad (7)$$

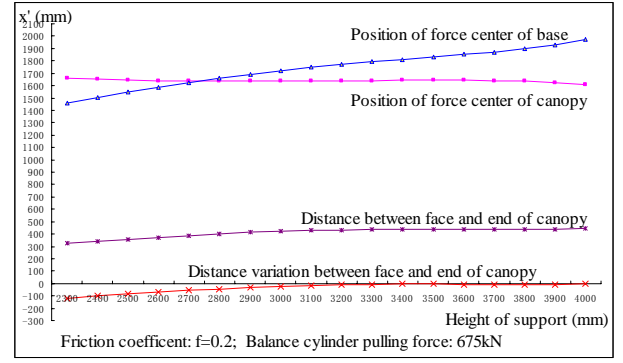
The force balance equations of base are:

$$\begin{cases}
 P \sin(\beta) - F_1 \sin(\alpha_3) - F_2 \sin(\alpha_4) - fQ_2 = 0 \\
 Q_2 - P \cos(\beta) - F_1 \cos(\alpha_3) - F_2 \cos(\alpha_4) = 0 \\
 P \cos(\beta)L_{15} - P \sin(\beta)(L_{11} - L_{13}) + F_1 \cos(\alpha_3)L_{14} \\
 + F_1 \sin(\alpha_3)(L_{12} - L_{13}) - Q_2x_2 - fQ_2L_{13} = 0
 \end{cases}
 \quad (8)$$

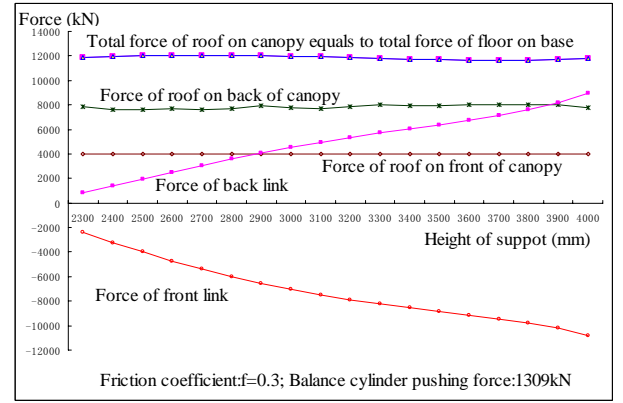
The computer program and software are developed to derive solutions to simulation problems. The simulations, done by computer software, are able to calculate the interact forces of the base, the linkages, the shield and the canopy of the powered support at different heights and positions.

#### 4.3 Dynamic Simulation

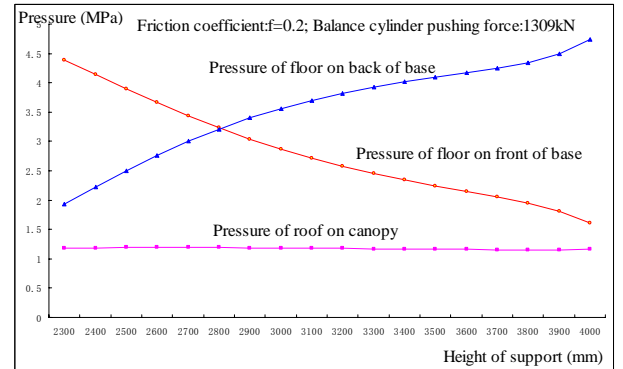
For ZF12000/23/40D type powered supports, the interact forces of base, linkages, shield and canopy of



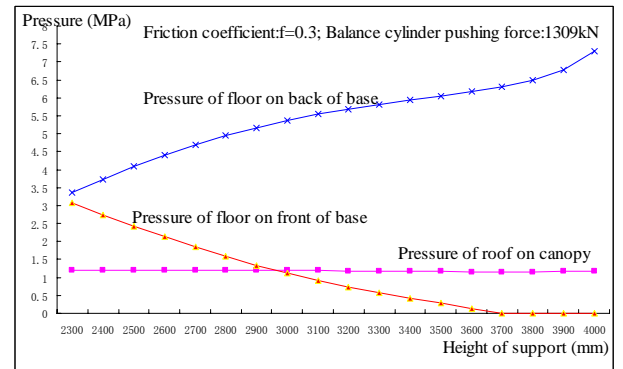
(a) Force Centers and Canopy Position



(b) Force Variation of Different Parts of Support



(c) Pressure Variations on Canopy and Base



(d) Pressure Variations on Canopy and Base

Fig. 6 Dynamic Simulation



the powered support at different heights and positions are calculated and simulated. Some of the results are shown in Fig. 6.

Figure 6 (a) shows the center positions of total normal force of roof acted on canopy and that of floor acted on base, distance between the front end of canopy and face of the coal wall, and the distance variations between the front end of canopy and face of the coal wall. In this simulation example, the calculation friction coefficient of roof and floor acted on the powered support is set as 0.2, and the force of balance cylinder of powered support is 675 kN, which is polling force acted on canopy and shield. The simulation results indicate that the center positions of total normal force of roof acted on canopy and that of floor did on base change with different trends as the height of powered support increasing, which the center point of total normal force of roof acted on canopy moves backward slightly, and that of floor acted on base moves forward gradually.

Figure 6 (b) shows the forces variation of main parts of powered support. In this simulation example, the calculation friction coefficient of roof and floor acted on the powered support is set as 0.3, and the force of balance cylinder of powered support is 1309 kN, which is pushing force acted on canopy and shield. The simulation results indicate that the total normal force of roof acted on canopy equals that of floor acted on base, which is about 12000 kN, same as working resistant force. In order to investigate the affectivity of canopy of powered support to the roof, the calculation is done respectively for both ends of canopy. The results indicate that the normal force of roof on the back end of canopy varies approximately from 3950 kN to 4000 kN as the height of powered support increasing from 2300 mm to 4000 mm, and of roof on the front end of canopy varies approximately from 7900 kN to 8000 kN as the height of powered support increasing from 2300 mm to 4000 mm. The forces of linkages change obviously as the height variation of powered support. The pushing force of front link increases sharply from 2400 kN to 10800 kN as the height of powered support increasing from 2300 mm to 4000 mm. The polling force of back link increases sharply from 800 kN to 9000 kN as the height of powered support increasing from 2300 mm to 4000 mm.

Figure 6 (c) shows the pressures variation of canopy and base. In this simulation example, the calculation friction coefficient of roof and floor acted on the powered support is set as 0.2, and the force of balance cylinder of powered support is 1309 kN, which is pushing force acted on canopy and shield. From the simulation results, it can be found that the average pressure value of canopy is almost constant from 1.15 to 1.19 MPa in the whole working height of the powered support. However, the pressure values of different ends of base change evidently. The pressure of front end of base decreases from 4.39 to 1.61 MPa, and the pressure of back end of base increases from 1.93 to 4.73 MPa.

Comparing with Fig. 6 (c), Figure 6 (d) shows the pressure variations of canopy and base. In this simulation example, the calculation friction coefficient of roof and floor acting on the powered support is set as 0.3, and the force of balance cylinder of powered support is 1309 kN, which is pushing force acted on canopy and shield. From the simulation results, it can be found that the average pressure value of canopy is almost constant from 1.19 to 1.21 MPa in the whole working height of the powered support. However, the pressure values of different ends of base change evidently. The pressure of front end of base decreases from 3.06 to 0.00 MPa, and the pressure of back end of base increases from 3.37 to 7.29 MPa, which must be paid attention in the design of powered support working under wet floor conditions.

The simulation makes foundations for the structure parameter optimizing design of the powered support.

## 5. MODELING AND DESIGN

The integrated mechanized coal mining technique significantly reduces the risks in underground working, increases productivity, and improves efficiency. At the same time, the integrated mechanized coal mining technique also brings about more requirements concerning the design, manufacture and application of the mining equipments. In order to meet engineering demand, 3D solid model of powered support are built based on the integrated mechanized coal mining in Pingsu, and virtual assembly prototype are created as shown in Fig. 7.

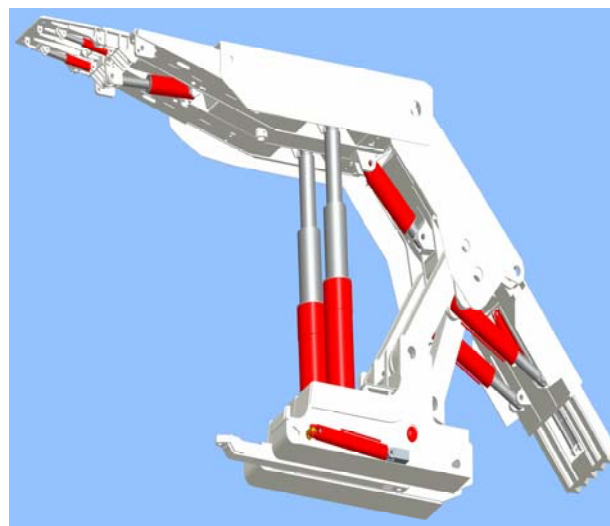


Fig. 7 3D Modeling and Design of Powered Support

The 3D graphics clearly and precisely display the parts, structures and the relationships of the powered support during movement, and simulate the powered support operation. Power supports designed on the basis of the dynamic, graphic simulation are performing with great excellence in Pingsu mines of Zhongmei Group in China.



## 6. CONCLUSIONS

(1) The kinematics equations of powered support are presented and the optimum model is built. Through simulation and optimization, the maximum distance variation between the face and canopy tip is minimized to 20mm, improving the original design greatly.

(2) Force analysis is made for powered support, and force balance equations of main parts are deduced. Computer program and software are developed to derive solutions to calculate the interact forces of the base, the linkages, the shield and the canopy of the powered support.

(3) The dynamic simulations are made for powered support under different working conditions. The results show the variation of loads of the roof and the floor acting on the canopy and the base, and the variation forces of the front and the back linkages as the height of powered support increases.

(4) The dynamic simulations indicate the effect of various friction coefficient between the roof and the canopy, and the floor and the base on loads. The pressure distribution on base and floor varies greatly, in which the maximum pressure is 7.29 MPa while the friction coefficient is 0.3, and the maximum pressure is 4.73 MPa while the friction coefficient is 0.2.

(5) The simulation makes foundations for the powered support of the structure parameter optimizing design. The simulation based design of powered supports have achieved great success in the application of integrated mechanized coal mining in Pingsu mines of Zhongmei Group in China.

## ACKNOWLEDGMENTS

This research is sponsored by “The Fundamental Research Funds for the Central Universities” (No.2010YJ04), China University of Mining and Technology (Beijing) “211” and “985” Project founded by Education Ministry of China. The authors would like to acknowledge and express their thanks for the support.

## REFERENCES

- Costa, J. F., Zingano, A. C., Koppe, J.C., 2000. Simulation — An Approach to Risk Analysis in Coal Mining. *Exploration and Mining Geology*, 9(1), 43-49.
- Dong, Z., Chang, H., Wang, S., 2001. Virtual Reality of Integrated Mechanized Top-Coal Caving Machine in Longwall. *Proceedings of The 29<sup>th</sup> International Symposium on Computer Applications in The Mineral Industries*. pp. 387-390. May 25-27, Beijing (Beijing, China)
- Dong, Z., Li, H., Sun, M., 2004. Virtual Design of A New Type Hydraulic Support. *Chinese Journal of Mechanical Engineering (English Edition)*, 17, 205-208.
- Kelly, M., 1999. Developing Coal Mining Technology For The 21st Century. *Proceedings of The '99 International Symposium on Mining Science and Technology*. pp. 3-7. August 29-31, Beijing (Beijing, China)
- Knessel, H.C.W., Mischo, H., 1999. High Performance Longwall Extraction in Large Depth. *Proceedings of The '99 International Symposium on Mining Science and Technology*. pp. 31-38. August 29-31, Beijing (Beijing, China)
- Li, J., Chen, H., Zhang, J., 2007. Design and Realization of Virtual Reality System of Fully Mechanized Mining Face. *Proceedings of The 2007'International Symposium on Safety Science and Technology*. pp. 42-49. April 16-19, Jiaozuo (Henan, China)
- Qan, M., He, F. and Miao, X., 1996. The System of Strata Control around Longwall Face in China. *Proceedings of The '96 International Symposium on Mining Science and Technology*, pp. 15-18. October 16-18, Xuzhou (Jiangsu, China)
- Singh, T. N., Singh, B., 1985. Model Simulation Study of Coal Mining Under River Beds in India. *International Journal of Mine Water*, 4 (3), 1-10.

## AUTHORS BIOGRAPHY

**Dong, Zhifeng**, Ph. D, Professor of Department of Mechanical and Electronic Engineering of China University of Mining and Technology. Majors in mechanical engineering and mining engineering; possesses great interest in machine design, virtual reality, CAD, and CAE.

**Liu, Guozhu**, Ph. D candidate in China University of Mining and Technology. Vice manager in chief of the Beijing Mining Machinery Company Limited of China Zhongmei Group. Majors in mechanical engineering and mining engineering; possesses interest in powered support design and manufacturing.

**Wang, Jian**, Vice manager in chief of the Beijing Mining Machinery Company Limited of China Zhongmei Group. Majors in mechanical engineering and mining engineering; possesses interest in powered support design and manufacturing.

**Qiang, Jiangang**, Manager in chief of the Beijing Mining Machinery Company Limited of China Zhongmei Group. Majors in mining machinery; possesses great interest in powered support design and manufacturing.

**Zhang, Shouxiang**, Vice director of powered support design institute of Beijing Mining Machinery Company Limited of China Zhongmei Group. Majors in mining machinery; possesses interest in powered support design and manufacturing.

**Bao, Dongsheng**, Vice director of powered support design institute of Beijing Mining Machinery Company Limited of China Zhongmei Group. Majors in mining machinery; possesses interest in powered support design and manufacturing.

# IMPACT OF OCCUPANTS BEHAVIOR ON BUILDING ENERGY USE: AN AGENT-BASED MODELING APPROACH

Elie Azar<sup>(a)</sup>, Carol C. Menassa<sup>(b)</sup>

<sup>(a)</sup>PhD student at the University of Wisconsin-Madison  
2231 Engineering Hall 1415 Engineering Drive, Madison, WI, 53706  
PH +1 (608) 262-3542; FAX +1 (608) 262-5199

<sup>(b)</sup>M.A. Mortenson Company Assistant Professor in Construction Engineering and Management, Department of Civil and Environmental Engineering at the University of Wisconsin-Madison  
2318 Engineering Hall 1415 Engineering Drive, Madison, WI, 53706  
PH +1 (608) 890-3276; FAX +1 (608) 262-5199

<sup>(a)</sup>[eazar@wisc.edu](mailto:eazar@wisc.edu), <sup>(b)</sup>[menassa@wisc.edu](mailto:menassa@wisc.edu)

## ABSTRACT

Energy modeling techniques used during the design phase of buildings are failing to accurately predict energy use during the buildings operational phase. The main reason behind this discrepancy is the misrepresentation of the role and impact of occupants' energy consumption characteristics on building energy use. Current energy modeling tools represent occupants as static elements, overlooking their different and changing energy use behaviors over time. These behavioral aspects significantly affect energy consumption levels, hence the need for a new energy modeling approach that overcomes the limitations of traditional energy modeling tools by better accounting for occupants actions and behaviors to predict buildings energy consumption. Therefore, this paper presents a new agent-based modeling approach that accounts for the diverse and dynamic energy consumption patterns among occupants, in addition to the potential changes in their energy use behavior due to their interactions with each other and with the building environment.

**Keywords:** building energy consumption, occupant behavior, behavior change, agent-based modeling

## 1. INTRODUCTION

According to the United Nations Environmental Program, buildings are responsible for 30 to 40 percent of global energy use, and a similar percentage of green house gas emissions (UNEP-SBCI 2007). Over the life-cycle of buildings, more than 80 percent of the energy consumed occurs during the operational phase to meet various energy needs such as heating, ventilation, and air conditioning (HVAC), lighting, water heating, and various equipment loads. Moreover, in commercial buildings, more energy is often used during non-working hours than during working hours mainly due to occupancy related actions (Masoso and Groblera 2010).

A number of studies emphasize the role that building occupants play in affecting the energy consumption in buildings and the anticipated savings in energy usage if occupant behavior was influenced over time. A change in behavior could lead to energy savings in excess of 40 percent in some cases, leading to significant economical and environmental benefits (Emery and Kippenhan 2006; Meier 2006; Staats, Leeuwen, and Wit 2000). The Building Sector has therefore a large potential for delivering long-term, significant and cost-effective greenhouse gas emissions reductions, leading governments to shape their policies and programs to reduce energy use during the operational phase of buildings (UNEP-SBCI 2007).

A number of empirical and energy simulation models exist and are widely used in the building sector to predict energy consumption during the operational phase of buildings. Common software programs are eQuest, Energy Plus, Energy-10, and IES (IES 2011; SBCI 2010; EnergyPlus 2009; eQuest 2009). These tools are used at different stages of the life-cycle of buildings, but most importantly during the design phase to optimize the selection and the sizing of mechanical and electrical systems. Consequently, accurate energy predictions are essential to avoid common equipment over-sizing issues, typically resulting in excessive and unnecessary energy use throughout the life of the building under study (Crawley, Hand, Kummert, and Griffith 2008). Furthermore, the need for accurate energy estimates is growing with the increasing demand for Life-Cycle Analysis (LCA) and Life-Cycle Cost Analysis (LCCA), performed to assess the environmental footprint and the economic value of buildings.

However, current energy modeling tools are failing to accurately predict energy use. Their estimates are typically deviating by more than 30 percent from actual energy consumption levels and this difference can even reach a value of 100 percent in particular cases such as

laboratory buildings with high process loads (Yudelson 2010; Dell’Isola and Kirk 2003; Soebarto and Williamson 2001).

Although several limiting factors such as the complexity of buildings, weather, and variations in building schedule might affect the accuracy of energy estimates, studies are showing that these deviations are mainly attributed to misunderstanding and underestimating the important role that the occupant’s energy use characteristics play in determining energy consumption levels; the term ‘occupant’s energy use characteristics’ being defined as the presence of people in the premises and the actions they perform (or do not perform) to influence the level of energy consumption (Hoes, Hensen, Loomans, DeVries, and Bourgeois 2009; Turner and Frankel 2008). As a matter of fact, current commercial energy modeling software are accounting for occupancy in a static way by making simplistic assumptions about building occupants and their energy consumption behavior.

More specifically, the first assumption made by these tools is that all occupants have similar energy use patterns and consume energy at the same rates. This assumption results from the limited number of occupancy-related inputs that can be entered into the models, imposing a common schedule for all occupants and a common energy consumption pattern. Consequently, these tools are ignoring the results of many studies showing that building occupants typically have diverse energy use behaviors, which significantly affect the accuracy of the generated energy use estimates (Hoes, Hensen, Loomans, DeVries, and Bourgeois 2009; Turner and Frankel 2008).

The second assumption made by typical energy models is that occupants’ behavior remains constant over time. Here again, several studies are showing that occupants might change their energy use characteristics over time by adopting more energy efficient practices or on the contrary, adopt bad consumption habits due to the often called ‘rebound effect’. As an example of the ‘rebound effect’, occupants might tend to use more electric lighting following the installation of energy saving bulbs, assuming that their actions will have less impact on the environment. Such types of behavior change negatively impact energy consumption by increasing energy use (Sorrell, Dimitropoulos, and Sommerville 2009; Jackson 2005). On the other hand, factors such as energy conservation campaigns/trainings that encourage energy use reduction or financial incentives that incentivize energy savings typically lead to positive changes in energy consumption behavior. Another important factor is the ‘word of mouth’ or the ‘peer-to-peer’ effect, which is considered to be a very influential channel of communication. Originally used in the marketing field to promote new commercial products, the ‘word of mouth’ factor in terms of energy use is defined as the influence that each occupant exerts on the other occupants sharing the same building environment to change their energy consumption habits. Lastly, feedback techniques have also proven to induce

energy use reduction by providing occupants with information about their energy consumption levels. These better informed occupants typically tend to save energy, especially when they are given access to the energy use levels of the neighboring offices or rooms in the same building (Peschiera, Taylor, and Siegel 2010; Allsop, Bassett, and Hoskins 2007; Staats, Harland, and Wilke 2004).

In brief, there is a need to account for building occupants in a dynamic way by considering and modeling their different energy consumption characteristics in addition to the potential changes in their behavior over time.

## **2. OBJECTIVES**

The main objective of this paper is to present a new energy modeling approach to better predict energy use during the operational phase of commercial buildings. The proposed framework needs to account for different occupants’ energy consumption patterns, their change in behavior over time, and finally simulate the resulting impacts on building energy use. This dynamic modeling of occupants is expected to result in more accurate estimates when compared to traditional energy modeling techniques that overlook the significant impact of occupancy actions and interactions on building energy use.

## **3. AGENT-BASED MODELING**

Several simulation tools were identified in literature that are capable of modeling social and behavioral systems. The most common simulation methods are: Discrete Events (DE), System Dynamics (SD), and Agent-Based Modeling (ABM). While DE and SD are considered to a certain extent as centralized structures requiring the user to define the global system behavior (systems), ABM is decentralized where the modeler defines behavior at an individual level. In this ‘bottom-up’ modeling method, the global behavior emerges as a result of many individuals, each following its own behavior rules, interacting and communicating with each other and with their environment (Gilbert 2008; Edmonds, Hernandez, and Troitzsch 2007; Borshchev and Filippov 2004). More specifically, agent-based models consist of individual agents, commonly implemented in software as objects. Agent objects have states and rules of behavior. Running such models simply amounts to instantiating an agent population, letting the agents interact, and monitoring what happens (Axtell 2000).

For this purpose, this research investigated the use of ABM as a technique capable of simulating almost all behavioral aspect of agents, which represent the building occupants (Gilbert 2008; Axtell 2000). As detailed in the upcoming sections, agents, or building occupants, are assigned attributes that define their specific energy consumption characteristics and patterns. These characteristics might change over time due to external events (e.g., energy conservation

training), or due to the interactions with other occupants (e.g., peer-to-peer influence).

#### 4. METHODOLOGY

Four main steps were required to achieve this research's study objectives: (1) Define different occupancy energy use characteristics and identify factors that cause them to change, (2) build an agent-based simulation model, (3) generate energy use rates for each type of energy use behavior, which are imported into the simulation model, and finally (4) verify the built model through a numerical example.

##### 4.1. Occupants energy use characteristics and influencing factors

The first step consisted of defining different energy consumption behaviors, which is essential for building accurate energy estimation models (Yu, Fung, Haghighat, Yoshino, and Morofsky 2011; Hoes, Hensen, Loomans, deVries, and Bourgeois 2009; Clevenger and Haymaker 2006). Therefore, three categories of occupants were considered. 'High Energy Consumers' (HEC) represent occupants that over-consume energy. 'Medium Energy Consumers' (MEC) are the average energy consuming occupants. Finally, 'Low Energy Consumers' (LEC) represent occupants that use energy efficiently. These assumptions were made based on a study by Accenture (2010) that classified energy consumers in different countries around the world into eight different categories based on their attitude toward energy management programs. However, after discussions with industry professionals, and because of the main focus on energy modeling, it was assumed in this paper that three categories of occupants are adequate to observe differences in energy consumption levels.

Occupants in the three proposed categories mainly differ by the way they use the building energy systems, resulting in different energy use levels. To understand and quantify these differences, an intensive literature review was performed to understand how HEC, MEC, and LEC use each of the main building energy systems such as lighting, equipment/computers, and heating, ventilation, and air conditioning systems (HVAC). For instance, the study by Bourgeois, Reinhart, and MacDonald (2006) was used to identify the variations in the light switching patterns of occupants in commercial buildings. Regarding equipment/computer use, data collected by Sanchez, Webber, Brown, Busch, Pinckard, and Roberson (2007) and Webber, Roberson, McWinney, Brown, Pinckard, and Bush (2006) were used to determine common rates of office equipment use for different occupancy patterns. Similarly, the studies of Davis and Nutter (2010) and Wang, Federspiel, and Rubinstein (2005) were considered to study occupants' presence in their offices, which significantly affects the energy use levels.

After defining different energy consumption characteristics, it was important to identify the factors that might affect these behaviors and cause them to

change over time. Three main factors were chosen for analysis in this paper.

First, energy conservation trainings/workshops were considered, which are informational events that educate occupants about energy saving practices and encourage them to reduce their energy consumption. So, after attending such events, a portion of the occupants is expected to reduce its energy consumption. This is translated into a conversion of some HEC to MEC, and some MEC to the LEC category.

Second, the 'rebound effect' represents the opposite type of behavior change, where occupants counter react to energy reduction initiatives and increase their energy use. In that case, a portion of the LEC becomes MEC, and some MEC convert to HEC.

Finally, the peer-to-peer influence represents the influence that occupants sharing a certain building environment have on each other to change their energy consumption patterns. So, each category of occupants (HEC, MEC, and LEC) might influence occupants from the other categories to change their behaviors, and adopt its energy consumption patterns. So, in this type of interaction, three possible changes in behavior or conversion of occupants can occur as illustrated in Figure 1. The first line shows the case where HEC are influencing MEC and LEC. This change in characteristics is gradual where some MEC get converted to HEC, and some LEC become MEC. Similarly, the second and third lines of Figure 1 respectively show the cases where MEC and LEC are actively converting other occupants. The influence of each category on the others depends on the number of persons in this category in addition to its Level of Influence (LI), which is entered by the user. More details on these conversions are provided in the upcoming sections.



Figure 1: 'Peer-to peer' interactions and conversions

##### 4.2. Agent-Based model

The next step consisted of building an agent-based model to simulate the above-mentioned interactions and predict energy use.

The agent-based software that was chosen for this research is 'Anylogic', which is widely used in the industry (XJ Technologies 2009; Borshchev and Filippov 2004). The choice of 'Anylogic' was mainly due to its Java-based environment that allows the user to develop custom Java codes, and integrate them in pre-built simulation blocks (XJ Technologies 2009). This was essential in this research to optimize and customize the proposed model in order to simulate the complex behavioral aspects of building occupants. The

proposed model flowchart is shown in Figure 2 where a five-step iterative process was defined.

First, energy consumption rates for HEC, MEC, and LEC are imported into the agent-based model to be used in the following stages (Step 1). Details on obtaining these rates are shown in the upcoming section.

Step 2 simulates the interaction of agents and the potential change in behavior due to the 'word of mouth' or peer-to-peer effect. So, for the first time step (e.g., first month), HEC, MEC, and LEC sharing the same simulation environment (e.g., office) interact and try to influence each others to change behavior. The chances of success for each category are dependent on two variables: (1) the number of persons in this category at the current time step, and (2) its Level of Influence (LI) on other categories. Initially, three LIs are entered by the user before the beginning of the simulation:  $LI_{HIGH}$ ,  $LI_{MEDIUM}$ , and  $LI_{LOW}$ . For instance, a  $LI_{HIGH}$  of 5 percent/person/month means that each HEC person has a 5 percent chance of successfully converting another occupant every month. So, a high number of HEC at a specific time step or a high  $LI_{HIGH}$  results in high pressure on the other occupants sharing the same environment to increase their energy consumption and become HEC. At the end of Step 2, the model updates and stores the numbers of HEC, MEC, and LEC for the current time step.

Next, the model moves to Step 3 to check if an energy conservation event (e.g., training, workshop, etc.) is scheduled for this time step. Here again, the user can schedule 'events' and specify their effectiveness before running the model, then track their impact on occupancy behavior and energy use. For instance, an event scheduled for month 12 with an efficiency of 30 percent result in the conversion of 30 percent of HEC to MEC and 30 percent of the MEC to LEC when the simulation time reaches 12 months. So, for each time step, the model checks if any 'event' is scheduled and updates the number of HEC, MEC, and LEC.

In step 4, the model checks if a rebound effect is occurring at this time step, leading to a change in behavior and an increase in energy consumption. In this case, the conversion of occupants occurs towards the high energy consumption categories, where some LEC become MEC and some MEC convert to HEC. Here again, the occurrence time and the effectiveness of the rebound effect is previously specified by the user.

The updated numbers of HEC, MEC, and LEC are then combined to the energy consumption rates obtained from Step 1, and total energy consumption levels are calculated for the current time step (Step 5). Once this iteration is completed, the model moves to the next time step and keeps repeating the cycle until the total simulation time is reached.

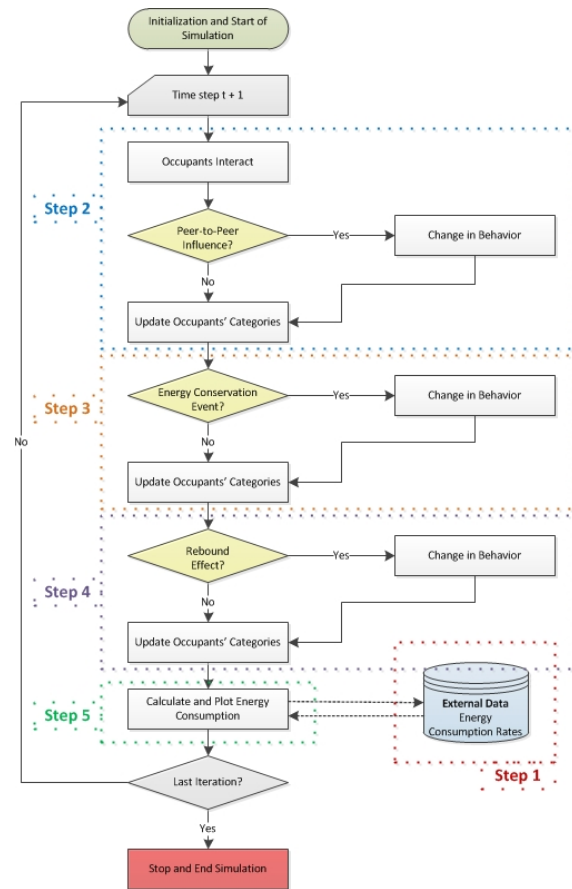


Figure 2: Model's flowchart

#### 4.3. Energy use rates for different types of energy use behavior

This section presents the proposed method that was used to obtain energy consumption rates for HEC, MEC, and LEC using traditional energy modeling software (e.g., EnergyPlus, eQuest, etc.). In this study three sets of simulations were particularly needed, each set having specific inputs representing the different energy characteristics of the three defined categories of occupants. In general, two types of inputs are required to build these models: (1) building related inputs which are the same for all the simulations, and (2) different occupancy related inputs that will lead to the three different energy use rates.

So, the first step consists of defining the building environment under study accommodating occupants who consume energy over time through their daily activities. Common inputs related to the building under study are determined such as the building type and size, floor plan layout, construction materials, HVAC equipment, lighting systems, miscellaneous equipment (e.g., computers), and hot water supply (Dell'Isola and Kirk 2003).

The next step consists of defining occupancy related parameters, leading to the difference between the energy consumption levels of HEC, MEC, and LEC. These differences are generated in traditional energy

simulation software by varying occupancy related parameters such as ‘equipment rates of use’ or ‘building operating schedules’. For instance, to illustrate how HEC leave their computers ON more frequently than MEC and LEC during building non-operating hours, parameters such as equipment use rates can be increased for the HEC simulation to represent this over-use of energy. Similarly for the rest of the energy consumption sources (e.g., lighting, HVAC, etc.), specific inputs are used to customize each of the three simulations and translate the differences in behavior into differences in energy consumption levels. The obtained rates are then imported into the simulation model as was shown in Step 1 of Figure 2, to be used for the total energy use calculation of the building under study.

It is important to note that the input parameters are determined on a case by case basis, depending on the building environment and on the specific occupancy characteristics. A detailed example of the proposed method is presented in the following section.

## 5. MODEL VERIFICATION THROUGH A NUMERICAL EXAMPLE

This section highlights the capabilities of the model through a numerical example and presents some of the sensitivity analyses that were performed to verify the model.

### 5.1. Environment description

First, an experimental energy simulation model was built for the purpose of this study using eQuest. This conceptual 2000 square feet (186 square meters), two-story university type building, is located in the city of Madison, Wisconsin (See Figure 3). The first floor contains two open space offices accommodating 16 and 10 graduate students respectively. The second floor contains one classroom with a maximum capacity of 50 students.

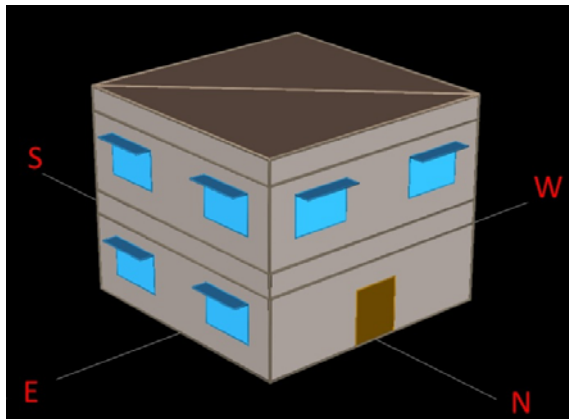


Figure 3: Building environment in eQuest

### 5.2. Inputs and assumptions

Six energy consumption sources were considered in this example: (1) HVAC heating, (2) HVAC cooling, (3) area lighting, (4) task lighting, (5) equipment

(computers), and (6) hot water supply. For each occupant categories (HEC, MEC, and LEC), energy consumption rates were obtained using eQuest by running different experiments using specific inputs that reflect their behavioral differences (Refer to Section 4.3 for more details about the method used). More specifically, three types of inputs were varied in the simulations to differentiate between the energy use behaviors of HEC, MEC, and LEC: (1) lighting schedules, (2) equipment schedules (computers), and (3) hot water use.

In this example, the parameters were first derived based on several studies in literature on occupants and the different energy consumption patterns they adopt (Masoso and Grobler 2010; Mahdavi, Mohammadi, Kabir, and Lambeva 2008; Sanchez, Webber, Brown, Busch, Pinckard, and Roberson 2006; Webber, Roberson, McWinney, Brown, Pinckard, and Bush 2006). Some parameters were also suggested by eQuest, which follows the California’s Title 24 building code requirements (eQuest 2009). Additional occupancy related assumptions had to be made about the differences between the high, medium, and low energy use levels. These hypotheses were then reviewed and confirmed by industry experts and were as a result considered acceptable. A summary of the used parameters is shown in Table 1.

Table 1: Input Parameters Variation

	HEC	MEC	LEC
<b>Lighting Schedules</b>	30% of time running after building operating hours	10% of time running after building operating hours	0% of time running after building operating hours
<b>Equipment Schedules</b>	60% of time running after building operating hours	20% of time running after building operating hours	0% of time running after building operating hours
<b>Hot Water Use</b>	20% more than MEC	1.20 Gallon/Pers/Day	20% less than MEC

So, by using the above-shown sets of parameters, energy consumption rates for HEC, MEC, and LEC were obtained, and then imported into the proposed agent-based model.

### 5.3. Agent-Based model outputs

The next step consisted of executing the steps of the flowchart shown in Figure 2 to simulate occupancy and predict the building’s energy consumption levels. An example of the proposed model’s output is shown in Figure 4, showing the change in behavior in one of the offices of the building (upper graph), along with the resulting changes in energy consumption (lower graph). Similar graphs are generated for each of the building’s rooms, allowing to track occupancy behavior and energy consumption changes both on a room and on a whole building level.

At the start of the simulation, and for this particular office, 5 of the students were assumed to be HEC, 6



MEC, and 5 LEC, reflecting average energy use characteristics. LEC were assumed to have a higher LI than the other categories, showing a strong influence of energy efficient occupants on others. The occupants of the studied room were also assumed to attend an energy conservation event at month 12 encouraging them to reduce their energy consumption and adopt energy saving practices. In this example, the efficiency was set at 50 percent, which means that after attending the event, 50 percent of the occupants were expected to reduce their energy consumption. Also, a rebound effect with a 25 percent efficiency was scheduled for month 24, resulting in an increase in the occupants' energy consumption, and a conversion from the LEC and MEC categories to the MEC and LEC categories respectively. It is important to note that the specific choice of parameters in this example was made to highlight the capabilities of the model. All of the parameters can be specified by the user through a user-friendly interface to customize the model and better represent the specific building and occupants under study.

Figure 4 illustrates how the occupants of the office were changing behavior over the 36 months simulation time due to their interactions with each other, and also due to the external events that were scheduled for months 12 and 24. So, by sharing the same office, the 16 occupants were influencing each other according the LIs that were entered for this particular example. This resulted in a change in behavior represented by the change in the numbers of HEC, MEC, and LEC over time. Moreover, the energy conservation event at month 12 caused a sudden conversion of occupants towards the LEC with their number increasing from 6 to 9. The opposite type of conversion occurred at month 24, with a drop in the number of LEC and an increase in HEC and MEC.

These changes in energy behavior were reflected on the energy consumption levels of the office. More specifically, as the room occupants were becoming LEC, energy consumption (electric and gas energy combined) was decreasing, with a total drop of 12 percent as shown in the lower graph of Figure 4. It is important to note that over the 36 months studied period, the peaks of the total energy consumption curve occurred during the winter season. This was expected in a cold weather region such as Wisconsin, where heating loads are significant and typically exceed air-conditioning cooling loads required during the warmer seasons.

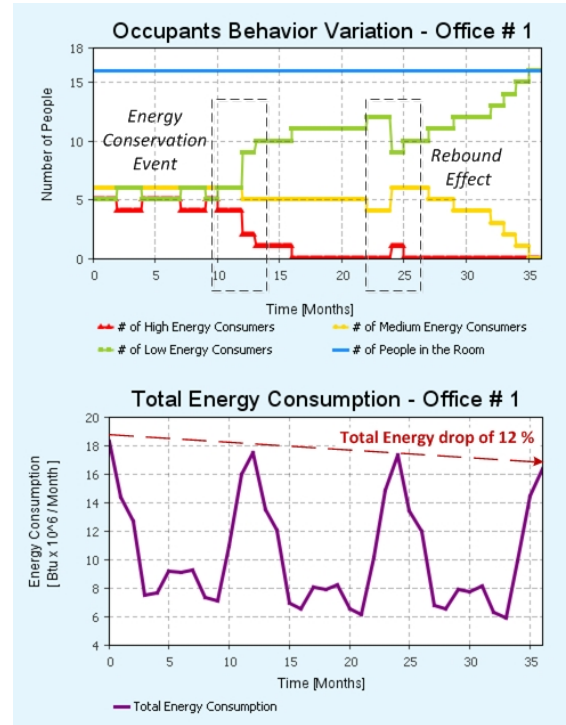


Figure 4: Agent-Based model output (Office)

As was previously mentioned, similar graphs are generated for all of the other rooms in the building. For instance, Figure 5 shows the change in behavior and the resulting change in energy consumption for the classroom area of the building in Figure 3. Unlike in an office where occupants interact on a daily basis and influence each other, students in a classroom have less influence on each other to change behavior. As result, it was assumed in this example that the change in behavior of occupants can only be induced by independent events such as energy conservation events or the rebound effect. This is illustrated in Figure 5 where changes in behavior only occurred when events were scheduled at months 12 and 24.

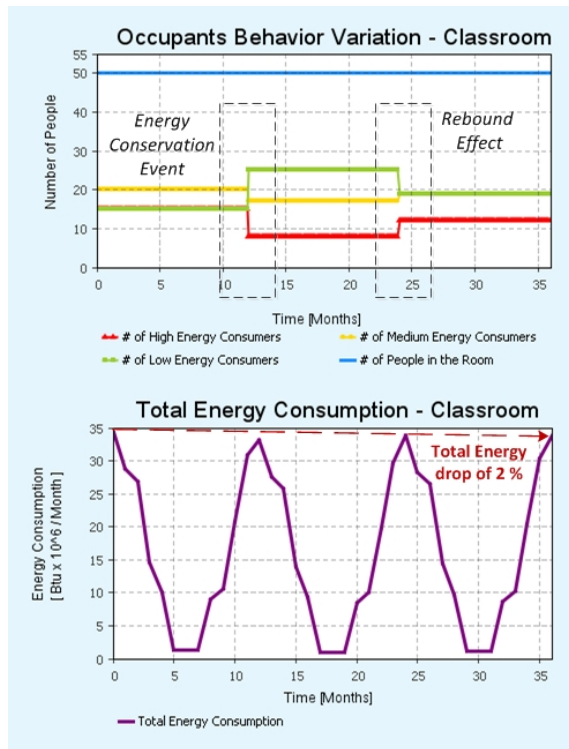


Figure 5: Agent-Based model output (Classroom)

In this example, the relatively low drop in energy consumption of 2 percent for the classroom can be attributed to two main reasons. First, by the end of the simulation time, only 19 of the 50 occupants were LEC, hence the remaining 31 occupants were consuming energy at medium or high levels. Second, the occupants of a classroom have a low level of control over the classroom energy consumption when compared to office occupants that use computers, task lights, etc. So, even if behavior changes in a classroom, the resulting impact on energy use remains limited.

So, by simulating occupancy behavior in all of the rooms of the studied building, the proposed model performs a whole building analysis and predicts the total energy consumption levels.

#### 5.4. Sensitivity analyses

This section shows an example of the sensitivity analyses that were performed to test the model's reaction to changes in input parameters. In accordance with the study's objectives, a specific emphasis was put on the role played by occupants by varying parameters related to their behavior and interactions, and ultimately tracking the resulting changes in energy use.

Four different scenarios were considered in this analysis, and compared to the base case model built using eQuest (See Table 2). This base case represents traditional energy modeling software programs that do not allow for different nor changing occupancy characteristics throughout the simulation time. The first scenario represents the extreme case where all of the occupants are HEC with no peer-to-peer influences or

events scheduled to change the high energy consumption behavior of the occupants. In scenario 2, the occupants are evenly divided between HEC, MEC, and LEC, however, the occupants are expected to convert to the HEC category since the highest LI was given to HEC and a rebound effect is scheduled for the 12<sup>th</sup> month. Scenario 3 is the opposite of scenario 2 with the highest LI given to the LEC, in addition to an energy conservation event scheduled for the 12<sup>th</sup> month. Finally, scenario 4 represents an extreme case with all occupants being LEC and no peer-to-peer influence or events to change behavior.

Table 2: Sensitivity analysis inputs variation

	Initial Energy Use Behavior of Occupants	Peer-to-peer Levels of Influence (LI)	Energy Conservation Event	Rebound Effect
eQuest Base Case	All MEC	none	none	none
Scenario # 1	All HEC	none	none	none
Scenario # 2	Evenly divided between HEC, MEC, and LEC	Highest for HEC	none	Scheduled at t = 12 months
Scenario # 3	Evenly divided between HEC, MEC, and LEC	Highest for LEC	Scheduled at t = 12 months	none
Scenario # 4	All LEC	none	none	none

The results of this sensitivity analysis are summarized in Figure 6 where Scenario 1 resulted in the highest energy use level that is 13 percent higher than the eQuest base case. On the other hand, scenario 4 resulted in the lowest energy consumption level, 17 percent lower than the base case. These results were expected since Scenarios 1 and 4 represented the extreme cases with all the occupants being HEC and LEC respectively throughout the simulation time. Scenarios 2 and 3, in turn, showed variations in energy consumption by plus 8 percent and minus 11 percent. These numbers were also expected since HEC had the highest LI in Scenario 2 in addition to a rebound effect scheduled, while in Scenario 3, LEC were advantaged with a high LI and an energy conservation event. Finally, by comparing the two most extreme cases, Scenarios 1 and 4, a total net difference of 30 percent was observed.

The results from this sensitivity analysis in addition to the other analyses that were performed confirmed that the model is behaving in logical manner to changes in inputs. As a result, the model was considered verified.

Moreover, the significant variations in energy use that were observed confirmed the importance of the impact of occupants on building energy estimation.

More variations are also expected when more types of occupants' interactions are considered, and more control is given for occupants over the building's energy systems.

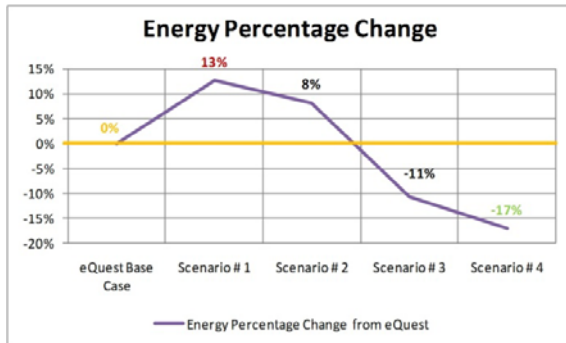


Figure 6: Percentage change in energy from eQuest

## 6. CONCLUSION

In conclusion, traditional energy simulation software programs are accounting for occupancy in a simplistic way, overlooking the impact of occupants' behavior and actions on building energy use. This is causing their energy estimates to significantly deviate from actual energy consumption levels (Yudelso 2010; Clevenger and Haymaker 2006; Soebarto and Williamson 2001).

This paper presented a new agent-based modeling approach to energy estimation by modeling occupancy in a dynamic way, accounting for both the differences between occupants' energy use characteristics and the changing aspect of these characteristics over time. The proposed approach was then tested and verified through a numerical example, which showed that significant changes in energy predictions can be obtained when occupancy is modeled in a dynamic way.

After successfully verifying the agent-based model, the next step is to validate it by comparing its energy estimates to numbers from actual buildings in operation. This is essential to make sure that the proposed model simulation numbers are consistent with the actual electricity, gas, and water consumption levels of the specific building under study. Once validated, the model can then be disseminated for real life applications.

Finally, the proposed methodology adds a new and important dimension to the energy modeling field by accounting for occupants and their impact on energy use. Consequently, this method overcomes the main limitations of current energy modeling software and results in more realistic and accurate energy predictions, which are essential to the design of more green and energy efficient buildings.

## REFERENCES

- Accenture (2010). Understanding Consumer Preferences in Energy Efficiency. Available from: [http://www.accenture.com/NR/rdonlyres/AA01F184-9FFC-4B63-89BD-7BFA45397F13/0/Accenture\\_Utility\\_Study\\_What\\_About\\_Consumers\\_Final.pdf](http://www.accenture.com/NR/rdonlyres/AA01F184-9FFC-4B63-89BD-7BFA45397F13/0/Accenture_Utility_Study_What_About_Consumers_Final.pdf) [accessed 2 July 2011].
- Allsop, D. T., Bassett, B. R., Hoskins, J. A. 2007. Word-of-Mouth Research: Principles and Applications. *J. Advert. Res.*, 47(4), 398-411.
- Axtell, R. 2000. *Why Agents?: On the Varied Motivations for Agent Computing in the Social Sciences*, Center on Social and Economic Dynamics, Brookings Institution, Washington, DC, USA.
- Borshchev, A., and Filippov, A. 2004. From system dynamics and discrete event to practical agent based modeling: Reasons, techniques, tools. *Proc., 22<sup>nd</sup> International Conference of the System Dynamics Society*. July 25-29, Oxford, United-Kingdom.
- Bourgeois, D., Reinhart, C., MacDonald, L. (2006). Adding advanced behavioral models in whole building energy simulation: a study on the total energy impact of manual and automated lighting control. *Energy and Buildings Journal*, 38 (7), 814-823.
- Clevenger, C. and Haymaker, J. (2006). The Impact of the Occupant on Building Energy Simulations, *Proceedings of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, Montreal, Canada.
- Crawley, D. B., Hand, J. W., Kummert, M., Griffith, B. T. 2008. Contrasting the Capabilities of Building Energy Performance Simulation Programs. *Build. Environ.*, 43(4), 661-673.
- Davis, J. A. and Nutter, D. W. (2010). Occupancy diversity factors for common university building types, *Energy and Buildings Journal*, 42 (9), 1543-1551.
- Dell'Isola, A. J., and Kirk, S. J. 2003. *Life Cycle Costing for Facilities*, Reed Construction Data, Kingston, MA, USA.
- Edmonds, B., Hernández, C., Troitzsch, K. G. 2007. *Social Simulation Technologies, Advances and New Discoveries*, Information Science Reference, Hershey, PA, USA.
- Emery, A. F., and Kippenhan, C. J. 2006. A Long Term Study of Residential Home Heating Consumption and the Effect of Occupant Behavior on Homes in the Pacific Northwest Constructed According to Improved Thermal Standards. *Energy*, 31(5), 677-693.
- EnergyPlus 2009. Input/Output Reference: The Encyclopedic Reference to EnergyPlus Input and Output. <http://apps1.eere.energy.gov/buildings/energyplus/pdfs/inputoutputreference.pdf> [accessed 30 June 2011].
- eQuest 2009. Introductory Tutorial, version 3.63. Available from: [http://doe2.com/download/equest/eQ-v3-63\\_Introductory-Tutorial.pdf](http://doe2.com/download/equest/eQ-v3-63_Introductory-Tutorial.pdf) [accessed 30 June 2011].
- Gilbert, G. N. 2008. *Agent-Based Models*. London, United-Kingdom: Sage Publications, Inc.
- Hoes, P., Hensen, J. L. M., Loomans, M., De Vries, B., Bourgeois, D. 2009. User Behavior in Whole Building Simulation. *Energy Build.*, 41(3), 295-302.
- Integrated Environmental Solutions (IES) 2011. Introducing the virtual environment. Available from: <http://www.iesve.com/software> [accessed 30 June 2011].
- Jackson, T. 2005. *Motivating Sustainable Consumption: A Review of Evidence on Consumer Behaviour and Behavioural Change*, Centre for Environmental Strategy, University of Surrey, Surrey, United-Kingdom.
- Mahdavi, A., Mohammadi, A., Kabir, E., Lambeva, L. (2008). Occupants' Operation of Lighting and Shading Systems in Office Buildings. *Journal of Building Performance Simulation*, 1(1), 57-65.
- Masoso, O. T., and Grobler, L. J. 2010. The Dark Side of Occupants' Behaviour on Building Energy use. *Energy Build.*, 42(2), 173-177.
- Meier, A. 2006. Operating Buildings during Temporary Electricity Shortages. *Energy Build.*, 38(11), 1296-1301.
- Peschiera, G., Taylor, J. E., Siegel, J. A. 2010. Response-Relapse Patterns of Building Occupant Electricity Consumption Following Exposure to Personal, Contextualized and Occupant Peer Network Utilization Data. *Energy Build.*, 42(8), 1329-1336.
- Sanchez, M., Webber, C., Brown, R., Busch, J., Pinckard, M., Roberson, J. (2007). Space Heaters, Computers, Cell Phone Chargers: How Plugged In Are Commercial Buildings?. *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings*, Less is More, En Route to Zero Energy Buildings.
- Sustainable Building Industry Council (SBICI) 2010. Energy-10@ Software. <http://www.sbicouncil.org/energy-10-software> [accessed 6 February 2011].
- Soebarto, V. I., and Williamson, T. J. 2001. Multi-Criteria Assessment of Building Performance: Theory and Implementation. *Build. Environ.*, 36(6), 681-690.
- Sorrell, S., Dimitropoulos, J., Sommerville, M. 2009. Empirical Estimates of the Direct Rebound Effect: A Review. *Energy Policy*, 37(4), 1356-1371.
- Staats, H., Harland, P., Wilke, H. A. M. 2004. Effecting Durable Change: A Team Approach to Improve

- Environmental Behavior in the Household. *Environ. Behav.*, 36(3), 341-367.
- Staats, H., van Leeuwen, E., Wit, A. 2000. A Longitudinal Study of Informational Interventions to Save Energy in an Office Building. *J. Appl. Behav. Anal.*, 33(1), 101-104.
- Turner, C., and Frankel, M. 2008. *Energy Performance of LEED for New Construction Buildings*, New Buildings Institute, Vancouver, WA, USA.
- United Nations Environment Programme. 2007. Buildings can play key role in combating climate change. Available from: <http://www.unep.org/Documents/Multilingual/Default.asp?DocumentID=502&ArticleID=5545&l=en> [accessed 8 May 2011].
- Wang D., Federsipel, C. C. and Rubinstein, F. (2005). Modeling occupancy in single person offices, *Energy and Buildings Journal*, 38 (2), 121-126.
- Webber, C. A., Roberson, J. A., McWinney, M. C., Brown, R. E., Pinckard, M. J. and Bush, J. F. (2006). After-hours power status of office equipment in the USA. *Journal of Energy*, 31(14), 2823-2838.
- XJ Technologies (2009). Anylogic Overview. Available from: <http://www.xjtek.com/anylogic/overview/> [accessed 10 June 2011].
- Yu, Z., Fung, B. C. M., Haghighat, F., Yoshino, H., Morofsky, E. (2011). A Systematic Procedure to Study the Influence of Occupant Behavior on Building Energy Consumption. *Energy Build.*, 43(6), 1409-1417.
- Yudelson, J. 2010. *Greening Existing Buildings*. New York: NY Green Source/McGraw-Hill.

## AUTHORS BIOGRAPHY

**Elie Azar** is a PhD student in the Civil and Environmental Engineering Department at the University of Wisconsin-Madison. He received his Bachelor degree in Mechanical Engineering from Polytechnique Montreal, and his Master's degree in Civil and Environmental Engineering from the University of Wisconsin-Madison.

**Carol C. Menassa** is an M.A. Mortenson Company Assistant Professor in the Civil and Environmental Engineering Department at the University of Wisconsin-Madison. She received a Master of Science in Finance and a Ph.D. in Civil Engineering from the University of Illinois at Urbana-Champaign.

# SHEETING PROCESS MODELLING AND RHEOLOGICAL ANALYSIS OF AN OLIVE-OIL-EMULSION-BASED PUFF PASTRY

Francesca R. Lupi<sup>(a)</sup>, Noemi Baldino<sup>(b)</sup>, Lucia Seta<sup>(c)</sup>, Domenico Gabriele<sup>(d)</sup> Bruno de Cindio<sup>(e)</sup>

<sup>(a)</sup> Laboratory of Rheology and Food Engineering, Department of Engineering Modeling, University of Calabria, Cubo 39 C, Via P. Bucci, 87030 Arcavacata di Rende (CS), Italy

<sup>(a)</sup> [francesca.lupi@unical.it](mailto:francesca.lupi@unical.it) <sup>(b)</sup> [noemi.baldino@unical.it](mailto:noemi.baldino@unical.it) <sup>(c)</sup> [lucia.seta@unical.it](mailto:lucia.seta@unical.it), <sup>(d)</sup> [d.gabriele@unical.it](mailto:d.gabriele@unical.it), <sup>(e)</sup> [bruno.decindio@unical.it](mailto:bruno.decindio@unical.it)

## ABSTRACT

Different foodstuffs, like pastries, are commercialised in the form of laminated sheets; sheeting is, in several cases (i.e. puff pastries), the main process step in which fat/dough rheological matching is achieved. Pastries are baked goods produced to complement the flavour of the fillings and to provide them with a casing. Their ingredients include hard fats hydrogenated with catalytic processes. Unfortunately, these processes lead to the formation of trans fats dangerous for consumers' health. Therefore, hydrogenated fats are nowadays replaced by healthy fats, like olive oil, properly structured to be used as solid fat replacers.

The present work deals with the modelling of the sheeting process of olive-oil-based puff pastries; this allows their final height to be predicted. The rheological parameters necessary to set the model were obtained by the analysis of a pastry produced with water-in-olive-oil emulsions, structured with an organogelator agent, and optimised from a rheological point of view.

Keywords: rheology, puff pastry, W/O emulsions, organogels, sheeting, rolling.

## 1. INTRODUCTION

Rolling or sheeting between co-rotating rolls is used by baking industry as a dough-forming process for a wide range of products, such as cookies, crackers, pizza, bread and pastry (Mitsoulis E. et al 2009). In many operations rolling is well understood: metal sheeting, rolling of polymer, called calendaring, but the roller extrusion of food materials, however, is a broad area of industry activity in which current practice still frequently relies on approximations and expensive pilot plant testing (Peck M. C. et al 2006). This is due to the need to have good experimental values of the material properties (a heterogeneous system), but as well as a constitutive equation capable to fit the material behaviour. Being the material composite and showing time dependency the constitutive equations are often proposed in a differential form, thus a simultaneous integration with the conservation principles is needed

The aim of the current study is to model both from rheological and numerical point of view the puffy pastry dough, to create a tool capable to avoid trial and error

approach to the determination of the process variables.

Puff pastry is a dough spread with solid fats and repeatedly folded and sheeted. Thus, it consists of alternating layers of dough and solid-like fat, designed to obtain the proper rheological matching. In fact, the characteristic friable texture is due to the vapour released by either butter or dough during cooking ('puffing' process). Pastry fats (margarine) are W/O emulsions structured in the oil phase. They derive their consistency from a fat crystal network of saturated triglycerides (TAGs) (Cavillot et al 2009).

Traditionally, saturated TAGs can be obtained by hydrogenation, which also produces undesired trans fatty acids (TFAs), considered as the main responsible of heart diseases. Because of this, there is a growing market interest on the design of new pastries produced with a low saturated fat content, and, most of all, without trans fatty acids (Simovic et al 2009; de Cindio B. and Lupi F.R., 2011). It should be noticed that saturated fat substitution needs a different rheological fitting capable to maintain the desired consistency. One of the most promising techniques used to achieve the proper rheology of pastry fats, is the oil phase structuration with organogels (Marangoni 2009). In this work, different structured water-in-olive-oil emulsions were produced. Emulsions optimisation was achieved evaluating the effect of the organogelator amount and the water content on the system rheology, in order to reach properties similar to those of a commercial margarine.

Once the proper fat mechanical characteristics were optimised, a puff pastry was obtained with it. A rheological characterisation was done in order to obtain the setting parameters for the puff pastry sheeting process. The rheological and fluidodynamic modelling of the last step of the sheeting process, is necessary to control the final thickness of the pastry exiting from the rolls because the presence of a significant die swell effect, due to the normal stresses developed into the material during process. The velocity and pressure profiles inside the pastry (considered as an homogeneous material) passing thorough the rolls were evaluated according to a modified Upper Converted Maxwell Model (UCM) constitutive equation.

The sheeting process is traditionally modelled by the simplified Gaskell theory (Tanner R. I., 2000),



according to which, the elastic effects of viscoelastic materials can be neglected, and therefore the normal stresses acting on the rolls are ignored. In addition it is assumed that the lubrication theory holds, thus a unique component of the velocity vector is considered. On the contrary according to the viscoelastic material behaviour, when sheeting pastries, the normal stresses must be taken into account. In this work a more general model is presented with the insertion of normal stresses and removing also the other simplification usually considered in the Gaskell theory. Thus experimental determination and numerical simulation are considered to predict the final height of a pastry from which final desired organoleptic properties depend. The model was solved by the finite element method by using the commercial software COMSOL Multiphysics 3.5.

## 2. MATERIALS AND METHODS

### 2.1. Emulsions ingredients and manufacture

The raw materials used for emulsions preparation were: distilled water and a virgin olive oil (De Santis, Italy) as main constituents of the two phases, cocoa butter (Icam S.P.A., Italy), monoglycerides of fatty acids (Myverol 18-04 K, Kerry Group, Ireland) as organogelator, and NaCl (Panreac, Spain), at 0.1 M aqueous concentration, in order to identify the type of emulsion (W/O or O/W) through electric conductivity measurements.

The structured oil phase was prepared by adding contemporary the cocoa butter and the organogelator to the oil at 70°C, under continuous agitation. Afterwards, a fast cooling process to the gently stirred oil phase was performed, quenching it in a thermostatic cold bath. Different emulsions were produced varying the total organogelator concentration and the aqueous phase content. All samples were prepared leaving the cocoa butter/oil ratio constant. The aqueous phase was obtained by dissolving the NaCl in water, at room temperature, and later mixing it to the cold oil phase using a rotor-stator turbine at 7600 rpm for 30 s (Ultra-Turrax T 50, IKA, Germany).

The composition of each sample and its identification is reported in Table 1.

Table 1: Emulsions ID and composition [% wt/wt]

Sample ID	Oil	Water	Cocoa Butter	Myverol
Emu 1	75	12	10	3
Emu 2	68	20	9.1	2.9
Emu 3	74.12	12	9.88	4
Emu 4	76.76	12	10.24	1

Emulsion– stability was also studied, evaluating the evolution of their rheological properties with time (all samples were studied at 24h, 48h, 72h and 1 week from preparation). An emulsions having a similar rheological behaviour, compared to those of a commercial pastry margarine (i.e. “Vallè”, France), was used to prepare a hand-made puff pastry. Moreover, as benchmark, a commercial puff pastry (i.e. Cà Bianca, Italy) was also studied in order to compare the differences between its

characteristics and those of the hand-made pastry. For the preparation of the puff pastry feed sheets, the dough was kept from the mixed batch immediately after mixing and rolled out with a wooden cylinder on a kitchen board according to the so-called “Danish method” (Harte 2003). After this step the sheet was reduced to a lower thickness by a lab scale sheeting machine.

### 2.2. Rheological characterisation

The rheological characterisation of the emulsions (the proposed fat substitute) was performed with a Small Amplitude Oscillation Analysis (SAOTs) at 20°C using a controlled-strain rheometer ARES-RFS (TA Instruments, U.S.A.) equipped with serrated plates ( $\varnothing=50$  mm, gap 2 mm). Frequency sweep tests were performed within the linear viscoelastic region, in the frequency range 0.1-10 Hz.

The characteristics of both commercial and hand-made pastry were analysed with a controlled-stress rheometer DSR500 (Rheometric Scientific, USA), using a serrated plate geometry ( $\varnothing=25$  mm, gap 1.5 mm). The viscous flow behaviour of pastries was determined in the shear rate range  $0.0001-1$  s<sup>-1</sup>, and stress relaxation tests were performed at 20°C. Owing to the slippage problems, steady shear viscosity was determined only at very low shear rates with creep tests (at 20°C and in the range of stresses 400-1300 Pa).

This complete characterization was necessary because, while the SAOTs give information about the mechanical characteristics of samples, allowing their rheological optimisation, the flow characterisation and the relaxation time (evaluated from stress relaxation tests) were necessary to obtain the material parameters introduced by constitutive equation discussed below.

By means of these data it was possible to determine an optimum recipe of fat replacer to produce a pastry similar to the commercial margarine.

A Stress Relaxation test of puff pastries was performed with the same rheometer equipped with a serrated plate geometry ( $\varnothing=25$  mm) at 20°C, imposing a constant stress value within the linear viscoelastic region. Thus a relaxation time was determined.

### 2.3. Rheological data analysis and modelling

Structured emulsions can be described as weak systems, consequence of the development of a three-dimensional network of organogels molecules and cocoa butter crystals with rheological “units” connected by weak bonds. In these cases, the structured oil phase entraps both the oil solvent and the water drops, yielding to a stable water-in-oil emulsion similar to other organogel products reported in the literature (Marangoni 2009).

Owing a linear trend in log-log of both moduli, the rheological data can be interpreted by a “weak-gel model” (Gabriele et al. 2001):

$$G^*(T, \omega) = \sqrt{(G')^2 + (G'')^2} = A(T) \cdot \omega^{\frac{1}{2(T)}} \quad (1)$$

where  $G^*$  is a complex modulus,  $\omega$  the oscillation frequency,  $T$  the temperature and  $z$  the network extension, related to the number of interacting rheological units within the 3-D network, and  $A$  is the strength of the interactions. When  $A$  increases, the interaction forces within the network increase, whilst a high  $z$  value indicates a large number of interacting units cooperating and increasing the network connectivity.

Creep data, reported in terms of compliance versus time, were analysed by using a one-element Kelvin – Voigt viscoelastic model (Steffe, 1996)

$$J(t) = J_0 + J_1 \left( 1 - e^{-\frac{t}{\lambda_1}} \right) + \frac{t}{\eta} \quad (2)$$

where  $J_0$  is the instantaneous compliance,  $J_1$  is the retarded compliance,  $\lambda_1$  is the retardation time of the Kelvin component, and  $\eta$  is the viscosity. Thus the shear viscosity can be computed as the reciprocal value of the slope of the linear behaviour attained after the retardation time for any applied stress  $\tau_0$ , and reported as shear rate function as:

$$\dot{\gamma} = \frac{\tau_0}{\eta} \quad (3)$$

To extend the experimental range, dynamic and shear parameters were related through the so-called Cox–Merz rule ~~applies~~ (Cox and Merz, 1958) stating, as a rule of thumb, that shear and dynamic complex viscosity for polymer systems coincide:

$$\dot{\gamma} = \omega \quad (4)$$

where  $\dot{\gamma}$  is the shear rate and  $\omega$  the oscillation frequency. According to the literature (Peressini et al. 2002) for food systems it was found that the equivalence between complex and shear viscosity ~~can~~ holds when assuming

$$\dot{\gamma} = a_{\dot{\gamma}} \omega \quad (5)$$

where  $a_{\dot{\gamma}}$  is a shift factor of order of magnitude 0.01. Thus the complex viscosity can be computed as

$$\eta^*(\omega) = \frac{G^*(\omega)}{\omega} \quad (6)$$

and it can be reported in the same plot together with the steady shear viscosity. Both rheological parameters showed a very similar slope, thus it was possible to overlap those curves by a horizontal shift, estimating  $a_{\dot{\gamma}}$ . The obtained flow curve exhibited a power-law behaviour in the investigated shear rate range

$$\tau = k \cdot \dot{\gamma}^n \quad (7)$$

in which  $k$  can be considered as an index of the material

“consistency” and  $n$ , the flow index, can be calculated as the slope of the curve and its value is always positive and lower than one.

Finally, stress relaxation data were analysed in order to evaluate the relaxation time  $\lambda$  of the material, according to a single element Maxwell model. The relaxation of stress after the application of a sudden strain can be described by the following equation (Steffe, 1996).

$$\sigma = \sigma_0 \cdot \exp\left(-\frac{t}{\lambda}\right) \quad (8)$$

From the initial slope it is possible to find the  $\lambda$  value.

#### 2.4. Sheeting process modelling

The flow is governed by the conservation equation of mass and momentum under isothermal conditions that for an incompressible material read:

$$\rho \cdot \left( \frac{\partial \underline{v}}{\partial t} + \underline{v} \cdot \nabla \underline{v} \right) = -\nabla P + \nabla \cdot \underline{\underline{\tau}} \quad (9)$$

$$\nabla \cdot \underline{v} = 0 \quad (10)$$

Where  $\underline{v}$  is the velocity vector,  $P$  is the pressure,  $\rho$  is the density of the pastry and  $\underline{\underline{\tau}}$  is the extra stress tensor and the body forces not considered.

Many complex fluids of industrial interest are assumed to be viscoelastic because they exhibit contemporary viscous and elastic behaviour under strain. For these materials the extra stress may be written as an Upper convected Maxwell model (UCM)

$$\underline{\underline{\tau}} + \lambda \overset{\nabla}{\underline{\underline{\tau}}} = 2\eta \underline{\underline{D}} \quad (12a)$$

Where  $\underline{\underline{D}}$  is the rate of strain tensor and  $\overset{\nabla}{\underline{\underline{\tau}}}$  is the time upper convected derivative defined as:

$$\overset{\nabla}{\underline{\underline{\tau}}} = \frac{\partial \underline{\underline{\tau}}}{\partial t} + (\underline{v} \cdot \nabla) \underline{\underline{\tau}} - [(\nabla \underline{v}) \underline{\underline{\tau}} + \underline{\underline{\tau}} (\nabla \underline{v})^T] \quad (13)$$

Where  $\underline{\underline{D}}$  is the rate of strain tensor, which is defined as:

$$\underline{\underline{D}} = \frac{1}{2} [\nabla \underline{v} + (\nabla \underline{v})^T] \quad (14)$$

It should be noticed that the symmetry of the stress tensor implies that the off-diagonal extra stress components are equal, i.e.  $\tau_{ij} = \tau_{ji}$ . When considering a transient xy planar flow, then the only non zero components of the extra stress are  $\tau_{xy}$ ,  $\tau_{xx}$ ,  $\tau_{yy}$  and the velocity shows only the two components  $v_x$  and  $v_y$ .

The values of the two material parameters introduced by eq.12 have been experimental determined as above described. Because the doughs show a shear-thinning behaviour, it was necessary to modify the UCM model, by inserting the variation of viscosity with shear rate, as reported in the following? Indeed, the

flow curve data were fitted with a power law model in the range between  $0.1\text{-}100\text{ s}^{-1}$  (eq. 7), whilst  $\eta_\infty$  and  $\eta_0$  were assumed equal to the viscosity value showed at 100 and  $0.1\text{ s}^{-1}$  respectively. The equation (12a) begins:

$$\underline{\underline{\tau}} + \lambda \frac{\nabla \cdot \underline{\underline{\tau}}}{\dot{\gamma}} = 2K \dot{\gamma}^{n-1} \underline{\underline{D}} \quad (12b)$$

## 2.5. Method of solution

The constitutive equation (eq.12b) must be solved together with the conservation equations (eq.s 9-10) and with appropriate boundary conditions.

Figure 1 shows the solution domain. Because of symmetry, only one half of the domain is considered. The boundary conditions at any surfaces are reported in the following for material flowing from left to right, and indicating  $\underline{\underline{t}}$  and  $\underline{\underline{n}}$  as tangential and normal versor:

1. along the inlet boundary AF, the only stress acting across the boundary is due to pressure force:  $\underline{\underline{\tau}} \cdot \underline{\underline{n}} = -p_{atm} \underline{\underline{n}}$ ;

2. symmetry along the centerline AB:  $v_y=0$ ,  $\tau_{xy}=0$ ;

3. along the exit free surface DC, tangential and normal stresses vanish:  $(\underline{\underline{\tau}} \cdot \underline{\underline{n}}) \cdot \underline{\underline{t}} = 0$ ,  $(\underline{\underline{\tau}} \cdot \underline{\underline{n}}) \cdot \underline{\underline{n}} = 0$  and no flow is passing through the surface being  $(\underline{\underline{v}} \cdot \underline{\underline{n}}) = 0$ , where  $\underline{\underline{n}}$ ,  $\underline{\underline{t}}$  are the normal and tangential versors to the surface;

4. along the outlet boundary BC, the pressure is the atmospheric pressure:  $\underline{\underline{\tau}} \cdot \underline{\underline{n}} = -p_{atm} \underline{\underline{n}}$ ;

6. along the entry free surface FE, tangential and normal stresses vanish  $(\underline{\underline{\tau}} \cdot \underline{\underline{n}}) \cdot \underline{\underline{t}} = 0$ ,  $(\underline{\underline{\tau}} \cdot \underline{\underline{n}}) \cdot \underline{\underline{n}} = 0$  and no flow is passing through the surface being  $(\underline{\underline{v}} \cdot \underline{\underline{n}}) = 0$ ;

7. drag flow along the roll walls from point E to point D (tangential velocity  $v_t = \underline{\underline{t}} \cdot \underline{\underline{U}}$ , normal velocity  $v_n = \underline{\underline{n}} \cdot \underline{\underline{U}}$ ) (Mitsoulis E. 2008), assuming no slip.

The reference pressure is set to atmospheric pressure. For the simulations we have used the roll radius is 0.05 m and the velocity is 0.05m/s. The  $H_f$  entry height is set to 0.03m, while the half minimum gap is set to 0.003 m. The numerical solution is obtained with the finite element method (FEM), using the program COMSOL Multiphysics 3.5. The free surfaces at the inlet and outlet are found in a coupled way with the ALE method, implemented in COMSOL too. Because the boundaries of the computational domain are moving in the time, it was necessary a method to deform the mesh. In COMSOL Multiphysics, it is possible to control the movement of the interior nodes by letting the mesh movement determined by physical deformation variables.

The technique for mesh movement is an arbitrary Lagrangian-Eulerian (ALE) method. In the special case

of a Lagrangian method, the mesh movement follows the movement of the physical material. Such a method is often used in solid mechanics, where the displacements often are relatively small.

When the material motion is more complicated, like in a fluid flow model, the Lagrangian method is not appropriate. For such models, an Eulerian method, where the mesh is fixed, is often used—except that this method cannot account for moving boundaries.

The ALE method is intermediate between the Lagrangian and Eulerian methods, and it combines the best features of both—it allows moving boundaries without the need for the mesh movement to follow the materia. The triangular mesh used is shown in-Figure 2 and give 19056 unknown degrees of freedom. More elements have been concentrated near the attachment and detachment points from the roll, where most of the changes are expected. The entering sheet is usually set with a length major then  $H_f$  of at least 3 order of magnitude and the exiting sheet with a length of  $4H$  to guarantee a fully developed profile at entry and exit (Mitsoulis E. 2008, Mitsoulis E. 2009).

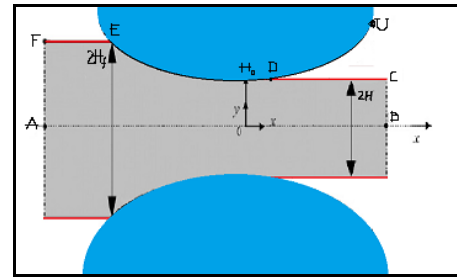


Figure 1: Flow domain and boundary conditions for the 2-D FEM analysis.

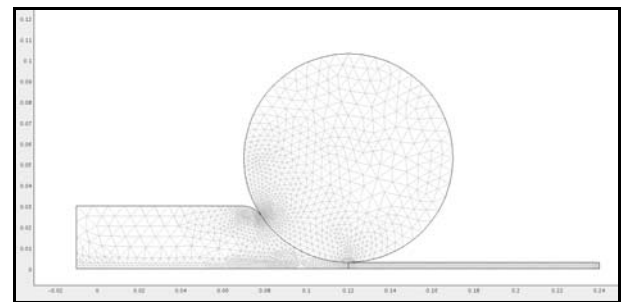


Figure 2: Finite element mesh used in the computation.

## 3. RESULTS AND DISCUSSION

### 3.1. Rheological analysis

According to the frequency sweep tests, all the samples showed a typical weak gel behaviour (data not shown); the weak gel model parameters of samples analysed after 72h from the preparation are listed in table 2, and compared to the margarine parameters.

Table 2: Weak Gel Model Parameters (72h from the

preparation)

Sample	$A$ [ $\text{Pa s}^{1/z}$ ]	$z$ [-]
Margarine Vallè	$43000 \pm 200$	$14 \pm 0.6$
Emu1	$86000 \pm 800$	$9.4 \pm 0.4$
Emu2	$14900 \pm 100$	$9.1 \pm 0.4$
Emu3	$30700 \pm 100$	$10.3 \pm 0.2$
Emu4	$6600 \pm 100$	$7.6 \pm 0.6$

Samples Emu 1, Emu 3 and Emu 4 were prepared varying the Myverol content in the oil phase. Figure 3 shows the effect of Myverol on the rheological characteristics of the emulsions after 24h, 72h and 8 days (labelled '8d' in figure 1) after the preparation. All the samples underwent a maturation of the network strength until 72h after preparation, (increase of  $A$  until a constant value was reached), while the network extension (parameter  $z$ ) was constant after 48h. Emu 1 (3% of Myverol) showed the highest  $A$  value. In fact, a further addition of Myverol (Emu 3) implicated a lower cocoa butter content, and a less consistent emulsion (lower  $A$  value), while a lower amount (Emu 4) gave the lowest values of both the weak gel model parameters. It is worthy noticing that Myverol gives structure to the system (higher crystallisation degree). Nevertheless, despite the little increase of  $z$  for Emu 3 with respect to the other samples, higher values of  $A$  are preferable for our aims, also considering the comparison with the commercial margarine (Lupi et al. 2011).

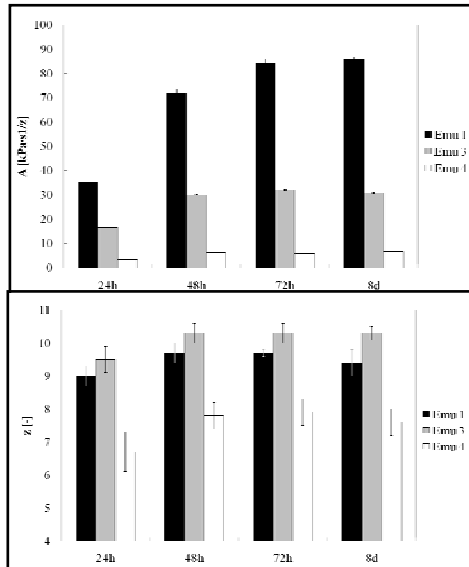


Figure 3: Weak Gel model parameters for W/O emulsions at different times from preparation.

Thus, for our scopes, the best emulsion is Emu 1: it is characterised by the higher  $A$  value and it results stable (no more variations of both the parameters) after 72h from its production. Considering these results, the formulation of Emu 1 was varied increasing the total amount of water from 12 to 20%. Emu 2 showed a low value of  $A$  with respect to Emu 1, but a similar parameter  $z$ . It seems that an increase in water content does not affect the crystallisation, but it decreases the

strength of the interactions in the network.

From a rheological point of view, the best pastry fat is Emu1. This sample was used to prepare an hand-made puff pastry, characterised and compared to a commercial one. Both the SAOTs (table 3) and the flow tests (figure 4) revealed differences between the two pastries, obviously due to the different ingredients and preparation processes. Anyway, the hand-made pastry was cooked (at  $200^\circ\text{C}$ ), and the consequent 'puffing' process resulted satisfying, giving the characteristic friable texture.

Table 3: Weak Gel Parameters For Pastries

Sample	$A$	$z$
commercial puff pastry	$88000 \pm 1300$	$4.6 \pm 0.2$
hand-made puff pastry	$127520 \pm 2480$	$3.1 \pm 0.1$

The setting parameters used to implement the modelling software for the sheeting process are the relaxing time  $\lambda$  and the power law parameters  $k$  and  $n$  (eq. 7):

$$\begin{aligned}\lambda &= 0.400 \pm 0.005 \text{ s} \\ k &= 5670 \pm 200 \text{ Pa} \cdot \text{s}^{1/z} \\ n &= 0.27 \pm 0.004\end{aligned}$$

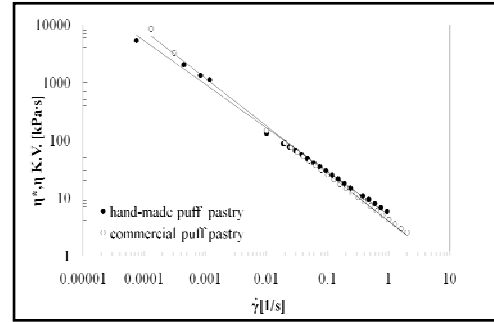


Figure 4: Flow Curves Of Commercial And Hand-Made Puff Pastry ( $\eta$  K.V. is referred to eq. 2)

### 3.2. Sheeting modelling

We present results from the two-dimensional isothermal analysis of rolling of puff pastry. The velocity profiles and the value of Pressure and Normal stresses along the coordinate  $x$  is reported in the Figure 5 and 6.

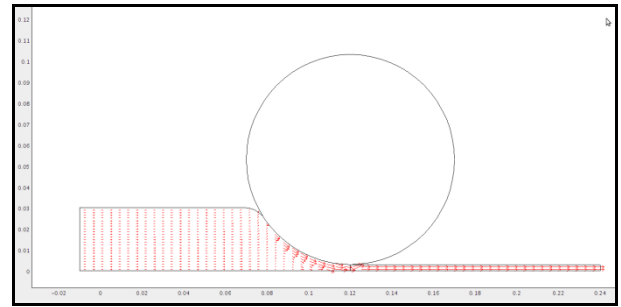


Figure 5: Arrow Velocity Profile in rolling of puff pastry.

The velocity profile along the  $x$ -coordinate shows variations of concavity due to the pressure profile. In

fact, the speed has a reversal point of the concavity at the attachment point. This change is due to the complete adherence of the sheet to the roll and then to the drag force that pushes forward the dough. As the material passes through the nip, it accelerates. The magnitude of the velocity at the nip is almost 3 times higher than the entry.

Then, after the maximum pressure value, the velocity increases due to the pressure gradient to begins constant, showing a typical plug velocity profile due to the existence of the free surfaces (Mitsoulis E. 2008, Mitsoulis E. 2009). Its speed at the exit is higher its speed at the entry, because of the mass conservation, namely, as the thickness decreases, the speed increases.

Looking at the pressure profile (Figure 6), it is possible to observe that the maximum pressure value is at 0.11 m, then before the minimum gap, and have a maximum value of 0.11 MPa. The pressure profile is slightly asymmetric, in agreement with the literature, due to the non-Newtonian material (Mitsoulis E. 2008, Mitsoulis E. 2009) and as expected, the biggest changes occur in the nip region.

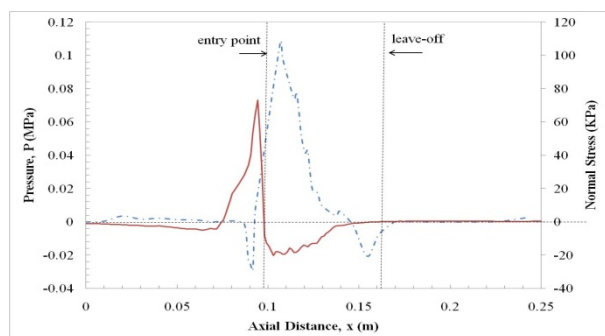


Figure 6: Pressure Profile in rolling of puff pastry.

Finally, it is clear that the points of attachment and detachment give the difficulty in the calculation with FEM.

The swelling of the puffy pastry is found at the exit of the nip of the roll, due to the material memory that completely relaxes the normal stresses because is not any longer confined. The point of detachment was obtained at a value of about 0,1455 m. The value of swelling, calculated as  $H/H_0$  is equal to 1.8.

The normal stress  $\tau_{xx}$  varies along the calendaring exhibiting its maximum value at the attachment point and a minimum value at the maximum of the pressure (Figure 6).

The model validation was not performed, but it is currently under experimental investigation.

#### 4. CONCLUSIONS

Rheological modelling allowed, through the optimization of the formulation of water-in-olive-oil emulsions, structured with an organogelator agent, to characterize and obtain a puff pastry with rheological characteristics similar to the commercial one. The rheological parameter were used to model a rolling process. The rolling model, using an UCM modified

constitutive equation, was implemented in Comsol Multiphysics 3.5. The velocity, pressure and normal stress profile were found and the swell and detachment point were predicted by the model.

The present results are a first step towards a better understanding of rolling of viscoelastic materials. This analysis coupled with rheological data for a UCM modified model, and operating data for the process, can be useful to predict the final sheet height and swell.

#### REFERENCES

- Cavillot, V., Pierart, C., Kervyn De Meerendré, M., Vincent, M., Paquot, M., Wouters, J., Deroanne, C., Danthine, S., 2009. Physicochemical Properties Of European Bakery Margarines With And Without Trans Fatty Acids. *Journal of Food Lipids*, 16, 273-286.
- de Cindio, B., Lupi, F. R., (2011), Part 2, Chap. 15 – Saturated fat reduction in pastry in Talbot G., *Reducing saturated fats in foods*, Woodhead Publishing Limited (Great Abington, Cambridge, UK), in press.
- Simovic, D. S., Pajin, B., Seres, Z., Filipovic, N., 2009. Effect of low-trans margarine on physicochemical and sensory properties of puff pastry. *International Journal of Food Science and Technology*, 44, 1235–1244.
- Marangoni, A. G., 2009. Novel strategies for nanostructuring liquid oils into functional fats. *Proceedings of the 5th International Symposium on Food Rheology and Structure*, pp. 15-18. June 2009, (Zurich, CH)
- Burkhardt, M., Kinzel, S., & Gradzielski, M., 2009. Macroscopic properties and microstructure of HSA based organogels: Sensitivity to polar additives. *Journal of Colloid and Interface Science*, 331, 514–521
- Levine, L., Corvalan, C.M., Campanella, O.H., Okos M.R., 2002. A model describing the two-dimensional calendaring of finite width sheets. *Chemical Engineering Science* 57, 643-650.
- Lupi, F.R., 2010. Rheology of highly-concentrated-in-oil emulsions. Thesis (PhD). University of Calabria.
- Harte, J.B., 2003. Pastry Products/Types and Production. In: Caballero, B., Trugo, L., Finglas, P., *Encyclopaedia Of Food Science And Nutrition*. Amsterdam: Academic Press, 4407-4412.
- Gabriele, D., de Cindio, B., D'Antona, P., 2001. A weak gel model for foods. *Rheologica Acta*. 40, 120-127.
- Steffe, J.F., 1996. *Rheological methods in food process engineering*, Freeman Press. Michigan (Michigan, USA).
- Peressini, D., Sensidoni, A., Pollini, C. M., Gabriele, D., Migliori, M., de Cindio, B., (2002). *Filled-snacks production by co-extrusion-cooking. Part 3. A rheological-based method to compare*

- filler processing properties*. Journal of Food Engineering. 54, 227–240.
- Cox, W. P. and Merz, E. H., 1958. *Correlation of dynamic and steady flow viscosities*. Journal of Polymer Science 28, 619-622.
- Mitsoulis, E., 2008. *Numerical simulation of calendaring viscoplastic fluids*. Journal of Non-Newtonian Fluid Mechanics. 154, 77-88.
- Mitsoulis, E., Hatzikiriakos, S. G., 2009. *Rolling of bread dough: Experiments and Simulations*. Food and Bioproducts Processing. 87, 124-138.
- Peck, M. C., Rough, S.L., Barnes, J., Wilson, D. I., 2005. *Roller extrusion of biscuit dough*. Journal of food engineering. 74, 431-450.
- Tanner, R. I. 2000. *Engineering rheology*, Oxford University Press Inc. (New York, USA).

## ACKNOWLEDGEMENT

The University of Calabria and the Food Science and Engineering Interdepartmental Center of University of Calabria and L.I.P.A.C., Calabrian Laboratory of Food Process Engineering (Regione Calabria APQ- Ricerca Scientifica e Innovazione Tecnologica I atto integrativo, Azione 2 laboratori pubblici di ricerca mission oriented interfiliere) are gratefully acknowledged.



# NEW EVOLUTIONARY ALGORITHM BASED ON PARTICLE SWARM OPTIMIZATION AND ADAPTIVE PLAN SYSTEM WITH GENETIC ALGORITHM

Pham Ngoc Hieu<sup>(a)</sup>, Hiroshi Hasegawa<sup>(b)</sup>

<sup>(a)</sup>Functional Control Systems - Graduate School of Engineering and Science  
Shibaura Institute of Technology, Japan

<sup>(b)</sup>Department of Systems Engineering and Science  
Graduate School of Engineering and Science  
Shibaura Institute of Technology, Japan

<sup>(a)</sup>[pnh112@gmail.com](mailto:pnh112@gmail.com), <sup>(b)</sup>[h-hase@shibaura-it.ac.jp](mailto:h-hase@shibaura-it.ac.jp)

## ABSTRACT

To reduce a large amount of calculation cost and to improve the convergence to the optimal solution for multi-peak optimization problems with multi-dimensions, we propose a new method of Adaptive plan system with Genetic Algorithm (APGA). This is an approach for improving Particle Swarm Optimization (PSO) using APGA. The new strategy based on PSO operator and APGA (PSO-APGA) to improve the convergence towards the optimal solution. The PSO-APGA is applied to some benchmark functions with 20 dimensions to evaluate its performance.

Keywords: Particle Swarm Optimization, Genetic Algorithm, Adaptive System, Multi-peak problems.

## 1. INTRODUCTION

The product design is becoming more and more complex for various requirements from customers and claims. As a consequence, its design problem seems to be multi-peak problem with multi-dimensions. The Genetic Algorithm (GA) (Goldberg 1989) is the most emergent computing method has been applied to various multi-peak optimization problems. The validity of this method has been reported by many researchers (Digalakis and Margaritis 2000, Sakuma and Kobayashi 2001, Li and Kirley 2002). However, it requires a huge computational cost to obtain stability in the convergence to an optimal solution. To reduce the cost and to improve stability, a strategy that combines global and local search methods becomes necessary. As for this strategy, current research has proposed various methods (Mahfoud and Goldberg 1992; Goldberg and Voessner 1999; Hiroyasu, Miki and Ogura 2000; Miki, Hiroyasu and Fushimi 2003). For instance, Memetic Algorithms (MAs) are a class of stochastic global search heuristics in which Evolutionary Algorithms-based approaches (EAs) are combined with local search techniques to improve the quality of the solutions created by evolution. MAs have proven very successful across the search ability for multi-peak functions with multiple dimensions (Smith, Hart and Krasnogor 2005).

These methodologies need to choose suitably a best local search method from various local search methods for combining with a global search method within the optimization process. Furthermore, since genetic operators are employed for a global search method within these algorithms, design variable vectors (DVs) which are renewed via a local search are encoded into its genes many times at its GA process. These certainly have the potential to break its improved chromosomes via gene manipulation by GA operators, even if these approaches choose a proper survival strategy.

To solve these problems and maintain the stability of the convergence to an optimal solution for multi-peak optimization problems with multiple dimensions, Hasegawa et al. proposed a new evolutionary algorithm (EA) called an Adaptive Plan system with Genetic Algorithm (APGA) (Hasegawa 2007).

Particle Swarm Optimization (PSO), first introduced by Kennedy and Eberhart (2001) is one of the modern meta-heuristics algorithms. It has been developed through simulation of a simplified social system, and has been found to be robust in solving optimization problems. Nevertheless, the performance of the PSO greatly depends on its parameters and it often suffers from the problem of being trapped in the local optimum. To resolve this problem, various improvement algorithms are proposed. It is proven to be a successful in solving a variety of optimal problems.

In this paper, we proposed a new strategy for optimization using PSO and APGA (PSO-APGA) to converge to the optimal solution.

This paper is organized in the following manner. The concept of PSO is described in Section 2, concept of APGA is in Section 3. Section 4 explains the algorithm of proposed strategy (PSO-APGA), and Section 5 discussed about the convergence to the optimal solution of multi-peak benchmark functions. Finally, Section 6 includes some brief conclusions.

## 2. PARTICLE SWARM OPTIMIZATION

PSO is a robust stochastic optimization algorithm which is defined by the behavior of a swarm of particles in a

multidimensional search space looking for the best solution (Kennedy and Eberhart 2001, Clerc 2005).

We are concerned here with gbest-model which is known to be conventional PSO. In this model, each particle which make up a swarm has information of its position  $x_i$  and velocity  $v_i$  ( $i$  is the index of the particle) at the present in the search space. Each particle aims at the global optimal solution by updating next velocity making use of the position at the present, based on its best solution has been achieved so far  $p_{id}$  and the best solution of all particles  $p_{gd}$  ( $d$  is the dimension of the solution vector), as following equation:

$$v_{id}^{t+1} = wv_{id}^t + c_1 r_1 (p_{id}^t - x_{id}^t) + c_2 r_2 (p_{gd}^t - x_{id}^t) \quad (1)$$

where  $w$  is inertia weight;  $c_1$  and  $c_2$  are cognitive acceleration and social acceleration, respectively;  $r_1$  and  $r_2$  are random numbers uniformly distributed in the range  $[0.0, 1.0]$ .

In our strategy, the discrete version of PSO designed for binary optimization has been adapted, which developed by Kennedy and Eberhart (1997). Discrete PSO is composed of the binary variable, so the velocity must be transformed into the change of probability. In this version of PSO, the velocity update is same as continuous PSO, but for adjusting the new position the probability function of particle velocity was used as follows:

$$S(v_{id}^t) = \frac{1}{1 + e^{-v_{id}^t}} \quad (2)$$

$$x_{id}^{t+1} = \begin{cases} 1, & \rho \leq S(v_{id}^t) \\ 0 \end{cases} \quad (3)$$

where  $S(v_{id}^t)$  and  $\rho$  are sigmoid limiting transformation and random number selected from a uniform distribution in  $[0.0, 1.0]$  respectively.

### 3. ADAPTIVE PLAN SYSTEM WITH GENETIC ALGORITHM

#### 3.1. Formulation of the optimization problem

The optimization problem is formulated in this section. Design variable, objective function and constrain condition are defined as follows:

$$X = [x_1, \dots, x_n] \quad (4)$$

$$-f(X) \rightarrow \text{Max} \quad (5)$$

$$X^{LB} \leq X \leq X^{UB} \quad (6)$$

where  $X^{LB} = [x_1^{LB}, \dots, x_n^{LB}]$ ,  $X^{UB} = [x_1^{UB}, \dots, x_n^{UB}]$  and  $n$  denote the lower boundary condition vectors, the upper boundary condition vectors and the number of design variable vectors (DVs) respectively. A number

of DV's significant figure is defined, and DV is rounded off its places within optimization process.

#### 3.2. APGA

The APGA concept was introduced as a new EA strategy for multi-peak optimization problems. Its concept differs in handling DVs from general EAs based on GAs. EAs generally encode DVs into the genes of a chromosome, and handle them through GA operators. However, APGA completely separates DVs of global search and local search methods. It encodes control variable vectors (CVs) of AP into its genes on Adaptive system (AS). Moreover, this separation strategy for DVs and chromosomes can solve Memetic Algorithm (MA) problem of breaking chromosomes (Smith, Hart and Krasnogr 2005).

The conceptual process of APGA is shown in Figure 1. The control variable vectors (CVs) steer the behavior of adaptive plan (AP) for a global search, and are renewed via genetic operations by estimating fitness value. For a local search, AP generates next values of DVs by using CVs, response value vectors (RVs) and current values of DVs according to the formula:

$$X_{t+1} = X_t + NR_t \cdot AP(C_t, R_t) \quad (7)$$

where  $NR$ ,  $AP()$ ,  $X$ ,  $C$ ,  $R$ ,  $t$  denote neighborhood ratio, a function of AP, DVs, CVs, RVs and generation, respectively. In addition, for a verification of APGA search process, refer to ref. (Hasegawa 2007).

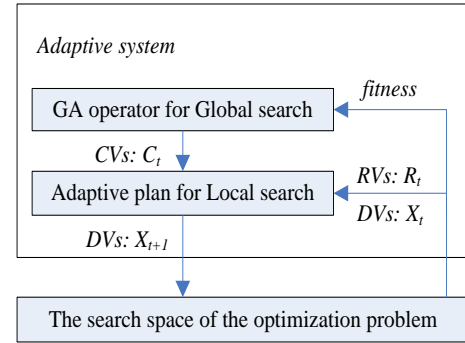


Figure 1: Conceptual Process Of APGA

#### 3.3. Adaptive Plan (AP)

It is necessary that the AP realizes a local search process by applying various heuristics rules. In this paper, the plan introduces a DV generation formula using a sensitivity analysis that is effective in the convex function problem as a heuristic rule, because a multi-peak problem is combined of convex functions. This plan uses the following equation:

$$AP(C_t, R_t) = -Scale \cdot SP \cdot \text{sign}(\nabla R_t) \quad (8)$$

$$SP = 2C_t - 1 \quad (9)$$

where  $Scale$ ,  $\nabla R$  denote the scale factor and sensitivity of RVs, respectively.

A step size  $SP$  is defined by CVs for controlling a global behavior to *prevent* it falling into the local optimum.  $C = [c_{i,j}, \dots, c_{i,p}]$ ;  $0.0 \leq c_{i,j} \leq 1.0$  is used by (9) so that it can change the direction to improve or worsen the objective function, and  $C$  is encoded into a chromosome by 10 bit strings with two values (0 and 1). In addition,  $i, j$  and  $p$  are the individual number, design variable number and its size, respectively.

### 3.4. GA Operators

Selection is performed using a tournament strategy to maintain the diverseness of individuals with a goal of keeping off an early convergence. A tournament size of 2 is used.

Elite strategy, where the best individual survives in the next generation, is adopted during each generation process. It is necessary to assume that the best individual, i.e., as for the elite individual, generates two behaviors of AP by updating DVs with AP, not GA operators. Therefore, its strategy replicates the best individual to two elite individuals, and keeps them to next generation. As shown in Figure 2, DVs of one of them ( $\blacktriangle$  symbol) is renewed by AP, and its CVs which are coded into chromosome aren't changed by GA operators. Another one ( $\bullet$  symbol) is that both DVs and CVs are not renewed, and are kept to next generation as an elite individual at the same search point.

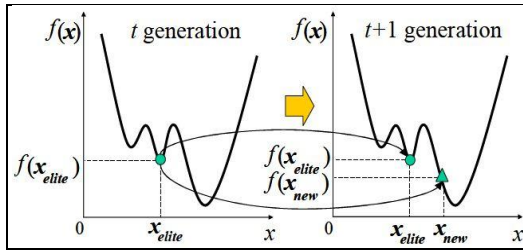


Figure 2: Elite Strategy

In order to pick up the best values of each CV, a single point crossover is used for the string of each CV. This can be considered to be a uniform crossover for the string of the chromosome. Mutation are performed for each string at mutation ratio on each generation, and set to maintain the strings diverse.

At following conditions, the genetic information on chromosome of individual is recombined by uniform random function.

- One fitness value occupies 80% of the fitness of all individuals.
- One chromosome occupies 80% of the population.

If this manipulation is applied to general GAs, an improved chromosome into which DVs have been encoded is broken down. However, in the APGA, the genetic information is only CVs used to make a decision for the AP behavior. Therefore, to prevent from falling into a local optimum, and to get out from the condition of being converged with a local optimum,

a new AP behavior is provided by recombining the genes of the CVs into a chromosome. And the optimal search process starts to re-explore by a new one. This strategy is believed to make behavior like the re-annealing of the Simulated Annealing (SA).

### 4. NEW EVOLUTIONARY ALGORITHM

Many of optimization techniques called meta-heuristics including PSO are designed based on the concept. In case of PSO, when a particle discovers a good solution, other particles also gather around the solution. Therefore, they cannot escape from a local optimal solution. Consequently, PSO cannot achieve the global search. However, APGA that combines the global search ability of a GA and an Adaptive Plan with excellent local search ability is superior to other EAs (MAs) (Hasegawa 2007, Tooyama and Hasegawa 2009, Pham and Hasegawa 2010). With a view to global search, we propose the new evolutionary algorithm based on PSO and APGA named PSO-APGA, as shown in Figure 4.

To improve the multi-point search capability of APGA, applying neighborhood range control is used. The distance for a search point can be changed by controlling  $NR$  for determining the neighborhood range.  $NR$  is adjusted by following sigmoid function:

$$NR = \frac{1}{1 + \exp\left(\beta \cdot \frac{inv - (individual/2)}{individual}\right)} \quad (10)$$

$$0.0 \leq NR \leq 1.0 \quad (11)$$

where  $\beta$ ,  $inv$  denote the gain of the sigmoid function and the individual number, respectively.

PSO-APGA aims at getting direction from particle swarm to adjust into adaptive system. This strategy introduces a handling sign of the gain  $\beta$  for assignation of neighborhood range control. The velocity update is same as discrete PSO operator, in which a moving in a state space restricted to local search and global search on each dimension, in terms of the changes in probabilities that a bit will be in one state or the other. Such a situation, in which the individual searches its own neighborhood area without performing a global search, can generally occur at any time in search process. Therefore, it cannot escape from the local optimum solution. To solve this problem, this method employs sign of the gain  $\beta$  using the probability function of particle velocity as follows:

$$\begin{cases} \beta > 0, & \rho \leq S(v_{id}^t) \\ \beta < 0 \end{cases} \quad (12)$$

As the assignation step shown in Figure 2, the individuals are allocated small  $NR$  values to perform a local search efficiently, and the individuals are allocated large  $NR$  values to search the global area in the design space of DVs by  $\beta < 0$ . On the other hand, just before

converging to the global optimum solution, individuals can gather in the neighborhood area of the elite individual by  $\beta > 0$ .

In this approach, PSO and APGA run individually. The iteration is run by PSO operator and the velocity update is given as initial parameter for APGA process.

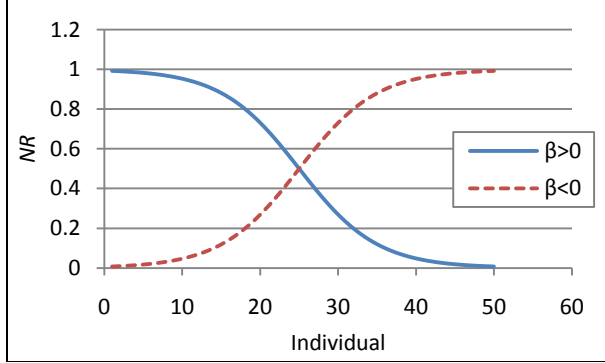


Figure 3: Neighborhood Range Control

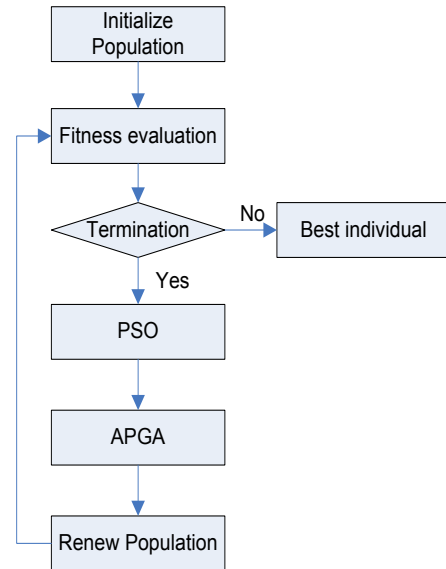


Figure 4: Flow-chart Of PSO-APGA

## 5. NUMERICAL EXPERIMENTS

To show the effects of PSO-APGA whether particles can escape from a local optimal solution and find the global optimal solution, we compare with other methodologies for the robustness of the optimization process. These experiments are performed 20 trials for every function. The initial seed number is randomly varied during every trial. In each experiment, the inertia weight  $w$  is set 0.9, and the acceleration coefficients  $c_1$  and  $c_2$  are set by fixed value of 2.0. The GA parameters used in solving benchmark functions are set as follows: selection ratio, crossover ratio and mutation ratio are 1.0, 0.8 and 0.01 respectively. The population size is 50 individuals and the terminal generation is 5000<sup>th</sup> generation. The sensitivity plan parameters in (8) for normalizing functions are listed in Table 3.

### 5.1. Benchmark Functions

For PSO-APGA, we estimate the stability of the convergence to the optimal solution by using three benchmark functions with 20 dimensions: Rastrigin (RA), Griewank (GR) and Rosenbrock (RO). These functions are given as follows:

$$RA = 10n + \sum_{i=1}^n \{x_i^2 - 10\cos(2\pi x_i)\} \quad (13)$$

$$GR = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad (14)$$

$$RO = \sum_{i=1}^n 100(x_{i+1} + 1 - (x_i + 1)^2)^2 + x_i^2 \quad (15)$$

Table 1 shows their characteristics, and the terms epistasis, multi-peak, steep denote the dependence relation of the DVs, presence of multi-peak and level of steepness, respectively. All functions are minimized to zero, when optimal DVs  $X = 0$  are obtained. Moreover, it is difficult to search for their optimal solutions by applying one optimization strategy only, because each function has a different complicated characteristic. In Table 2, their design range, the digits of DVs are summarized. If the search point attains an optimal solution or a current generation process reaches the termination generation, the search process is terminated.

Table 1: Characteristics Of The Benchmark Functions

Function	Epitasis	Multi-peak	Steep
RA	No	Yes	Average
GR	Yes	Yes	Small
RO	Yes	No	Big

Table 2: Design Range. Digits Of DVs

Function	Design range	Digits
RA	$-5.12 \leq X \leq 5.12$	2
GR	$-51.2 \leq X \leq 51.2$	1
RO	$-2.048 \leq X \leq 2.048$	3

Table 3: Scale Factor For Normalizing The Benchmark Functions

Function	Scale Factor
RA	10.0
GR	100.0
RO	4.0

### 5.2. Experiment Results

The experiment results are shown in Table 4. The success ratio of all benchmark functions is 100% with small computation cost. The solutions of all benchmark functions reach their global optimum solutions.

Next, Figure 5, Figure 6 and Figure 7 show diagrams for the average fitness of individual until PSO-APGA reaches the global optimum solutions, in the numerical experiment again to confirm above mentioned result.

Table 4: Number Of Generations By PSO-APGA

Trial	Function		
	RA	GR	RO
1	198	360	1164
2	200	332	1118
3	225	394	1020
4	208	340	1018
5	228	412	913
6	178	323	1213
7	239	413	900
8	221	412	791
9	170	349	1400
10	238	399	1260
11	248	347	1253
12	212	331	843
13	229	399	996
14	207	373	1113
15	210	369	1251
16	155	371	1150
17	236	330	1269
18	204	317	1039
19	216	361	1155
20	198	359	1306
Average	211	364	1108

### 5.3. Comparison

PSO-APGA was compared with basic PSO and H-APGA (Pham and Hasegawa 2010). The results of these methods are shown in Table 5. In the table, when success rate of optimal solution is not 100%, “-” is described.

In particular, it was confirmed that the calculation cost with PSO-APGA could be reduced for benchmark functions. And it showed that the convergence to the optimal solution could be improved more significantly.

In summary, from the comparison among methods shown in Figure 8, we can confirm that the search ability of PSO-APGA with multi-dimensions optimization problems is very effective, compared with that of basic PSO. However, it did not gain a high probability by H-APGA.

Additionally, we calculated the standard deviation, used in combination with the average result to describe the normal distribution of PSO-APGA and H-APGA with RO function. As a result, PSO-APGA achieved at the optimal solution, however it still has the larger variation than H-APGA.

Overall, the PSO-APGA was capable of attaining robustness, high quality, low calculation cost and efficient performance on many benchmark problems.

Table 5: Average Results With 20 Dimensions

Function	Basic PSO	H-APGA	PSO-APGA
RA	-	196	211
GR	2878	298	364
RO	2220	1088	1108

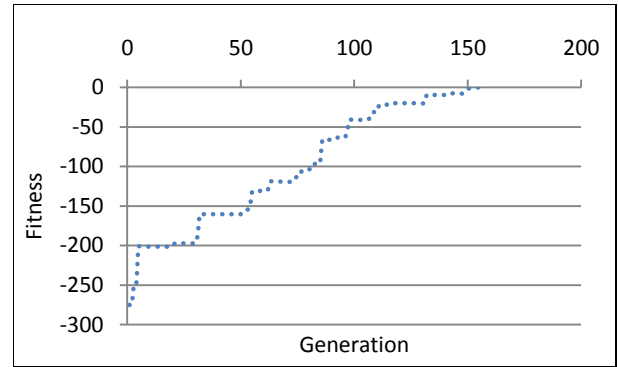


Figure 5: Elite Individual Fitness With RA

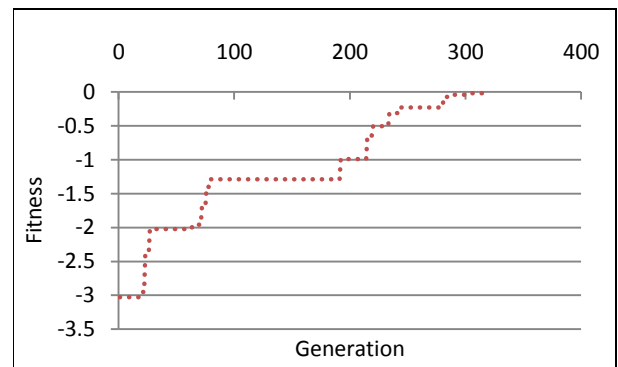


Figure 6: Elite Individual Fitness With GR

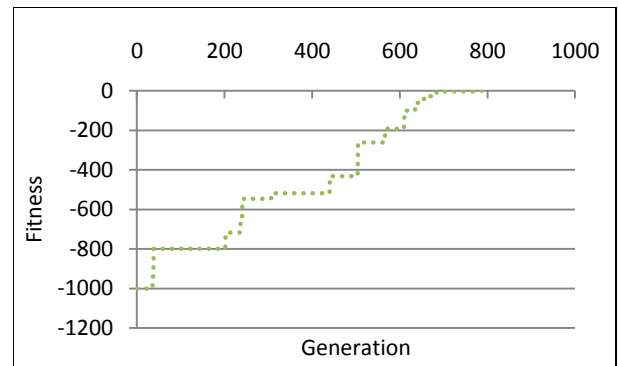


Figure 7: Elite Individual Fitness With RO

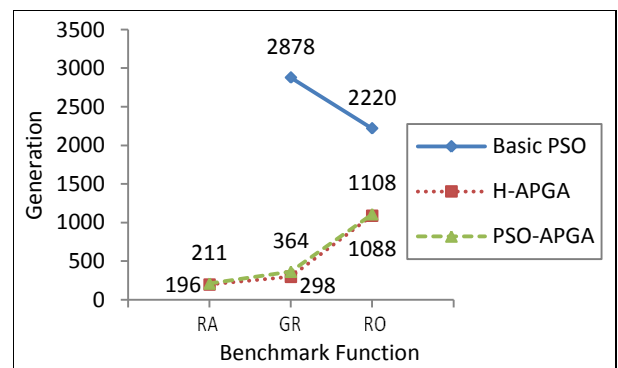


Figure 8: Comparison Among Methods

## 6. CONCLUSION

In this paper, to overcome the computational complexity, two efficient optimization evolutionary algorithms have been used. A new approach strategy based on PSO and APGA (PSO-APGA) has been proposed. Then, we verify the effectiveness of PSO-APGA by the numerical experiments performed three benchmark functions.

Moreover it was compared with basic PSO and H-APGA. As a result, we can confirm that the PSO-APGA reduces the calculation cost and improves the convergence to the optimal solution.

About the optimal solution such as minimum time and maximum reliability, it is a future work.

Finally, this study plans to do a comparison with the sensitivity plan of the AP by applying other optimization methods and optimizing the benchmark functions.

## REFERENCES

- Goldberg, D. E., 1989. *Genetic Algorithms in Search Optimization & Machine Learning*. Addison-Wesley.
- Digalakis, J. D., Margaritis, K. G., 2000. An experiment study of benchmarking functions for Genetic Algorithms. *Proceedings of IEEE Conference on Transactions*, 5:3810-3815.
- Sakuma, J., Kobayashi, S., 2001. Extrapolation-Directed Crossover for Real-coded GA: Overcoming Deceptive Phenomena by Extrapolative Search. *Proceedings of Congress on Evolutionary Computation (CEC 2001)*, 685-662.
- Li, X., Kirley, M., 2002. The Effects of Varying Population Density in a Fine-grained Parallel Genetic Algorithm. *Proceedings of Congress on Evolutionary Computation (CEC 2002)*, 2:1709-1714.
- Mahfoud, S. W., Goldberg, D. E., 1992. A genetic algorithm for parallel simulated annealing. *Parallel Problem Solving from Nature*, 2:301-310.
- Goldberg, D. E., Voessner, S., 1999. Optimizing global - local search hybrids. *IlliGAL report no. 99001*.
- Hiroyasu, T., Miki, M., Ogura, M., 2000. Parallel Simulated Annealing using Genetic Crossover. *Proceedings of the ISCA 13th International Conference on PDCS-2000*.
- Miki, M., Hiroyasu, T., Fushimi, T., 2003. Parallel Simulated Annealing with Adaptive Neighborhood Determined by GA. *IEEE International Conference on System, Man and Cybernetics*.
- Smith, J. E., Hart, W. E., Krasnogor, N., 2005. *Recent Advances in Memetic Algorithms*. Springer.
- Hasegawa, H., 2007. Adaptive Plan System with Genetic Algorithm based on Synthesis of Local and Global Search Method for Multi-peak Optimization Problems. *Proceedings of the 6th EUROSIM Congress on Modelling and Simulation*.
- Tooyama, S., Hasegawa, H., 2009. Adaptive Plan System with Genetic Algorithm using the Variable Neighborhood Range Control. *IEEE Congress on Evolutionary Computation (CEC 2009)*, 846-853.
- Pham, N. H., Hasegawa, H., 2010. Hybrid neighborhood control method of adaptive plan system with genetic algorithm. *Proceedings of the 7th EUROSIM Congress on Modelling and Simulation*.
- Kennedy, J., Eberhart, R., 1997. A discrete binary version of the particle swarm algorithm. *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4104-4108.
- Kennedy, J., Eberhart, R., 2001. *Swarm Intelligence*. Morgan Kaufmann Publishers.
- Clerc, M., 2005. *Particle swarm optimization*. ISTE.

## AUTHORS BIOGRAPHY

**Pham Ngoc Hieu** received the B.E. (2007) in Mechatronics from Hanoi University of Science and Technology (HUST), Vietnam. He got the double degree of Master from HUST and Shibaura Institute of Technology (SIT), Japan (2010). Currently, he is a doctor student at Functional Control Systems - Graduate School, SIT. His research interests include optimization design, robotics, and neuroscience.

**Hiroshi Hasegawa** received his B.E. (1992) and M.E. (1994) from Shibaura Institute of technology, Japan. He received PhD (1998) in Mechanical Engineering from Tokyo Institute of technology, Japan. He currently is Professor working at Shibaura Institute of Technology. He also is member of Japan Society of Mechanical Engineers, American Society of Mechanical Engineers and Japan Society for Simulation Technology. His research interests include computer-aided exploration, especially multi-peak optimization, robust design and multi-disciplinary optimization.



# THE ART OF DATA-DRIVEN MODELLING IN LOGISTICS SIMULATION

Rainer Frick

V-Research GmbH, Dornbirn, Austria

[rainer.frick@v-research.at](mailto:rainer.frick@v-research.at)

## ABSTRACT

The field of discrete event simulation used for logistics simulation is covered by many software tools and software libraries. Most of them offer libraries and functionality to model a simulation case and run experiments. Already small changes of the logistics case can result in considerable efforts of changing the simulation model. Data-driven simulation modelling provides the possibility to reduce adaptation work and makes simulation model more long-lasting. This work comes up with a reflection of state-of-the-art data-driven modelling techniques and reports on implemented sample cases in the world of logistics.

Keywords: Simulation, Modelling, Data-driven

## 1. INTRODUCTION

Discrete event simulation is part of operations research and is concerned with modelling dynamic systems. The system is described by time-dependent state variables and these variables are changed by occurring events. Simulation helps modelling more complex process correlations and can be used for comparative analysis (März 2008, Wenzel 2009).

In the logistics field simulation is used to model process flows to analyze cost structures, to discover bottlenecks or to visualize the business activities behind the scene, i.e. of transportation networks, warehouse activities or production flows.

To create a valid and significant simulation model in these areas can be a cost-intensive matter. Simulation know-how and knowledge in the specific area must be combined to shape an appropriate simulation entity and process model. And what happens if the underlying business case changes or enhances? This requires adaptations in the simulation model as well and therefore a simulation expert is needed again which increases costs.

Data-driven modelling of simulation studies provides more flexibility in terms of a changing environment.

Partly data-driven modelling can be found in many simulation studies. Normally order, product or demand

data is taken out of a data archive for their usage in a simulation. Layout and resource data in a simulation model are represented as simulation entities or objects which are most often created by drag and drop in a graphical editor.

## 2. THE APPROACH

Presently modelling of a simulation study is done by defining simulation entities in a simulation software tool with the help of a graphical editor (Figure 1) or implementing them directly in code with the specific programming language.

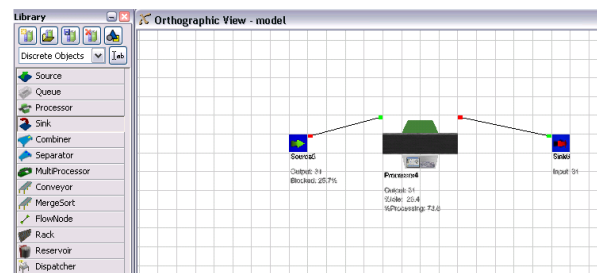


Figure 1: Model-driven modelling

As a sample can serve a warehouse where racks, forklifts and workers are placed in the simulation model and interconnected with each other (see Figure 2).

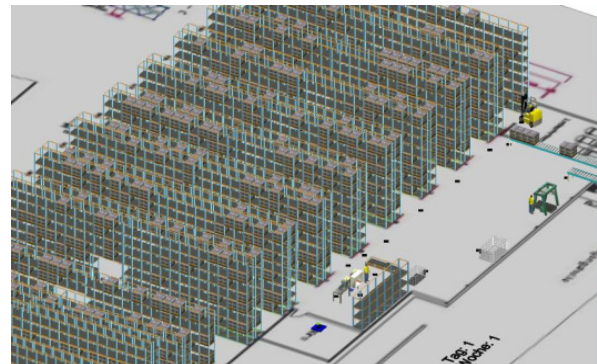


Figure 2: A modelled sample warehouse

Later changes of the model in Figure 2, i.e. placing more racks, mean changing the simulation model itself in the simulation software tool. This requires a simulation expert again, because the additional racks

have to be included into the warehouse layout and workflow.

A dynamic way of building the simulation model automatically can obviate this change loop. New simulation entities are added to a database and not directly to the simulation model. Therefore the basic concept of data-driven modelling is the separation of the simulation model from the used simulation software. All the data needed for a simulation study are lodged in a data archive, most often a database, and not in the simulation software. At runtime the data is read from the archive and the simulation model is created dynamically. In other words the simulation entities necessary to run a simulation with a specific simulation software tool are created with the help of a generation software tool out from the data archive (Wang 2008, Hassan 2009, Jensen 2007).

Figure 3 shows the software components needed most often for data-driven modelling of a simulation study. All the data necessary for the simulation model is stored in a data archive. For storing, changing and viewing this software an assistance software tool can be helpful. The generation software creates the simulation model able to run out from the archive. This model is carried out by a simulation software tool (Yang 2008).

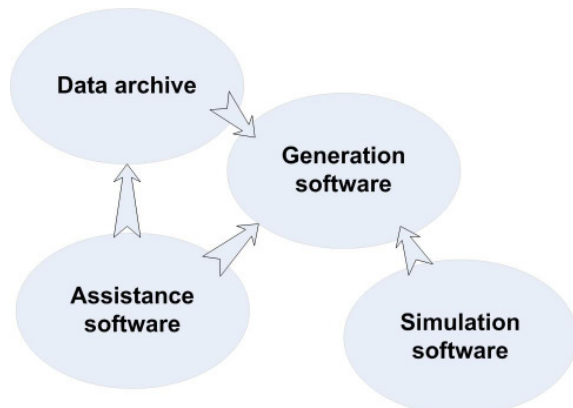


Figure 3: Components of a data-driven modelling system

As a data archive often a relational database or Xml files are used. The generation software is an individually implemented tool focused on the specific problem domain with knowledge of the used simulation software. It provides the simulation entities for the simulation software. The simulation software can be any commercial or non-commercial simulation tool with an application programming interface (API) (Bergmann 2010).

Figure 4 demonstrates the basic application cycle of a data-driven simulation. Common for all simulation studies is the phase of data collection and preparation. A data model is defined, a data archive created and the data stored in tables, files etc. Out of this data the

simulation model is generated automatically and it is executed by a simulation software tool. During simulation operation steps, tasks, events, times are logged into the data archive. With this logging information the results can be calculated and the result analysis is prepared. This result analysis leads to scenarios deduced from the original model. Alternative experiments can be carried out.

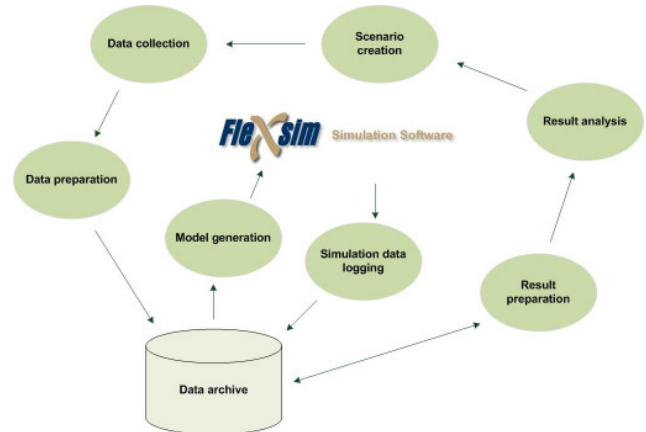


Figure 4: Simulation application cycle

### 3. PROCESS LEVEL

Data-driven modelling provides the simulation entities automatically to be used in the simulation model. But how about the process flows, the mode information and material find their way?

Three modes are possible.

First the process flow is not part of the data archive and the generation software. It is implemented in the simulation software and only the simulation entities are created automatically by the generation software.

Second the process flow is defined implicitly by creating the simulation objects. In other words the behaviour of these objects is implemented in their abstract definition, normally in a particular code class. This is common practice in agent-based simulations.

Third the process flow is generated automatically too and therefore must be expanded into the data archive. To model the process flow the assistance software has to be extended to define the interaction between the simulation entities; a process model editor is suggested to handle this. Therefore process units must be defined, each of them combined with some simulation entities. The process model editor is capable to combine these process units to a valid and complete process flow (see Figure 5). To establish alternative process arms conditions which interpret the state of simulation entities are needed.

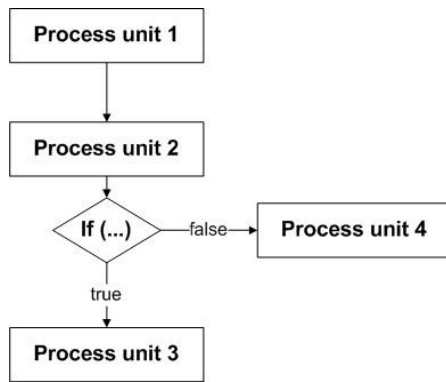


Figure 5: Sample process flow

#### 4. PROS AND CONS

Data-driven modelling offers more flexibility to the process of building simulation models. Changes in the simulation case can be handled by the generation software to build a model which is able to run. In some circumstances the modelling data can be taken directly from the enterprise planning software which means an integration of the simulation technique into the operational business. Some other less important pros can be identified like the relieving of the computer system by loading only necessary data at runtime, the usage of software design patterns by separating programming code into different aspects, consistency check of the data archive and a step towards to a modelling standard.

But ever more flexibility is caught in the generation software its complexity raises to a higher level which can make their implementation more difficult. Nevertheless it is worth the effort because the costs have to be seen in relation to the practicability of the resulting simulation models.

#### 5. LOGISTICS SAMPLES

In the scope of the authors daily work view simulation projects have been settled down with different data-driven principles behind them.

##### 5.1. Transportation network

For a less complex simulation problem it can be helpful to do the generation of the model in the simulation software tool itself. Therefore only a database access has to be existent in the tool; no assistance software tool (see Figure 3) is needed. A transportation network in Slovakia serves as a sample (Figure 6). The branches, transport resources and other necessary structural data is kept in a database, extracted by the generation part of the simulation software tool and brought to a simulation model. The process flow is hard-coded in the simulation tool itself in a special simulation library. The results of the simulation can be interpreted in key figures such as utilization of resources, shipment times and more (Levinson 2004).

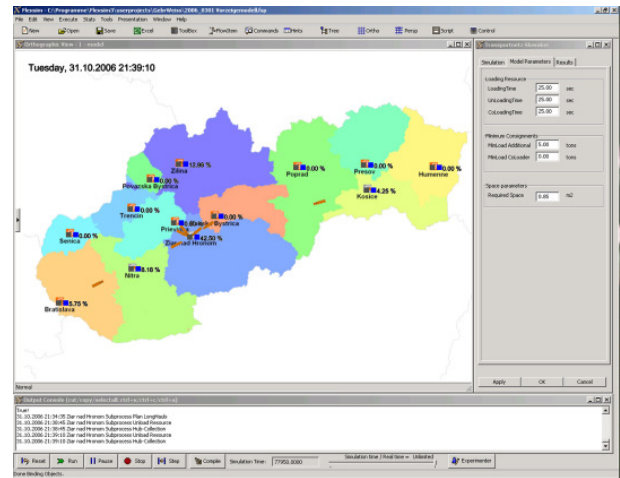


Figure 6: Transportation network sample

The data model is kept easily, the generation complexity straightforward and the process flow not very adaptable but a wide range of transportation networks with its long- and short-distance subtleties can be modelled and analyzed rapidly.

A further extension of this sample is to bring the process flow into the simulation objects it selves. This leads to more manageable and adaptable software code.

##### 5.2. Warehouse

With increasing complexity of the logistics case the data-driven generation of a simulation model becomes more complex as well. This can be seen in a warehouse simulation study (Figure 7). An assistance software tool is essential here and it can be seen that the generation part has to be separated from the simulation tool. This is a crucial advantage for further adaptations to the logistics situation.

Each the warehouse layout and the process flow are modelled outside of the simulation tool by a layout and a process model editor. The generation software combines the layout and the process flow to an executable simulation model.

The generation software in this sample inherits even more functionality by providing schedules for the assignment of warehouse resources and workers. This is useful if own implementations of schedules and strategies shall be applied.

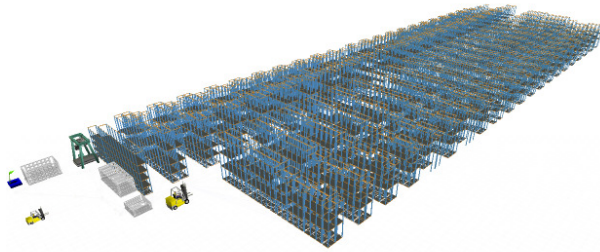


Figure 7: Warehouse sample

### 5.3. Production plant

A simulation software tool can be only a small part of the overall analysing and visualizing system. As seen in Figure 8 which demonstrates a production line. The modelling information about plant units and machines comes from a layout data archive and is translated into simulation objects. The tasks for the production line come from a production server. The simulation tool is responsible for the 3D visualization, the scheduling of the production tasks and the logging of operation data and time. The generation part is quite complex but very adaptable. New production lines can be modelled really quickly. Plant cycle times and machine cycle times can be analyzed via the logged data.

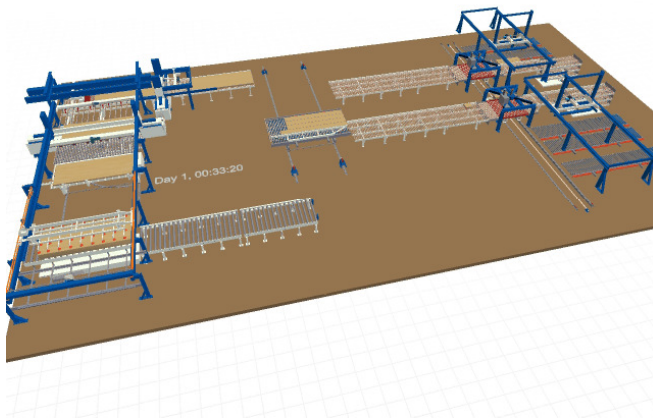


Figure 8: Production plant sample

## 6. CONCLUSION

The simulation of logistics cases helps to achieve a better understanding of the processes taking place. Key figures can be evaluated and different action alternatives can be compared to get a better performance in some respects.

A software framework and data-driven modelling reduce simulation model building time tremendously (Jensen 2007). An additional process editor enhances this technique and provides the capability of separating the structural model data from the operational flow and underlines the possibility of modelling distinct versions of process flow alternatives for a simulation study to compare them regarding some predefined key figures.

Taking altogether data-driven modelling can be a crucial advantage in changing logistics environment.

## REFERENCES

- David Levinson. The evolution of transport networks. Technical report, 2004.
- Sören Bergmann, Alexander Fiedler, and Steffen Straßburger. Generierung und Integration von Simulationsmodellen unter Verwendung des Core Manufacturing Simulation Data (cmsd) Information Model. *Integrationsaspekte der Simulation: Technik, Organisation und Personal*, 2010.
- Samer Hassan, Luis Antunes, and Juan Pavon. A data-driven simulation of social values evolution. *8th International Conference on Autonomous Agents and Multiagent Systems*, 2009.
- Sigrid Wenzel. Modellbildung und Simulation in Produktion und Logistik, Stand und Perspektiven. *ASIM Workshop*, 2009.
- Minghui Yang. Using data driven simulation to build inventory model. *Proceedings of the 2008 Winter Simulation Conference*, 2008.
- Sven Jensen. Eine Methodik zur teilautomatisierten Generierung von Simulationsmodellen aus Produktionsdatensystemen am Beispiel einer Job Shop Fertigung. *ASIM Workshop*, 2007.
- Catrina Kennedy, Georgios Theodoropoulos: Towards intelligent data-driven simulation for policy decision support in the social sciences. *School of Computer Sciences, University of Birmingham, UK*.
- Lothar März, Martin Saler: Ein Analyse-, Planungs- und Entscheidungsinstrument für Lagerlogistikanwendungen, *Advances in simulation for production and logistics applications, Markus Rabe (Hrsg.), Stuttgart 2008*.
- Sigrid Wenzel, Anke Jauss: Diskussion eines Benchmark-Verfahrens für den Vergleich von Simulationswerkzeugen in Produktion und Logistik, *Advances in simulation for production and logistics applications, Markus Rabe (Hrsg.), Stuttgart 2008*.
- Guixiu Qiao, Frank Riddick, and Charles McLean. 2003. New manufacturing modeling methodology: data driven design and simulation system based on XML. In *Proceedings of the 35th conference on Winter simulation: driving innovation (WSC '03)*. Winter Simulation Conference 1143-1148.

## AUTHORS BIOGRAPHY

Rainer Frick has studied Computer Science and is currently studying Business Administration at the University of Innsbruck. His main focus is developing software in the simulation and optimization field at V-Research company.



# EVALUATION OF CLIMATE-EFFICIENT CROSS-COMPANY LOGISTICS MODELS BASED ON DISCRETE EVENT SIMULATION

Christian Hillbrand <sup>(a)</sup>, Susanne Schmid <sup>(b)</sup>

<sup>(a,b)</sup> V-Research GmbH, Industrial Research and Development

(a) [christian.hillbrand@v-research.at](mailto:christian.hillbrand@v-research.at), (b) [susanne.schmid@v-research.at](mailto:susanne.schmid@v-research.at)

## ABSTRACT

Manufacturers shifting production facilities to low-wage countries are usually facing new challenges in transport and logistics management. In practice, two possible approaches to deal with these issues come into consideration: Building up cross-company logistics models within regional clusters and integrating intermodality into transport chains. These concepts do not only lead to an increase of efficiency and reduction of transports through bundling effects but also to a reduction of overall emissions. Yet, the implementation requires integral planning and evaluation models for a holistic perspective of the developed transport chains. In this paper we describe a novel approach for the analysis of multimodal cross-company logistics models based on discrete event simulation. The potential of the proposed simulation and evaluation model is illustrated in the form of a case study of an automotive cluster in the Romanian Timis region.

**Keywords:** simulation, evaluation, green logistics, cross-company logistics models, multimodal transport.

## 1. INTRODUCTION

The tendency for offshoring manufacturing activities to low-wage countries has been extensively researched in the relevant literature (cf. Mucchielli & Sucier 1997; Pennings & Sleuwaegen 1997; Kinkel & Maloca 2009). The main driver for this development is the opportunity to cut labor costs as the biggest share of production expenses. A study of 1663 German companies shows that in 2006, a share of 88% of all respondents stated this as the most important motive for production offshoring. Only 27% moved production facilities because of an intended market opening followed by 26% for avoiding capacity bottlenecks, 20% for the vicinity to customers and 11% because of tax advantages (Kinkel & Maloca 2009).

When analyzing the targets of these offshoring activities one can observe that a large share of companies choose to locate their manufacturing plants in countries where the cultural and geographical distance is not exceedingly large. Therefore about 55% of outsourced manufacturing operations are moved towards new EU member countries in Central and Eastern Europe (CEE). As figure 1 shows neighboring countries like the Czech Republic or Poland are

followed by Hungary, Slovakia and other CEE countries.

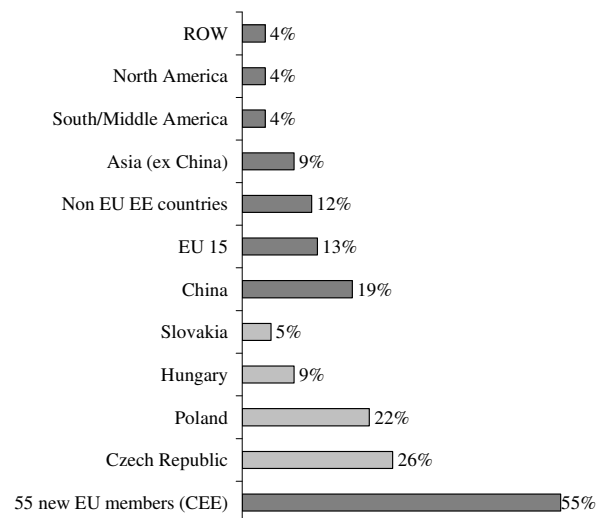


Figure 1: Target countries of production offshoring (cf. Kinkel & Maloca 2009)

One consequence of this specific offshoring behavior is that goods have to be moved back to Western European production facilities for further processing or final assembly. As the new suppliers are within a feasible distance for ground transportation, most of these goods are directly delivered by road-based means of transportation. However, other transport modes like railways or inland waterways are almost irrelevant in this context. This further leads to rising road traffic in transit countries which in turn has negative economic and environmental effects for the concerned regions. As an example, the external costs of road-based traffic (i.e. environmental costs, costs caused by accidents, noise costs) in Austria as a typical transit country amounts for EUR 112.- per transported ton and kilometer (Herry et al. 2007). In total this makes up a sum of EUR 2.4bn per year of which about 37% or EUR 890m are caused by transit traffic. These costs have to be covered by the public sector.

Therefore, from a macroeconomic point of view, the use of more efficient modes of transport like railway or ship is to be promoted. However, as these transport modes require high volumes for efficiency reasons, this leads to the necessity of co-ordination among

consignors. 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 of emissions, costs and logistical competitiveness are targeted on developing new sustainable and energy-efficient logistics models.

## 2. MODEL OF CROSS-COMPANY LOGISTICS NETWORKS

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 laws, goods, finance and information where spatial, quantitative, informational 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.

The mixture of logistics systems made up from the given basic structures is decided in the logistical network structure. The processes are designed when the logistical capacities are superimposed on this. The logistical 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 logistics 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 logistical 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 logistical network structure. The following forms may 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.
- 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 (Brauer 1982). The number of transports between sources and targets can be reduced by the setup of transshipment points from  $m \times n$  to  $m + n$ ,  $m$  and  $n$  being the number of source and target points (Campbell 1990; Simchi-Levi et al. 2007).

Bundled transport over the long run between two 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 of reducing logistics costs while keeping logistics 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 logistics models regarding ecology, economy or logistic competitiveness.

Transport bundling or cross-company logistics 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 logistics 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.

### 3. SIMULATION AND EVALUATION MODEL FOR COOPERATIVE MULTIMODAL LOGISTICS CHAINS

The analysis, evaluation, and comparison of the described multimodal logistics concepts for cross-company bundling base on economical, ecological and logistical key performance indicators. Obtaining these indicators is affected by a multiplicity of factors and there exist numerous interdependencies between the parameters. The specification of a specific resource and a specific routing for instance have influence on each other and can have diverse impact on goal criteria. 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.

For these reasons a simulation and evaluation model has been developed which allows system experts to model, analyze and evaluate co-operative multimodal logistics chains.

In the following chapters the steps **model generation** and **simulation & logging** as well as **result calculation** and **result analysis** are described in detail.

#### 3.1. Conceptual design and automatic generation of the Simulation Model

The model is implemented within the simulation environment Flexsim® (Nordgren, 2003). The conceptual design of the simulation model focuses on providing a generic, easily adaptable model which allows a quick and efficient modeling and analysis of multimodal transportation scenarios. For this reason individual logistics simulation modules (or classes) were developed which individually implement the behavior of the objects as well as the interaction with other objects. These classes include the previously described logistics concepts (point-to-point transportation, consolidation terminals, milk runs, etc.) and can be combined with other modules in order to build multimodal logistics chains.

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 manufacturing facility. This route has one or several transport resources of various types assigned and either implements a direct relation or an intermodal transport, which is usually routed through certain transshipment points. The simulation model can be automatically generated by dropping route-objects into to simulation environment as illustrated in figure 4.

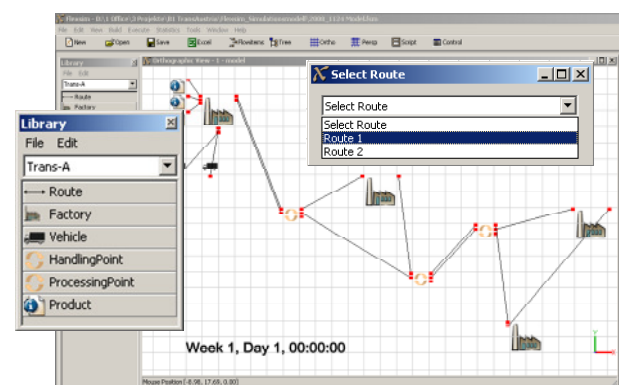


Figure 4: Generation of Simulation Model with Drag&Drop

A route defines the structural data of the model like number and types of plants, transshipment points or transport resources, or the connection between the simulation classes. This information is stored in a database. The behavior of the objects during the simulation run is implemented in the simulation modules. The classes are automatically created, parameterized, and connected with each other when a

route is dropped into the model. The simulation classes include the following:

- **Product:** The system load of the simulation model is determined by the output of products on the consignor's side, which is being processed in form of a tabulated production schedule.
- **Plant/Factory:** A plant serves as a starting point (source) to add products or as a destination to discharge items from the transport (sink). Product objects are created as flow items in the course of simulation, beginning at the starting plant.
- **Vehicle/Transport resources:** Transport resources are active parts within the simulation model. They can be defined as various resource types like trucks or trains.
- **Handling Point:** A handling or transshipment point defines a point on the route where the flows of goods are collected and consecutively assigned to the outgoing transports. This means, the whole transport resource is unloaded and the goods are temporarily stored at the handling point for further transportation.
- **Processing Point:** By means of processing points the transports can be delayed at certain locations on the route. For instance the time-consuming customs handling between frontiers can be modeled as a processing point.

With this modeling structure changes to the simulation parameters (varied transport cycles, use of various transport resources, etc.) can be easily made and a comparison of the effects can be analyzed with the evaluation model which will be discussed below.

### 3.2. Simulation Run and Data Logging

After the creation and parameterization of the generated simulation classes, the simulation run can be initiated. The process starts off with the creation of products at the starting points according to the defined production schedules (see figure 5). These generated flow objects have to be transported on the defined routes by the previously assigned transport resources to the destination plants, where they are destroyed. The transportation process is planned, 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 products are loaded onto the transport resource and then transported to the next stop on the route. If the next stop is a handling point, the products are unloaded and, if needed, prepared for the next shipment. In case of the next stop being a processing point, the idle time is applied and the transport continues. Upon arrival and unloading of the flow items at the destination plants, the respective product objects are destroyed in the simulation environment.

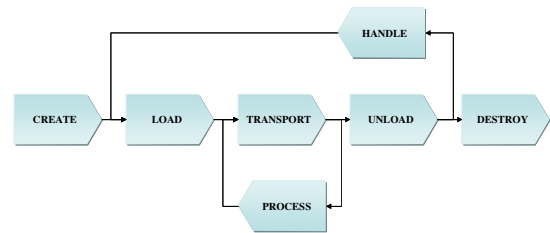


Figure 5: Simulation Processes for Flow Items

Throughout the simulation process, the running time and processing steps are logged and stored into a database. A historiography of the granulated simulation data in a database enables detailed and flexible analyses and a deduction of key performance indicators. With the evolved evaluation model which is discussed in the next section, the (logistic) results drawn from the simulation can be applied to calculations with data sources concerning emission behavior, for cost analyses, and for comparisons whether logistics targets have been met.

### 3.3. Evaluation of Simulation Results

The target criteria, which are relevant for the analysis of the designed co-operative multimodal logistics concepts, can be summarized as

- the minimization of emissions,
- the reduction of logistics costs, and
- the increase of logistical competitiveness.

Therefore, the developed evaluation model is based on the calculation of key performance indicators (KPIs) for the three main target dimensions: Emissions, costs and competitiveness.

Regarding the analysis of emissions, it is in particular the intermodality of the models, which plays an important role when planning multimodal logistics chains. For this purpose, a selection of the most harmful emissions – namely CO<sub>2</sub>, NO<sub>x</sub>, and amounts of particulates – is analyzed as key performance indicators. The emission levels are mainly dependent on the allocated type of transport resource and on the route, which means the distance covered by the predefined route profile. Diesel or electrical power consumption also play an important role in the output of emissions.

The cost evaluation model is somewhat more extensive and the assessment of total costs (1) can be subdivided into three individual cost calculations:

- transport costs (2),
- handling (i.e. transshipment) costs (3), and
- inventory costs (4).

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 forwarding companies. The road charges are particularly difficult to determine due to various systems in the individual countries which is also an important influencing factor for the selection of transportation routes.

$$C_{\text{Total}} = C_{\text{Transport}} + C_{\text{Handling}} + C_{\text{Inventory}} \quad (1)$$

$$C_{\text{Transport}} = C_{\text{Tsp.Truck}} + C_{\text{Tsp.Train}} + C_{\text{Tsp.Plane}} + C_{\text{Tsp.Ship}} \quad (2)$$

$$C_{\text{Handling}} = \sum (Q_{\text{Chg. of Resource}} \cdot C_{\text{Transshipment (Chg. of Resource)}}) \quad (3)$$

$$C_{\text{Inventory}} = C_{\text{Capital}} \cdot C_{\text{Warehousing}} \quad (4)$$

C...Cost, Q ... Quantity

The third criterion in evaluating simulation scenarios is the logistic competitiveness, which is made up of two key performance indicators:

- The ability to deliver: A measure of the extent to which the company can guarantee the logistical service requested by the customer - short delivery times compared to the competition are especially important for a high ability to deliver.
- The 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).

The calculations of the aforementioned key performance indicators are performed independently from the simulation runs of the multimodal transportation scenarios. The logged simulation processing steps which are described in the previous section only serve as input for the computation of the evaluation model. In addition to this historiography, route profiles (types of roads with assigned maximum speeds), altitude profiles (incline), consumption profiles (diesel and power consumption differentiated by resource types), road charge calculation systems and other relevant parameters which are stored in a database, serve as a foundation for the calculations.

As the evaluation criteria of these three dimensions as the result of a simulation study may cause certain trade-offs, an evaluation of a co-operative multimodal scenario of logistics chains can only be performed by comparing different scenarios and considering the pros and cons of all compromises. As an example, the fulfillment of logistical performance expectations will lead to the usage of flexible means of transport which are potentially more expensive and cause higher emissions due to an individual forwarding mode. The simulation and evaluation model offers a transparent and comparable analysis of the logistics system's dynamic behavior in all three dimensions and therefore provides a basis for decision making.

#### 4. HOLISTIC TRANSPORT OPTIMIZATION – A CASE STUDY

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.

##### 4.1. Designing simulation scenarios

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. 6). 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.

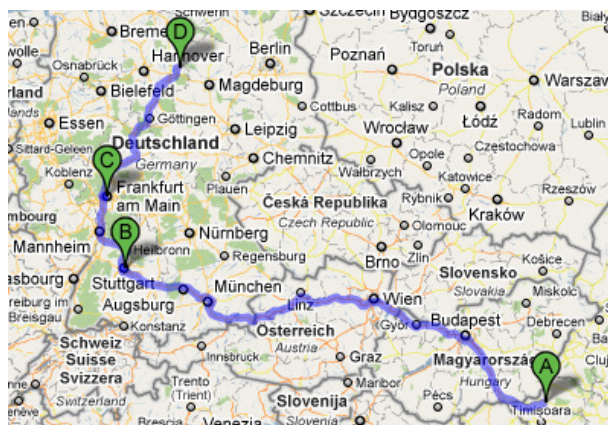


Figure 6: Main Leg in Scenarios 1 and 2

#### 4.2. Simulation run and evaluation of results

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 logistical 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<sub>2</sub> emissions by nearly 40% and NO<sub>x</sub> even by 50% (see fig. 7). 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.

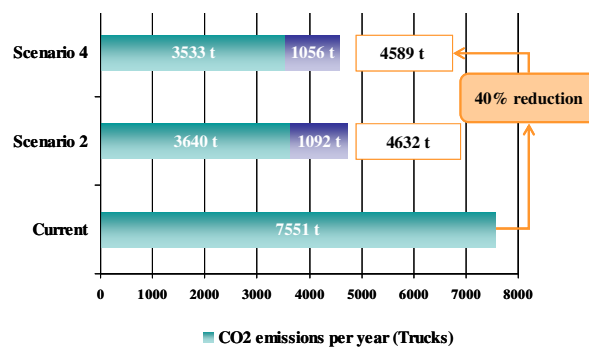


Figure 7: CO<sub>2</sub> emissions per year

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.

## 5. 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 logistical 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 logistical 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.
- Herry, M., Sedlacek, N., Steinacher, I. 2007. *Verkehr in Zahlen Österreich Ausgabe 2007*. Studie in Auftrag gegeben und herausgegeben vom Bundesministerium für Verkehr, Innovation und Technologie Abteilung V/Infra 5, Wien.
- Kinkel, S., Lay, G., Maloca, S. 2004. *Produktionsverlagerungen ins Ausland und Rückverlagerungen*. Report on the BMF research project no. 8/95, Fraunhofer-ISI, Karlsruhe.
- Kinkel, S., Malica, S. 2009. Drivers and antecedents of manufacturing offshoring and backshoring – a German perspective. *Journal of Purchasing and Supply Management*, 15, 154-165.
- Mucchielli, J.-L., Saucier, P. 1997. European industrial relocations in low-wage countries: policy and theory debates. In: Buckley, P.J., Mucchielli, J.-L. (Eds.), *Multinational Firms and International Relocation*. Cheltenham, UK, pp. 5-33.
- Nordgren, William B., 2003. *Flexsim Simulation Environment*. In Proceedings of the 2003 Winter Simulation Conference, Volume 1, eds. Stephen E. Chick, Paul J. Sánchez, David Ferrin, and Douglas J. Morrice, 197-200.
- Pennings, E., Sleuwaegen, L. 1997. International relocation: firm and industry determinants. *Economics Letters* 67, 179–186.
- 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.
- Simchi-Levi, D., Kaminsky, Ph., Simchi-Levy, E. 2007. *Designing and Managing the Supply Chain – Concepts, Strategies and Case Studies, 3<sup>rd</sup> edition*. 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.
- Vahrenkamp, R. 2007. *Logistik: Management und Strategien, 6.Auflage*. Oldenbourg Wissenschaftsverlag, München.
- Wiendahl, H.-P., Lutz, S. 2002. Production in Networks. *Annals of the CIRP* 51, No. 2, 573-586.

## AUTHORS BIOGRAPHY

Christian Hillbrand is head of the research area „Technical Logistics“ at V-Research, an Austrian competence center for industrial research and development. Furthermore he is director of the research program “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.



# BUILDING A SYSTEM FOR AUTOMATED MODELING AND SIMULATION OF PLANTS

Robert Schoech<sup>(a)</sup>, Susanne Schmid<sup>(b)</sup>, Christian Hillbrand<sup>(c)</sup>

<sup>(a, b, c)</sup> V-Research GmbH, Industrial Research and Development

<sup>(a)</sup> [robert.schoech@v-research.at](mailto:robert.schoech@v-research.at), <sup>(b)</sup> [susanne.schmid@v-research.at](mailto:susanne.schmid@v-research.at), <sup>(c)</sup> [christian.hillbrand@v-research.at](mailto:christian.hillbrand@v-research.at)

## ABSTRACT

Planning, implementation and operation of 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 tool 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 automated modeling and simulation tool, 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.

Keywords: Discrete Event Simulation, Emulation, Central Hierarchical Control Systems, Plant Engineering

## 1 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.

## 2 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. 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.

### 2.1 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 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.

## 2.2 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.

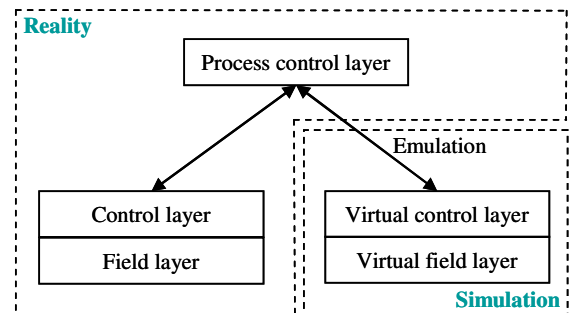


Figure 1: Central, hierarchical Control Systems

Classical plants are centrally and hierarchically controlled real time systems (see figure 1) (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).

### 3 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.

#### 3.3 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 2).

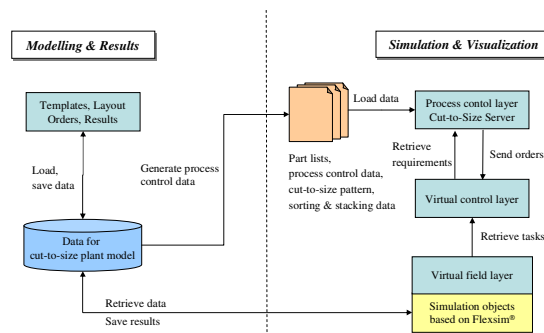


Figure 2: Software Architecture

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. Templates provide a set of unit based components and allow the specification of different types of models, depending on the goal of the modeler. The work flow of the plant is modeled implicitly by the coupling of the individual plant units. The modeled plant is persisted and process control data for the cut-to-size server is generated automatically. Depending on 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 which are 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 3). A single task describes an activity that imitates the physical mechanism which is executed on the simulation object.

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 3) and run time information is logged simultaneously. The received task sequences depend on bills of materials, process control data, cut-to-size patterns as well as sorting and stacking data.

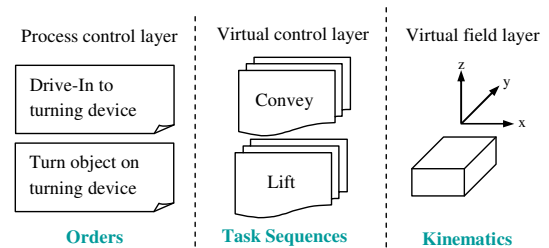


Figure 3: Order Handling

#### 3.4 Modeling process

Our approach to model large-scale plant systems is based on three complementary services called *plant unit template*, *unit-based system model* and *simulation model* (see figure 4).

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* depending on the defined *unit-based system model* description, the creation of all simulation objects and the assignment of values to all public properties.

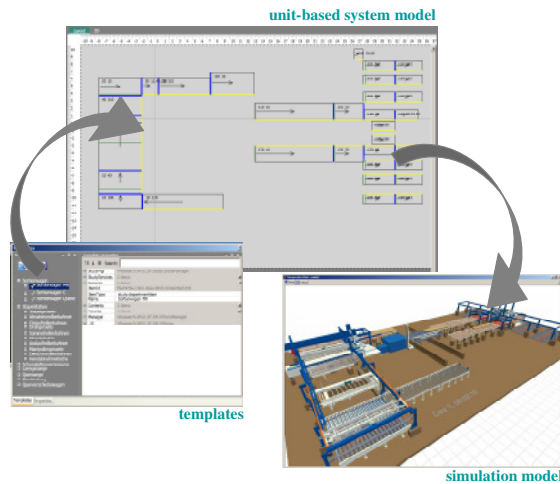


Figure 4: Simulation Services

### 3.5 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 instructions 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 and item operations relate to a panel, respectively parts of a panel.

The implementation of a *blocking rendezvous* pattern 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* of the *virtual control layer* and all activities are registered as tasks and saved in the event list. 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.

### 3.6 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 two kinds of 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 the utilization of plant units and devices (see figure 5) as well as utilization of plant units and devices.

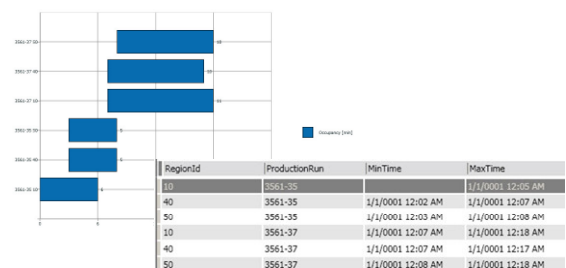


Figure 5: Analysis of Cycle Time of Simulation runs

## 4 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.

#### 4.1 Communication with 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. Depending on the order confirmation, new orders are triggered for processing next task sequences. The implementation of a blocking rendezvous 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.

#### 4.2 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 walkthroughs 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 is 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 comparison of real and virtual plants.

## 5 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.

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.
- Guenther, 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". *IEE 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.; and Wenzel, 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.



# DYNAMICS OF DIRECT TRANSPORTS IN AUTONOMOUSLY CONTROLLED PRODUCTION NETWORKS

B. Scholz-Reiter<sup>(a)</sup>, M. Görge<sup>(b)</sup>

<sup>(a),(b)</sup> BIBA – Bremer Institut für Produktion und Logistik GmbH

<sup>(a)</sup>[bsr@biba.uni-bremen.de](mailto:bsr@biba.uni-bremen.de), <sup>(b)</sup>[goe@biba.uni-bremen.de](mailto:goe@biba.uni-bremen.de)

## ABSTRACT

Autonomous control of logistic processes opens up new potentials to improve the handling of internal and external dynamics in production networks. These networks are characterized by geographically dispersed coupled transport and production processes. This coupling may influence the dynamic behavior of the complete network, due to interdependencies between production and transport quantities and time scales. This paper addresses the impact of two direct transport strategies in an exemplarily production network scenario with autonomous controlled production plants. Relevant parameters of both transport strategies, which determine the transport quantity and the transport time scale, are varied in different simulation experiments in order to explore their impact of the dynamic behavior of the entire network and its elements.

Keywords: production networks, autonomous control, simulation, transportation mode

## 1. INTRODUCTION

Modern logistic systems are exposed to various dynamical changing parameters in their internal and external environment. Especially logistic networks, e.g. production networks or supply chains, are affected by dynamical changes (Sydow 2006, Wiendahl and Lutz 2002). These dynamics may be caused by an increasing desire of customers for individualized goods or the demand for short delivery times and a strict adherence to due dates. On the other hand, internal factors can cause unfavorable dynamic behavior of logistic networks itself, e.g. interdependencies between transportation and production processes. Especially, in production networks with geographically dispersed production facilities, which are sequentially involved in the production process, the (temporal) coordination of transport and production processes gets more and more important (Sauer 2006). In this regard, additional tasks and challenges for production planning and control (PPC) arise, like the assignment of orders to plants. Under these highly dynamic and complex conditions, current PPC methods cannot cope with disturbances or unforeseen events in an appropriate manner (Kim and Duffie 2004). The implementation of autonomous

control is a promising approach to cope with increasing dynamics. This approach proposes decentralized coordination and decision-making of intelligent objects within a logistic system or network. It aims at improving the logistic performance due to flexible coping with dynamic complexity (Phillip et al. 2007). First approaches of autonomous control in production networks have been developed: these models have shown that autonomous control methods can improve the ability of a production network to handle dynamics, as well as the logistic performance of the network. Regarding the coupling of transport and production processes, these models revealed that complex interdependent dynamic effects can occur, which affect the logistic performance of the total system (Scholz-Reiter et al. 2009). Due to these coupled processes, the time scale of deliveries and transport quantities, as well as the logistic performance (measured in e.g. through put times, or work in process) of such networks are interrelated (Stadtler 2007). For example, a delivery according to a fixed transport schedule provides fixed delivery intervals, but the quantity depends on the production output of the previous production stages. Especially, the application of local autonomous control strategies requires knowledge about these interdependencies: autonomous control generates a more flexible and dynamic system behavior, which may lead to unfavorable dynamics in combination with a certain transport strategy.

Thus, this paper aims on investigating two different direct transport strategies: a fixed schedule strategy (FS) and a capacity based strategy (CS). These transport strategies are implemented in an autonomous controlled production network scenario. The model will be analyzed regarding varying transport quantities, varying time scales and the corresponding dynamic behavior of the network.

The paper is structured as follows: section 2 gives an overview about the concept of autonomous control in manufacturing. Section 3 focuses on production networks. The particular production network scenario will be presented in section 4. Section 4.1 describes the general structure of the network. Section 4.2 offers a description of the applied autonomous control methods. A detailed description of both direct transport strategies

provides section 4.3. Subsequently, the simulation results are presented in section 5. Finally section 6 gives a summary and an outlook with further research directions.

## 2. AUTONOMOUS CONTROL IN MANUFACTURING

The collaborative research centre 637 ‘Autonomous cooperating Logistic Processes: A Paradigm Shift and its Limitations’, which is founded by German research foundation, gives the following general definition of autonomous control: “Autonomous control describes processes of decentralized decision-making in heterarchical structures. It presumes interacting elements in non-deterministic systems, which possess the capability and possibility to render decisions independently. The objective of autonomous control is the achievement of increased robustness and positive emergence of the total system due to distributed and flexible coping with dynamics and complexity.” (Windt and Hülsmann 2007). According to this definition, the main idea of autonomous cooperating logistic processes is a shift of decision-making capabilities from the total system to its elements. In the context of production systems, or production networks, intelligent logistic objects are allowed to route themselves through the logistic network according to their own objectives (Wiendahl and Lutz 2002). The term intelligent logistic object is comprehensively defined. It covers physical objects (e.g., machines, parts, etc.) as well as immaterial objects like production orders (Windt 2006, Philipp et al. 2007).

Recent work on autonomous controlled production systems showed that the application of autonomous control improves the logistic target achievement as well as the handling of internal and external disturbances (Scholz-Reiter et al. 2009b, Armbruster et al. 2006, Scholz-Reiter et al. 2005). As far as production networks are concerned, first approaches also provided promising results (Scholz-Reiter et al. 2007, Dashkovskiy et al. 2011). Scholz-Reiter et al. 2009 modeled a transport dispatching rule with fixed time intervals. It was shown that increasing inter transport times (ITT) may cause sudden changes in the total system performance and lead to an unpredictable system behavior. In order to analyze these dynamic effects, this paper focuses on two transport strategies in a similar scenario. It aims on evaluation of these strategies and on identifying relevant dynamic effects. The next section gives a brief overview about production networks in general.

## 3. PRODUCTION NETWORKS

Production networks are company or cross-company owned networks of geographically dispersed production facilities. Production networks focus on the mutual use of common resources and integrated planning of value adding processes in the network (Wiendahl and Lutz 2002). On the one hand this allows the achievement of economies of scale through the joint planning and use

of production resources. On the other hand, these types of networks may react fast on internal or external disturbances due to redundancies of resources.

An integrated view on production planning and control and transports generates additional tasks: companies have to generate concepts for identifying new network partners, for the network design and for adjusting the PPC according to the network’s purpose (Sydow 2006). The interconnection of production facilities opens up potentials of dealing flexible with disturbances. However, this creates complex interdependencies between production planning and control of plants and coordination of transports, e.g. decisions about assigning parts to plants or planning of transports and transport capacity (Sauer 2006, Alvarez 2007). There are first approaches aiming at optimization of combined production and transport processes (Erengüc et al. 1999). Due to the high degree of structural complexity of these problems, a complete optimization of large problem instances seems to be difficult. Thus, in this paper direct transport connections between plants are assumed. Direct shipping describes transport processes, which connect two locations directly, without any kind of transshipments (Gudehus 2005). Literature provides several optimisation approaches concerning transport frequency, transport costs and inventory costs of senders and recipients (e.g., Bertazzi and Speranza 2005, Wagner 2006).

This paper adapts a fixed schedule based and a capacity based direct transport strategy for initiating transports between two plants (Gudehus 2005). It focuses rather on the dynamic implications of these strategies on the networks performance than on the concrete optimisation.

## 4. PRODUCTION NETWORK SCENARIO

A scalable production network scenario with  $j \times k$  different production locations is considered. This network comprises  $j$  different stages and  $k$  production plants per stage. Furthermore each plant in this scenario contains a shop floor with  $n$  parallel machines which are collocated on  $m$  different stages. Figure 1 depicts this scenario.

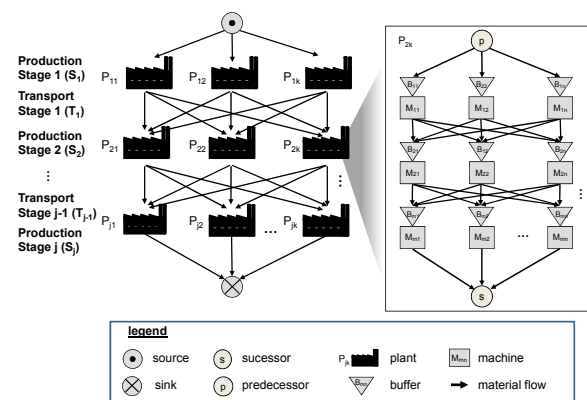


Figure 1: Production network scenario with  $j \times k$  plants and  $m \times n$  machines per plant

Each plant of a certain stage is connected to all plants on the succeeding stages via transport routes (see Figure 1). To analyze the impact of different direct transport strategies on the performance of an autonomous controlled network, a discrete event simulation model of a concrete scenario was build (similar to Scholz-Reiter et al. 2009). The following sections describe the parameterization of this scenario regarding: network structure, plant configuration and transport configuration.

#### 4.1. Network structure

This scenario contains six different plants on four stages. On the first and on the last stage only one plant is located, which function as a source or sink, respectively. Stages two and three comprise two parallel plants each. The transport distances between all plants are set equally to 140 km and the speed of a truck is set to 70 km/h. Accordingly, transports between plants have a duration of 2 h. Table 1 summarizes the transport connections and distances for this scenario.

Table 1: distance matrix [km]

from / to	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>
P <sub>1</sub>	-	140	140	-	-	-
P <sub>2</sub>	-	-	-	140	140	-
P <sub>3</sub>	-	-	-	140	140	-
P <sub>4</sub>	-	-	-	-	-	140
P <sub>5</sub>	-	-	-	-	-	140

The considered network is able to process three different job types (Type A, Type B and Type C). All jobs arrive at plant P<sub>1</sub> and have to pass all stages of the network up to plant P<sub>6</sub>. To model demand fluctuation, the arrival rate of parts is set as sine function (1). This function has a phase shift  $\varphi$  of 1/3 of a period for each job type, so that the maximal arrival rates of all job types do not cumulate.

$$\lambda(t) = \lambda_m + \alpha \cdot \sin(t + \varphi) \quad (1)$$

The mean arrival rate  $\lambda_m$  is set to 0.4 1/h. Due to this mean arrival rate in average every 2.5 h a part arrives. The amplitude  $\alpha$  determines the intensity of the arrival rate fluctuation. It is set to  $\alpha=0.2$  1/h which causes variations of the inter-arrival time between 1.5 h and 3 h.

#### 4.2. Plant configuration

Every plant in this scenario comprises a shop floor with three production stages and three parallel machines and buffers on each stage. Parts have to pass all stages of a plant. There are different processing times for each job

type on the parallel machines of a certain production stage. Table 2 summarises these processing times.

Table 2.: Processing times of all job types on different processing lines [h:mm]

Plant	P <sub>1</sub> , P <sub>6</sub>			P <sub>2</sub> , P <sub>3</sub> , P <sub>4</sub> , P <sub>5</sub>		
Line / Type	1	2	3	1	2	3
Type A	2:00	3:00	2:30	4:00	5:00	4:30
Type B	2:30	2:00	3:00	4:30	4:00	5:00
Type C	3:00	2:30	2:00	5:00	4:30	4:00

Two different autonomous control methods are implemented: the queue length estimator method (QLE) and the pheromone based method (PHE). Both methods enable decentralised autonomous decision-making of parts. Parts using the QLE method compare the buffer levels of each production line at a certain stage and calculate their actual workload. In order to reduce their own throughput time (TPT) parts choose the machine with the lowest workload (for a detailed description see Scholz-Reiter et al. 2005).

The second method (PHE) aims on reducing the TPT, as well. In contrast to the QLE method, this method is inspired by the process of ants marking possible routes to food sources with pheromone trails. After being processed, parts leave information about waiting and processing times at the machine as a kind of artificial pheromone. Following parts are able to detect this information and choose the machine with the lowest mean value of artificial pheromones. The natural evaporation of pheromones is emulated by taking the moving average of the last five parts (for a detailed description see Armbruster et al. 2006, Scholz-Reiter et al. 2007).

#### 4.3. Direct shipping strategies

Two simple direct shipping strategies are implemented in this scenario: fixed schedule transport strategy (FS) and capacity based transport initiation (CS). The FS initiates transports from one plant to the next plant according to fixed schedules, i.e. fixed transport intervals ( $TI$ ). In cases of constant production rates and demands the transported quantities can be easily calculated (Gudehus 2005). On the other hand, in cases of varying production output the transported quantity  $q$  depends on the production output in a certain time interval  $TI$ . Contrary the CS strategy initiates transports according to a predefined transport capacity. Here the transport starts, when the number of produced parts equals this capacity ( $C$ ). The transport quantity is constant and inter-transport time between transports depends on the production rate. Figure 2 visualises exemplarily the connection between both transport strategies.

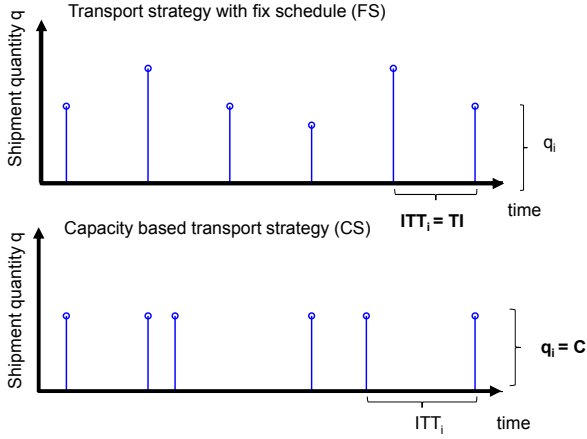


Figure 2: Shipment quantity and transport intervals of FS and CS

Figure 2 depicts two effects: the inter-transport times ( $ITT$ ) for the FS are constant and equal to the predefined  $TI$ , but the shipment quantity may differ according to the production rate. In the CS, case the transport quantity  $q$  equals the predefined capacity  $C$ , but the  $ITT$  may vary between transports.

One can expect that both transport strategies affect the production network differently. In order to analyse the dynamic impact of both transport strategies, different simulation experiments are set up.

## 5. SIMULATION AND RESULTS

The analysis of the impact of transport strategies on the network dynamics is organised as follows: section 6.1 focuses on the connection between inter-transport times, transport quantities and the logistic performance of the entire network. The following section 6.2 gives a detailed analysis of the performance of the network stages.

### 5.1. Transport quantities and transport intervals

In order to investigate the influence of relevant parameters of both transport strategies (FS and CS) on the performance of the total network, several simulation runs are conducted with different configurations of  $TI$  and  $C$ . Figure 3 (a) depicts the quantities, which are transported along all transport connections of the network using the FS strategy. Each point in Figure 3 (a) represents the average of transported quantities  $\bar{q}$  for a simulation run with a certain  $TI$  according to:

$$\bar{q}_{TI} = \frac{\sum q_i}{n}$$

Where  $\bar{q}$  denotes the mean transport quantity per transport,  $q_i$  the quantity of the  $i^{th}$  transport and  $n$  the number of transports during the simulation period. The  $TI$  is increased in steps of 0.5 h per simulation run.

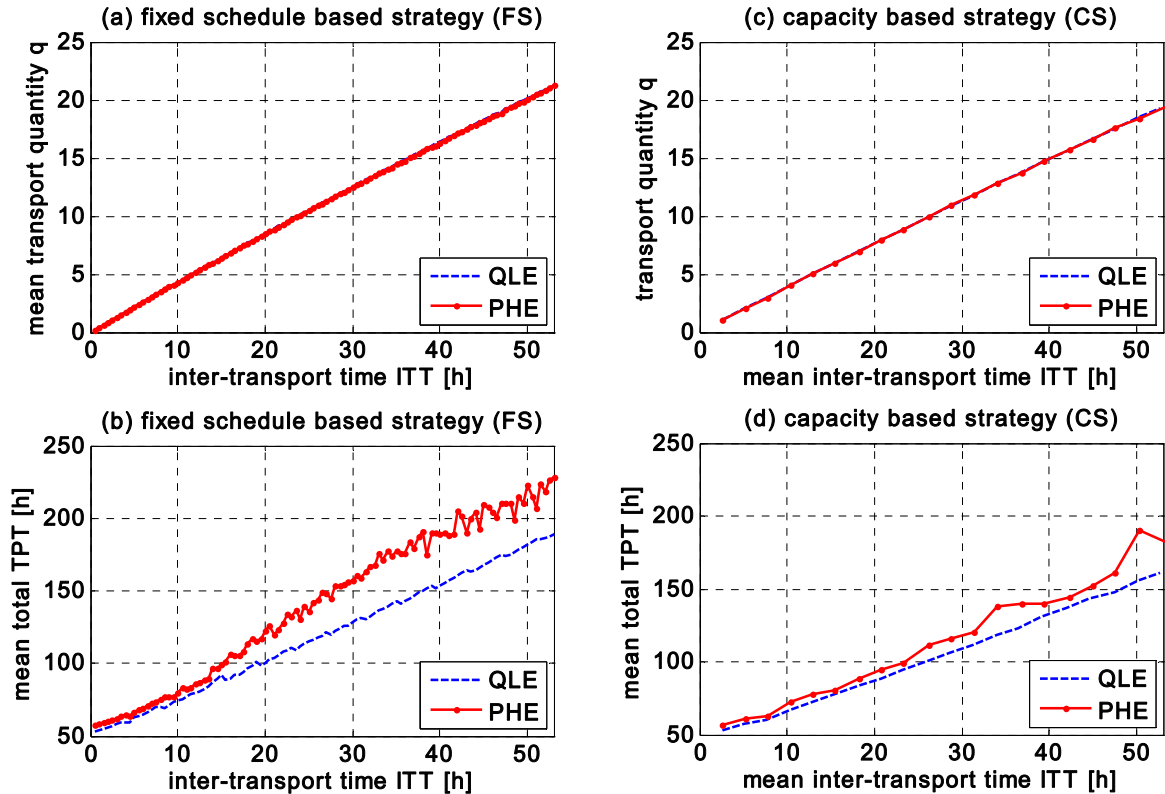


Figure 3: Mean transport quantities of FS against  $ITT$  (a), mean total TPT for FS against  $ITT$  (b), transport quantity of CS against mean  $ITT$  (c), mean TPT for CS against mean  $ITT$  (d)

Figure 3 (a) depicts a rising trend of transport quantities for an increasing  $TI$ . This can be explained by the production output of the pre-located plants, which sums up during the  $TI$ . Thus, transport quantity rises with the  $TI$ . Regarding the impact of the autonomous control strategies inside the plants on the mean transport quantity, one can notice, that the difference between both autonomous control methods in the mean transport quantities is very low. The curves for the QLE and the PHE methods overlap almost (Figure 3 (a)). Thus, it can be concluded that both methods do not affect the transported quantities in dependence of the  $TI$ . With regard to the performance of the entire network, Figure 3 (b) shows the mean total TPT for the corresponding  $ITT$ . In this context, the mean total TPT is measured as the mean time span that parts need to pass through the entire network. Comparing the mean TPT for the FS strategy and both autonomous control method in Figure 3 (b), one can find a significant difference between the QLE and the PHE method. The total TPT of PHE method is bigger for all  $TI$ . Especially higher  $TI$  values lead to a bigger difference of both methods. For the investigated range of  $TI$ , the average difference in total TPT between the QLE and the PHE method amounts 17.23%. In contrast to the average transport quantities, which seem to be independent from the autonomous control method, a correlation between the FS strategy and the autonomous control methods can be found here. It is assumed that both autonomous control methods have a different sensitiveness to a periodic transport interval  $TI$ . Especially the PHE method leads to a high total TPT in this case.

Contrary to the FS strategy, the CS strategy leads to transports with constant transport quantities and varying inter-transport times. Figure 3(c) and Figure 3(d) present the results of the CS. In particular Figure 3 (c) depicts the average inter-transport times  $\overline{ITT}$  for in dependence of increasing fixed transport quantities  $C$ :

$$\overline{ITT} = \frac{\sum ITT_i}{n}$$

Where  $ITT_i$  denotes the inter-transport times between transports,  $C$  the predefined quantity and  $n$  the amount of total transports. The quantity  $C$  is increased in steps of one piece per simulation run. To provide comparability to Figure 3 (a) the axis of dependent variable ( $\overline{ITT}_c$ ) and the independent variable ( $q$ ) are switched.

Figure 3 (c) shows that the impact of the QLE and the PHE method on the inter-transport times is low. Similar to Figure 3 (a) both curves are almost identical. Furthermore, Figure 3 (d) exposes that the total TPT for the CS in dependence of the corresponding mean  $\overline{ITT}$ . For the CS strategy, the QLE method performs better than the PHE method. In this case the relative difference between both methods is less than for the FS. However, it amounts 8.8% for the CS. This leads to the conclusion that the PHE method is more sensitive to fluctuation quantities arriving quantities (Figure 3 (b)) than to fixed quantities which have varying  $ITT$  (Figure 3 (d)). A

similar conclusion can be drawn for the QLE method, but here the sensitiveness is less compared to the PHE method. In general, both local autonomous control methods perform better, if the CS strategy is applied.

## 5.2. Network performance

The results of the previous section showed, that this autonomous controlled network performs best using the CS strategy. This section investigates the impact of both transport strategies on the single stages of the network. For the further analysis, parallel located plants are seen as a production stage  $S_n$  and parallel transport connections as a transport stage  $T_n$  (see Figure 1). This allows a stagewise analysis of performance measures and a comparison of their changes on the different stages.

Figure 4 presents exemplarily the cumulated TPT of all network stages during a simulation period of 120 days. Each fraction in Figure 4 represent the TPT of parts in a production or a transport stage. Generally, Figure 4 contains simulation results of the PHE method for different  $TI$  values for the FS ( $TI=7$  in Figure 4 (a),  $TI=50$  in Figure 4 (b)) and  $C$  values for the CS ( $C=3$  in Figure 4 (c),  $C=19$  in Figure 4 (d)). These values of  $TI$  and  $C$  are exemplarily chosen in order to compare situations with small inter-transport times and big inter-transport times. Furthermore, these parameters lead to comparable mean inter-transport times and quantities according to Figure 3 (a) and Figure 3 (c).

Comparing Figure 4 (a) and Figure 4 (c), both strategies show a similar dynamic behaviour regarding the TPT. There are only small differences of TPT between both strategies per stage. The average difference is 1.44 h per stage. The situation with  $TI=50$  and  $C=19$  contrasts this. A difference between the dynamic behaviour in Figure 4 (c) and Figure 4 (d) can be clearly seen. Indeed the TPT through all stages is in average 21.12 % higher for the FS compared to the CS. Particularly the performance of the last production stage differs for both transport strategies. In the CS situation the mean TPT of the last stage is 36.46 h lower than in the FS situation. On this stage, there is only one plant which consolidates the material flow of the previous stages. It can be assumed that due to this consolidation dynamic fluctuation of previous stages in TPT sum up on this last stage and change the performance dramatically. Additionally, the FS strategy seems to amplify this effect. The highest mean TPT is recorded for this strategy. In contrast to this, the CS strategy reacts different. The maximum mean TPT can be found on the production stage  $S_3$ , followed by the mean TPT of production stage  $S_4$ . This implies that in the CS situation the effect of dynamic fluctuation summation is distributed on the last two stages. Table 3 confirms this. It presents the mean TPT and the standard deviation of the TPT for all stages of the network for both autonomous control methods and both transport strategies. It presents furthermore the mean total TPT (tTPT) and the total standard deviation (std) of the production network.

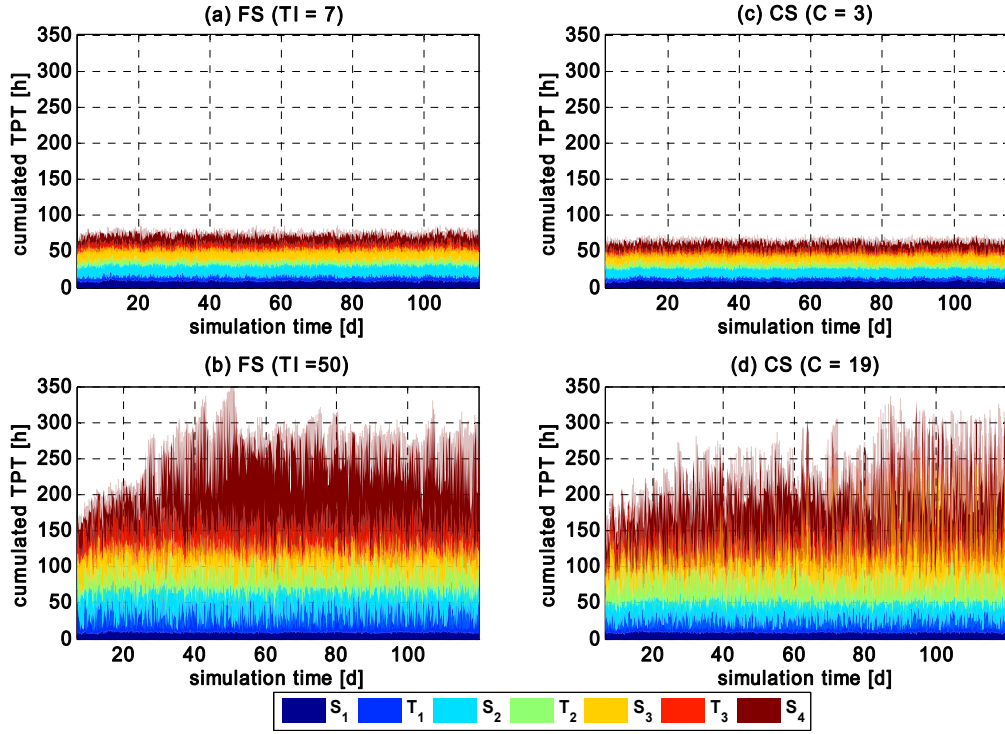


Figure 4: Cumulated TPT of all network stages for PHE against simulation time for FS and CS strategies: FS  $TI=7h$  (a), CS with  $C=3$  (b), FS with  $TI=50h$  (c), CS with  $C=19$  (d)

For the FS situation, the standard deviation of all production location, except for  $S_4$ , are similar, but the biggest value can be found on the last stage (see Table 3). In contrast to this, the maximum values of the standard deviation can be found on  $S_3$  and  $S_4$ . Furthermore, Table 3 presents the results of simulation runs with the QLE method and both transport strategies. Generally, the performance of the QLE method follows a similar trend concerning the impact of both transport strategies. The QLE method performs best in case of the application of the CS. In contrast to the PHE method, the effect of converging material flows on the last production stage is much lower. Especially the results

of the QLE method in combination with the FS with a transport interval of 50 h clarify that. The mean TPT of the last production stage is almost equal to every other network stages. The incoming workload can be processed better compared to the PHE method. Thus waiting and processing times in the last stage are lower.

Summarising these results concerning the logistic performance of single network stages, one can say, that the choice of the transport strategy influences the dynamic behaviour of the total network. The selected network scenario performs generally better using the CS strategy.

Table 3: Mean TPT and standard deviation of TPT of all network stages for all autonomous control methods (ACM) and transport strategies

Strategy	FS	FS	FS	FS	CS	CS	CS	CS
Parameter	$TI=7$	$TI=50$	$TI=7$	$TI=50$	$C=3$	$C=19$	$C=3$	$C=19$
ACM	QLE	QLE	PHE	PHE	QLE	QLE	PHE	PHE
avg. Q [p]	3.03	20.12	3.03	20.08	3	19	3	19
avg. ITT [h]	7	50	7	50	7.67	50.17	7.68	50.2
TPT S1 [h]	7.44	7.45	9.21	9.23	7.45	7.46	9.21	9.23
TPT T1 [h]	6.83	28.34	6.88	28.35	4.99	18.34	5.00	18.38
TPT S2 [h]	14.82	30.55	15.58	30.79	13.14	23.25	13.25	24.56
TPT T2 [h]	6.80	26.02	6.87	25.95	6.32	31.81	6.16	32.38
TPT S3 [h]	15.16	33.02	15.73	33.49	14.60	36.60	14.82	50.43
TPT T3 [h]	6.78	26.50	6.87	26.19	4.51	17.43	4.22	18.19
TPT S4 [h]	10.11	30.56	11.51	83.05	8.97	20.42	9.87	46.59
tIPT [h]	67.80	181.6	72.68	237.1	59.92	155.5	62.44	191.4
std S1 [h]	0.91	0.91	1.11	1.11	0.91	0.91	1.11	1.11
std T1 [h]	2.03	14.41	2.02	14.38	1.44	9.17	1.46	9.16
std S2 [h]	2.05	12.25	2.09	12.49	1.21	7.88	1.06	8.71
std T2 [h]	1.92	13.55	1.90	13.55	2.65	17.61	2.57	17.92
std S3 [h]	2.11	13.00	2.14	12.47	2.19	15.77	2.08	35.07
std T3 [h]	1.95	14.33	1.91	14.27	1.23	9.63	0.99	9.67
std S4 [h]	2.22	14.24	2.37	39.78	1.57	7.99	1.67	17.34
tstd [h]	3.56	18.49	4.34	46.04	3.85	22.26	3.91	25.64



## 6. SUMMARY AND OUTLOOK

A simulation model of an autonomous controlled production network scenario was introduced. It was shown that the transport performance, i.e. the inter-transport times and the transported quantities, of both strategies can be adjusted by varying the relevant parameters, i.e. transport interval and transport capacity. Furthermore it was shown that the logistic performance of the total network differs between both strategies, although the transport performance is equal. A deeper comparison of the performance of the single network stages revealed that the PHE method handles the incoming converging material flow on the last production stage better in the case of applying the CS strategy. In combination with the QLE method both transport strategies harmonise the material flow more, compared to the PHE method. In particular the QLE method performs superior in combination with the CS. Basically, this paper presented that interdependencies between transport and production processes can not be neglected in the design of autonomous controlled production networks. Thus, further research will focus on the design of autonomous control methods, which take these dynamical aspects into account. Additionally, design of autonomous control methods which enable an autonomous assignment of parts to plants seems to be promising. On the other hand investigations of intelligent buffering policies, which allow to damp dynamic variations, are possible fields of future research activities.

## ACKNOWLEDGMENTS

This research is funded by the German Research Foundation (DFG) as part of the Collaborative Research Centre 637 "Autonomous Cooperating Logistic Processes: A Paradigm Shift and its Limitations"

## REFERENCES

- Alvarez, E., 2007. Multi-plant production scheduling in SMEs. *Robotics and Computer-Integrated Manufacturing* 23 (6), pp. 608–613
- Armbruster, D., de Beer, C., Freitag, M., Jagalski, T., Ringhofer, C., 2006. Autonomous Control of Production Networks Using a Pheromone Approach. *Physica A* 363 (1), pp. 104–114
- Bertazzi, L., Speranza, M. G., 2005. Worst-case analysis of the full load policy in the single link problem. *International Journal of Production Economics* 93–94, pp. 217–224
- Dashkovskiy, S.; Görges, M.; Kosmykov, M.; Mironchenko, A.; Naujok, L.: Modeling and stability analysis of autonomously controlled production networks. In: *Logistics Research*, 3(2011)2, pp. 145-157
- Erengüc, S.S., Simpson, N.C., Vakharia A.J., 1999. Integrated production/distribution planning in supply chains: An invited review. *European Journal of Operational Research* 115(2), pp. 219–236
- Gudehus, T., 2005. *Logistik*. Springer, Berlin
- Kim, J. -H, Duffie, N.A., 2004. Backlog Control for a Closed Loop PPC System. *Annals of the CIRP* 53 (1), pp. 357–360
- Philipp, T., de Beer, C., Windt, K., Scholz-Reiter, B., 2007. Evaluation of Autonomous Logistic Processes—Analysis of the Influence of Structural Complexity. In: Hülsmann, M., Windt, K. (Eds.) *Understanding Autonomous Cooperation and Control in Logistics-The Impact on Management, Information and Communication and Material Flow*. pp. 303–324. Springer, Berlin
- Sauer, J., 2006. Modeling and solving multi-site scheduling problems. In: van Wezel, W., Jorna, R., Meystel, A. (Eds.) *Planning in Intelligent Systems: Aspects, Motivations and Method*. pp. 281-299. Wiley
- Scholz-Reiter, B., Freitag, M., de Beer, C., Jagalski, T., 2005. Modelling and Analysis of Autonomous Shop Floor Control, In: *Proc. of 38th CIRP International Seminar on Manufacturing Systems*. Florianopolis, Brazil
- Scholz-Reiter, B., de Beer, C., Jagalski, T., 2007. Selbststeuerung logistischer Prozesse in Produktionsnetzen. *Industrie Management* 23 (1), pp. 19–23
- Scholz-Reiter, B., Görges, M., Jagalski, T., Mehraei, A., 2009. Modelling and Analysis of Autonomously Controlled Production Networks. *Proceedings of the 13th IFAC Symposium on Information Control Problems in Manufacturing*. pp. 850-855. Moscow, Russia
- Scholz-Reiter, B., Görges, M., Philipp, T., 2009b. Autonomously controlled production systems-Influence of autonomous control level on logistic performance. *CIRP Annals - Manufacturing Technology* 58(1), pp. 395-398
- Stadtler H., 2005. Supply chain management and advanced planning-basics, overview and challenges. *European Journal of Operational Research, Supply Chain Management and Advanced Planning* 163(2), pp. 575-588
- Sydow, J., 2006. Management von Netzwerkorganisationen – zum Stand der Forschung, In: Sydow, J. (Ed.), *Management von Netzwerkorganisationen*. pp. 385-469. Gabler, Wiesbaden
- Wagner, B., 2006. *Hub and Spoke-Netzwerke in der Logistik*. Deutscher Universitäts-Verlag/GWV-Fachverlage, Wiesbaden ()
- Wiendahl H.-P, Lutz, S., 2002. Production in Networks. *Annals of the CIRP – Manufacturing Technology* 51 (2), pp. 1–14
- Windt, K., 2006. Selbststeuerung intelligenter Objekte in der Logistik. In: Vec, M., Hütt, M., Freund, A.

- (Eds.) Selbstorganisation – Ein Denksystem für Natur und Gesellschaft. Böhlau Verlag, Köln
- Windt, K., Hülsmann, M., 2007. Changing Paradigms in Logistics - Understanding the Shift from Conventional Control to Autonomous Cooperation and Control. In: Hülsmann, M.; Windt, K. (Eds.) Understanding Autonomous Cooperation and Control - The Impact of Autonomy on Management, Information, Communication, and Material Flow. pp. 4–16. Springer, Berlin

## AUTHORS BIOGRAPHY

**Bernd Scholz-Reiter** (born in 1957) holds a degree in Industrial Engineering from the Technical University Berlin (TUB). Following his PhD (TUB, 1990), Bernd Scholz-Reiter served as post-doctoral fellow in the department for Manufacturing Research at IBM T.J. Watson Research Center in Yorktown Heights, U.S.A. From 1994 to 2000 he held the Chair for Industrial Information Systems at the newly founded Brandenburg Technical University at Cottbus, Germany. In 1998 he founded the Fraunhofer Application Center for Logistic Systems Planning and Information Systems in Cottbus, which he headed until 2000.

Since November 2000 Bernd Scholz-Reiter is full professor and holds the new Chair for Planning and Control of Production Systems at the University of Bremen, Germany. Since 2002 he also serves as Managing Director of the BIBA – Bremer Institut für Produktion und Logistik GmbH (BIBA) at the University of Bremen.

Since July 2007, Bernd Scholz-Reiter is Vice President of the German Research Foundation (DFG).

**Dipl.-Wi.-Ing. Michael Görges** works as a research scientist at the BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen in the division Intelligent Production and Logistic Systems.

# ADVANCED DESIGN OF A STATIC DRYER FOR PASTA WITH SIMULATION TOOLS

Armenzoni Mattia<sup>(a)</sup>, Solari Federico<sup>(a)</sup>, Marchini Davide<sup>(a)</sup>, Montanari Roberto<sup>(a)</sup>, Ferretti Gino<sup>(a)</sup>

<sup>(a)</sup>Department of Industrial Engineering, Faculty of Engineering, University of Parma  
Viale G. P. Usberti, 181 - 43100 Parma

[mattia.armenzoni@gmail.com](mailto:mattia.armenzoni@gmail.com)

## ABSTRACT

In this work, a scientific approach was applied to the drying of pasta in order to obtain a simulative model for this particular industrial process.

At first, it was simulated the evolution of temperature and internal moisture inside a simple geometry product (spaghetti), using a Finite Difference Model (FDM) simulation, implemented in Microsoft Excel. In this way, it was possible to produce a spreadsheet that automatically calculates temporal progression of temperature and moisture in spaghetti of infinite length. Then, the results obtained were subsequently validated through the comparison with a real industrial case;

Later, it was analyzed a layout of static dryer for pasta, in this case various design parameter were determined through CFD.

Keywords: Static dryer, Pasta, FDM analysis, CFD analysis.

## 1. INTRODUCTION

Drying is one of the first technical operations of food preservation used, observed and studied by the industrial technology (Singh and Heldman, 2001). In particular, the drying of pasta is one of the oldest operation developed in food processing. In fact, the fresh pasta cannot be stored for a long time, and drying has represented for many centuries the unique option to increase product shelf life (Milatovich and Mondelli, 1990).

Secondly, due to heterogeneity of the dough base, the successful implementation of the process is essential to obtain a homogeneous product by aesthetic, chemical, physical and organoleptic point of view (Rizzo et al., 2010). In order to obtain the best final product, dryer must ensure the thermodynamic equilibrium that can lead the dough to lose its moisture as homogeneously as possible.

## 2. THE DRYING OF PASTA

In the sector of pasta, the drying method most frequently adopted is drying the product through hot air flow (Migliori et al. 2004). In detail, this process is the union of two distinct thermodynamic processes: the heat

flows from the drying media (the air) to the product and the moisture flows from the product to the drying media. The process is governed by transient thermal and mass transport phenomena, and the mechanism of moisture transfer depends on physical and chemical conditions of the product and of the drying media.

The initial different moisture content between product and heating media is the motor of the whole process, which looks for a thermodynamic equilibrium by the extraction of water from the food matrix to the drying media.

This extraction happens through two different mechanisms:

1. The internal water diffusion across the solid matrix of pasta. This movement brings the water at the product-air interface;
2. The evaporation of the moisture, in the form of vapor, from the interface to the external air (forced convection).

In this study it was reproduced an high temperature (HT) cycle of drying for long pasta (Milatovich and Mondelli, 1990).

This particular process is divided in two phases: pre-drying (air drying at 36 °C for 30 minutes) and drying (air drying at 76°C for 11-12 hours)

The work is divided in two different simulations of the drying process:

- At first, it was studied the product's internal flow of moisture, through the analysis of heat and mass transfer between the product and hot air;
- Secondly it was analyzed the layout of static dryer through external simulation of the thermal process.

## 3. INTERNAL SIMULATION

FDM was utilized in order to describe the phenomena that happens inside the product during a cycle of drying.

In order to implement this method, it was used the software MS Excel<sup>®</sup>, with the purpose to develop a predictive system that could become a future design tool. Thanks to this tool, generic users could recreate a drying's cycle, to verify the effect of this cycle on the

product, in terms of internal evolution of temperature and moisture.

### 3.1. Geometry of product

A simply geometry of the product was adopted for this analysis. It was considered a cylindrical geometry of pasta (spaghetti, for example), with 10 mm diameter and infinite length, as in Figure 1.

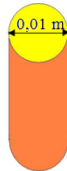


Figure 1 - Geometry of product

The FDM simulations based on the analysis of sequential points (also called nodes) taken from one of the diameters lying on a radial section of the product itself, as in Figure 2.



Figure 2 - Nodes of FDM analysis

In detail, the wall of the product is represented by the two external nodes, they are not at the interface product-air but at a distance of 0.5 mm inside the wall. The nodes of the FDM method are described by their position, identified with radial coordinates.

Table 1 - Coordinate of node

Node	Coordinates on diameter [mm]	Radial Coordinates [mm]
Wall1	0	5
Wall's node1	0,5	4,5
Node1	1	4
Node2	2	3
Node3	3	2
Node4	4	1
Product's core	5	0
Node6	6	1
Node7	7	2
Node8	8	3
Node9	9	4
Wall's node2	9,5	4,5
Wall2	10	5

Now, it's possible to create through MS Excel the spreadsheet formed with the ten nodes described in Table 1.

### 3.2. External conditions

For heat drying process, we will reproduce the effect of a hot air draught that laps against the product.

The data concerning the drying media to be included in the model are:

1. The flow velocity ( $v_{\infty}$ );
2. The temperature of air ( $T_{\infty}$ );
3. The humidity of air ( $\phi$ );

This three parameters associated with the maintenance time of the particular condition of flow in the chamber, are able to describe the drying cycle selected by the user.

### 3.3. The conditions of the product

The dough base presents obvious and disparate heterogeneity, from both physical and chemical point of view, compared with the internal structure of the solid matrix.

This heterogeneity can be seen along the three dimensions and during the time of the process.

For this study, in approximate way, the heterogeneity of product internal structure has not been considered.

The user must enter the initial data about pasta:

- The Temperature of product at initial time  $t_0$ , ( $T_p$ );
- The initial internal moisture of product ( $UT_{pasta}$ ), and the resulting water content of dough on dry basis ( $U_{pasta}$ ), derived by Eq.1.

$$U_{pasta} = \frac{UT_{pasta}}{(1 - UT_{pasta})} \quad (1)$$

In order to derive the product's internal temperature trend (the temperature trend in the nodes) we have divided the problem:

1. Study of thermal phenomena at the interface product-air;
2. Study of thermal phenomena inside the product.

### 3.4. Analysis of interface product-air thermal flow

#### 3.4.1. External convection

The first thermal flow which was analyzed is the forced external convection, due to the hot air flow and its interaction with the product.

This process was summarized by the Eq. 2;

$$q_{conv} = h \cdot S \cdot (T_{\infty} - T_p) \quad (2)$$

$h$  =: external heat exchange coefficient [ $J/m^2K$ ];

$S$  = contact surface [ $m^2$ ];

$q_{conv}$  = convection heat exchange [ $W$ ].

The variables contained in the formula were obtained through a fluid dynamic approach, using the Hilbert's formula to calculate the Nusselt's number and the external heat exchange coefficient (Migliori et al. 2004).

### 3.4.2. The external conduction

The wall's node exchange heat through conduction mechanism with the nearest internal node of product.

This process was summarized by Eq. 3:

$$q_{cond} = \frac{\lambda_p}{s_p} \cdot S \cdot (T_{N-1} - T_p) \quad (3)$$

$\lambda_p$  = thermal conductivity of dough [ $W/mK$ ]

$s_p$  = thickness of dough, the distance between the two nodes (called  $\Delta x$ ) [ $m$ ];

$S$  = surface of thermal exchange [ $m^2$ ];

$T_{N-1}$  = temperature of nearest internal node [ $K$ ];

$q_{cond}$  = conduction heat exchange [ $W$ ].

### 3.4.3. The evaporation

In this model, the water contained in the product at the beginning of the drying process (moisture content 30%) is partially "free moisture" and able to evaporate. In this time, the temperature of this water is the so called wet bulb temperature, and its pressure is equal to the saturation pressure.

During this period the evaporated water at the interface is continuously supplied by the internal mass flow and the heat power transferred from the air to the sample is fundamental in order to achieve the evaporation of water.

The heat flow that occurs from product to drying media is:

$$q_{evap} = h_m \cdot cl \cdot S \cdot (\rho_{v,T_{\infty}} - \rho_{v,T_{bb}}) \quad (4)$$

$h_m$  = mass transport coefficient [ $m/s$ ];

$cl$  = water latent heat of vapourisation [ $J/Kg$ ];

$S$  = surface of thermal exchange [ $m^2$ ];

$\rho_{v,T_{bb}}$  = vapor density at wet bulb temperature [ $kg/m^3$ ];

$\rho_{v,T_{\infty}}$  = vapor density at  $T_{\infty}$  [ $kg/m^3$ ];

Eq. 4 is valid until the moisture of product remains in saturation conditions.

In this phase, the process proceeds at a constant rate of water evaporation. During this period the water that evaporates at the interface product-air is continuously supplied by the internal moisture flow. In this time, the heat power transferred from the air to the dough equates the thermal power required for water evaporation.

At the end of this part of drying process, when the internal moisture content has decreased due to the drying process, the water concentration inside the product arises and the internal water flow decreases. In this second part of process (descending rate of drying) the thermal power required for water evaporation is not sufficient to balance the external heat transfer, and as a consequence of this, the interface temperature increases.

In this model, the mechanism is reproduced by changing the value of vapor density during the drying process. During the process constant rate, the vapor density at the interface equates the vapor density at wet bulb temperature. In the second phase the density decreases to the value of vapor density at the temperature of drying media ( $\rho_{v,T_{\infty}}$ ).

In this analysis it has been assumed a linear dependence between the internal humidity of product and the density of vapor:

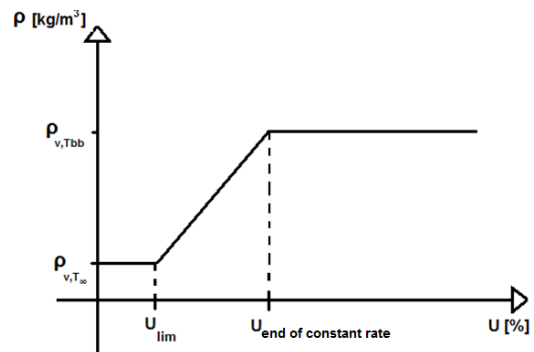


Figure 3 - vapor density at the interface

In this hypothesis, it has been researched the relation between the vapor density at the interface end the product internal moisture, explained by Eq. 5:

$$\rho_{v_i} = \rho_{v,T_{\infty}} + \frac{\rho_{v,T_{bb}} - \rho_{v,T_{\infty}}}{U_{end\ of\ constant\ rate} - U_{lim}} \cdot (U - U_{lim}) \quad (5)$$

$U_{end\ of\ constant\ rate}$  = internal moisture value corresponds with the end of the drying phase at constant rate of evaporation [%];

$U_{lim}$  = internal moisture value corresponds with the end of the moisture evaporation [%].

The user, through his knowledge about the dough base, determines the value of this specific moisture values.

When, at the wall's nodes, the moisture  $U_{lim}$  is reached, the external skin of product is completely dried for the model, and the thermal contribution of evaporation is nullified.

#### 3.4.4. Time dependence

During the time the heat flow that interests a particular node was described as follows:

$$\Delta q = \frac{\Delta x}{2} \cdot \rho_p \cdot C_p \cdot S \cdot \frac{[T_{t+1}^N - T_t^N]}{\Delta \tau} \quad (6)$$

$\Delta x$  ( $\Delta r$ ) = the distance node-interface [m];

$\rho_p$  = density of product [ $\text{kg/m}^3$ ];

$C_p$  = heat capacity of product [ $\text{J/kg} \cdot \text{K}$ ]

$S$  = surface of thermal exchange [ $\text{m}^2$ ];

$T_{t+1}^N$  = temperature of the N-node at time t+1 [K];

$T_t^N$  = temperature of the N-node at time t [K]

Eq. 6 must be equate with all the heat thermal flows related to the node:

$$\begin{aligned} \frac{\Delta r}{2} \cdot \rho_p \cdot C_p \cdot S \cdot \frac{[T_{t+1}^N - T_t^N]}{\Delta \tau} = & \left( \frac{T_t^{N-1} - T_t^N}{\frac{\Delta r}{\lambda_p}} \right) + \\ & + \left( \frac{T_\infty - T_t^N}{\frac{1}{h}} \right) + \left( h_m \cdot cl \cdot \left( \frac{P_{v,T_\infty}}{R \cdot T_\infty} - \frac{P_{v,T_{bb}}}{R \cdot T_{bb}} \right) \right) \end{aligned} \quad (7)$$

In the Eq.7,  $T_{t+1}^N$  is the unique unknown factor, thanks to recursive resolution process is possible to derive the trend of temperature along the time.

### 3.5. Analysis of internal thermal flow

#### 3.5.1. Internal conduction

In the internal nodes the heat flow is determined through conduction mechanism, similarly to external conduction. Differently, the equation must be referred to the cylindrical geometry, by using the cylindrical coordinates:

$$\rho_p \cdot C_p \cdot \frac{\partial T}{\partial \tau} = \lambda_p \cdot \frac{\partial^2 T}{\partial r^2} \quad (8)$$

In Eq. 8,  $r$  is the distance that divides the observed node and the center of the product (the core).

Also in this case, the study ends with the research of the temperature trend during the time.

### 3.6. Analysis of interface product-air mass flow

#### 3.6.1. The evaporation

The product loses his water content through the evaporation phenomena:

$$\dot{m} = h_m \cdot S \cdot (\rho_{v,T_\infty} - \rho_{v,T_{bb}}) \quad (9)$$

$\dot{m}$  = water mass transfer from product to drying media [ $\text{kg/s}$ ];

$h_m$  = external mass exchange coefficient [ $\text{m/s}$ ];

$S$  = surface of thermal exchange [ $\text{m}^2$ ];

$\rho_{v,T_\infty}$  = density of vapor at  $T_\infty$  [ $\text{kg/m}^3$ ];

$\rho_{v,T_{bb}}$  = density of vapor at wet bulb temperature, on the surface of product until saturation conditions remains [ $\text{kg/m}^3$ ].

Also for the Eq.9, the simplification concerning the density of vapor remains valid.

#### 3.6.2. The internal conduction

In order to completely study the phenomena related to the mass flow, the moisture diffusion between the wall's nodes and the nearest internal node of product must be analyzed:

$$\dot{m}_{cond} = D_{H_2O-pasta} \cdot S \cdot (U_t^N - U_t^{N+1}) \cdot \rho_p \quad (10)$$

$\dot{m}_{cond}$  = water mass transfer between two nodes [ $\text{kg/s}$ ];

$D_{H_2O-pasta}$  = water diffusivity in the dough [ $\text{m}^2/\text{s}$ ]

$S$  = surface of thermal exchange [ $\text{m}^2$ ];

$U_t^N$  = water content on dry basis of the N-node at time t  
 $\left[ \frac{\text{kg } H_2O}{\text{kg dry product}} \right]$

$U_t^{N+1}$  = water content on dry basis of the N-node at time t  
 $\left[ \frac{\text{kg } H_2O}{\text{kg dry product}} \right]$

$\rho_p$  = density of dough base [ $\text{kg/m}^3$ ]

In particular, an important factor for the Eq. 10, equation is the water diffusivity inside the product. This variable, depending on internal moisture content, is determined for water concentration higher than 22% w/w through experimental step function, and for water content less than 22% w/w through experimental continuous chemical function, not reported.



### 3.6.3. Time dependence

The mass flow which has been analyzed in wall's node it has been obtained through the analysis of conduction and evaporation, in order to research the internal moisture trend during time.

## 3.7. Analysis of internal mass flow

### 3.7.1. Internal conduction

As seen for the conduction of wall's node, the Fick's law guides the diffusion of moisture between two contiguous internal nodes in its massive terms.

Also in this analysis, the model researches the trend of internal moisture during the time of the process.

## 3.8. Implementation in MS Excel

With the previously equation, it has been possible to implement the predictive model through MS Excel.

In the spreadsheet there is a preliminary part where the user can set the simulation of the process. The user must insert:

1. The drying cycle data. In this part, the data of the different phases of the process are required, in particular the property of the drying air, as shown in Figure 4:

Insert the drying data:				
Phase	Minutes	Air Temperature (°C)	Air Moisture content [%]	Velocity [m/s]
1	0	36	2%	0,5
2	30	76	95%	0,5

Figure 4 - drying data

2. The dough data. The user through his knowledge must insert the data visualized in Figure 5:

Dough data		
Initial temperature [°C]		30 [°C]
Initial temperature [K]		303 [K]
External diameter		10 [mm]
Thermal conductivity		0,41 [W/(m*K)]
% of starch		0,1171 [%]
% gluten		0,2423 [%]
Initial moisture		0,3 [%]
Heat Capacity (Initial)		1179945,278 [J/kg*K]
Density (Initial)		1289,609433 [kg/m^3]

Figure 5 - data of dough

3. Moisture limits. In this part the user can set the values of moisture limit for the external evaporation, this values is summarized by Figure 6:

Moisture limit of constant rate fase	Total Moisture [% w/w]	Moisture on dry basis [%]
	19,00%	23,46%
Moisture limit	1,00%	1,01%

Figure 6 - values of moisture limit for the evaporation

4. The analysis data: in this part the user can set the temporal sequence of the calculation step. The value of  $\Delta\tau$  cannot be higher than  $\Delta\tau_{lim}$ , which is the value of  $\Delta\tau$  that makes unstable the model. This data are visualized in Figure 7:

Analysis Data	
$\Delta t$	10 [s]
$\Delta t_{lim}$	1772,41238 [s]

Figure 7 - Analysis data

## 3.9. Results

### 3.9.1. The comparison with industrial case

Through the implementation of the model in MS Excel it was possible to recreate a real industrial cycle.

In particular, the data inserted through the preliminary part reproduce a classic HT drying cycle for long pasta:

- First ventilation phase (30 minutes) where the product is ventilated by dry cold air (moisture content 2%, temperature 36°C);
- Second phase of drying, where spaghetti are dried by hot air (air moisture content 95%, temperature of 76°C).

This cycle, visualized in Figure 8, dries the product slowly, in order to preserve its organoleptic properties.

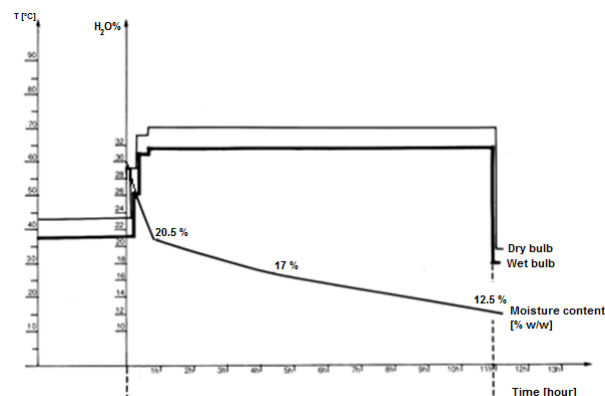


Figure 8 - internal moisture during HT drying cycle

In particular, one of the goals was to verify, through this model, the trend of internal moisture obtained by HT cycle.

The internal water content during the process, beginning from an internal value of 30%, reaches the value of 20,5% after 1 hour. This initial time of drying is due to the first process phase, that proceeds at a constant rate until the humidity of wall's node equates the value  $U_{lim}$ .

Later, the internal moisture reaches the value of 17% after 4-5 hours of process, in this period the most important phenomenon that interests the product is the internal mass flow, since the evaporation has entered in its descendent rate phase.

Finally, the product reaches the value of internal humidity of 12,5% (the maximum value of internal humidity is 12,5% for the Italian legislation) after 11-12 hours of drying process.

In the first instance, it was observed the temperature trend of the wall's node:

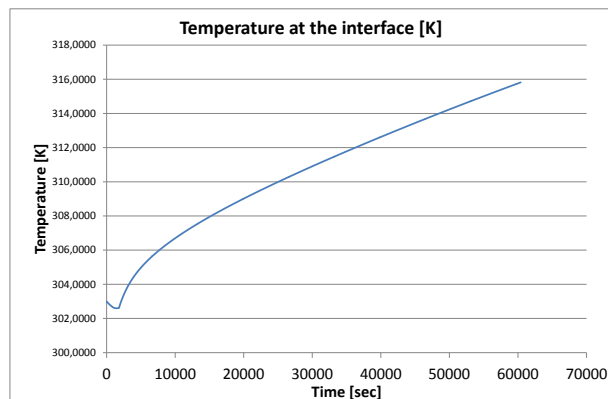


Figure 9 - temperature trend at the interface

From figure 9 it can be observed the two phases of the trend temperature at the interface product-air:

1. a first phase so-called “evaporative cooling” due to the contribute of the mechanism of evaporation, that impact on the temperature more than the conductive and convective terms;
2. in the second phase the temperature increases due to the convective heat exchange.

Secondly, it was analyzed the trend of internal moisture of nodes and the relative run of average internal humidity. This analysis is summarized by Figure 10:

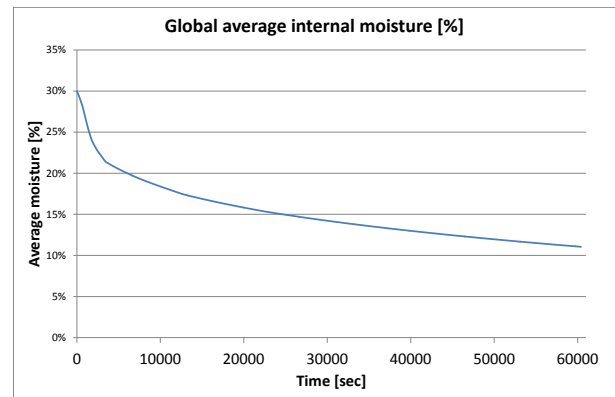
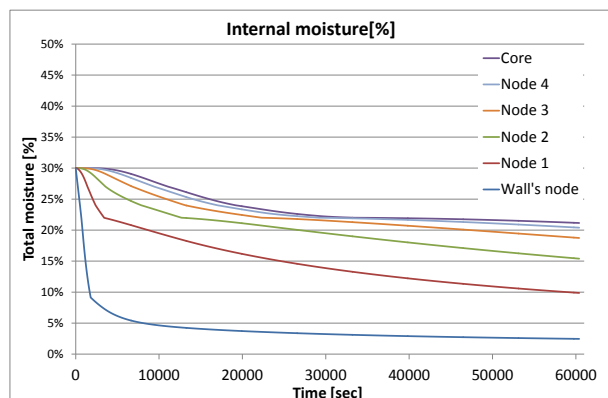


Figure 10- internal moisture trend

Thanks to the values of moisture previously identified, it has been possible to compare the results of the FDM simulation and the real industrial case. The results of the simulation are summarized in Figure 11:

Value	Time		
	Hours	Minutes	Seconds
0,205	0	51	50
0,170	3	35	0
0,125	12	30	10

Figure 11- characteristics value of cycle drying

The last phase of the process result slightly longer than the value identified for the industrial case.

Since the proposed values are highly dependent on the diffusivity coefficient, this variable was increased of 10%, obtaining the characteristic times shown in Figure 12.

Value	Time		
	Hours	Minutes	Seconds
0,205	0	52	10
0,170	3	28	20
0,125	11	44	40

Figure 12 - value of cycle drying with incremented diffusivity coefficient

Thanks to this control, the results of the model were verified, compared with the industrial case, taking into account the approximations used to create the model. In fact, with an incremented diffusivity coefficient, also the third part of process was reproduced with a good approximation.

#### 4. EXTERNAL SIMULATION

#### 4.1. The static dryer

In order to analyze a drying cycle of pasta through the simulation software TDyn, a simply layout of static dryer was selected for this study.

In this drying chamber, the product was introduced on perforated trays that were disposed along special supports.

This plants are often utilized in small and medium companies, thanks to its simplicity, its versatility and the low volume required.

The principal problem of this system is represented by the difficulties in obtaining the uniformity of the drying media inside the chamber.

This parameter is fundamental in order to achieve homogeneity of the process and therefore a better quality of product.

The CFD simulation was used to select the layout of static dryer of pasta capable of guaranteeing the best process uniformity.

The software TDyn, is able to predict with good approximation the evolution of product's temperature.

Thanks to this tools, the internal product's temperature was evaluated , through the analysis of the nodes that represent the pasta inserted in the room.

#### 4.2. Initial conditions

It was selected a particular static dryer provided with singular inlet and singular outlet.

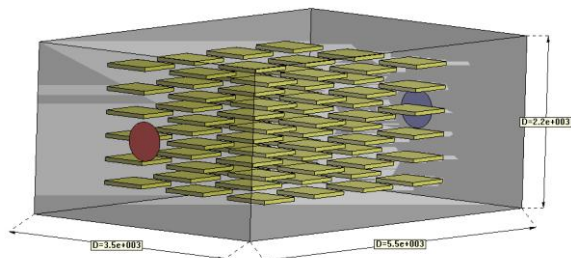


Figure 13 - static dryer

The design proposal in Figure 13 is one of the most used static dryer in small-sized company context.

The internal conditions of the drying chamber are settled through this parameter:

- The temperature of drying media: 353 K;
- The temperature of hot air: 293 K;
- Velocity of hot air at the inlet: 1 m/s;
- Surfaces of inlet and outlet: this parameters are settled to ensure a hot air flow of 1 m<sup>3</sup>/s;
- The geometry of drier: the geometry of drier is the same for all the simulation, it is a chamber 5500x3500x2200 mm;

- The disposition of product.

The simulation was programmed to proceeds to 5200 seconds, at the end of this period the temperatures of internal product's nodes were analyzed.

#### 4.3. Conclusions

In order to determine the quality of the drying process developed in the chamber, the statistical distribution of product's temperature was analyzed after 5200 seconds of process:

1. In order to evaluate the effectiveness of the process, the average temperature of product's nodes was analyzed. In fact, the biggest the average , the more drying process is effective,. Secondly, it was studied the value of median: the biggest the median, the further the drying stage is;
2. The efficiency of the process represents its homogeneity of product's drying. In order to evaluate this property, it was analyzed the value of standard deviation of product temperature . This parameter provides an indication of the distance between the maximum value of temperature and its average. In fact, the more the process dries homogeneously, the lower the standard deviation is.

	Singular inlet and singular outlet
Median	293,008
Average	293,1950103
Standard Deviation	1,275179069
Maximum	349,183

Figure 14- parameters of analysis

In addition, it was analyzed the statistical distribution of temperature of product after 5200 seconds, in order to appreciate graphically the evaluation of layout.

Initially, it was proposed the global distribution of temperature. Secondly, it was visualized the initial part of distribution ("the head") in order to appreciate the median value. Finally, it was showed the last part of distribution ("the tail"), in order to visualize the homogeneity of the process, represented by shorter tails of distribution.

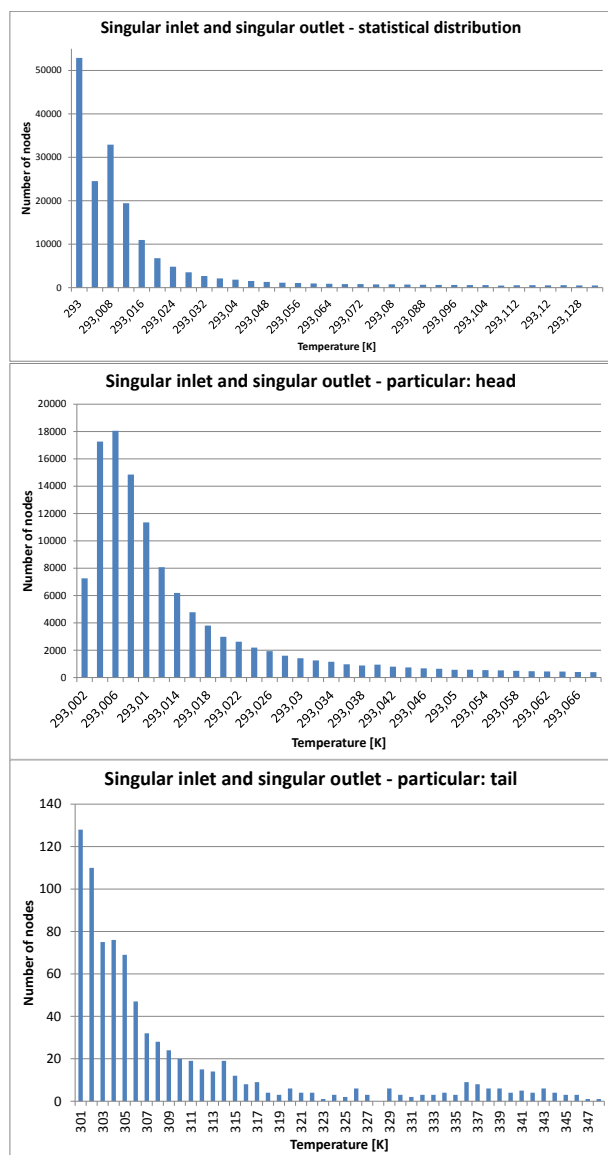


Figure 15 - statistical distribution of temperature

Thanks to parameters visualized in Figure 14 and the graphical analysis showed in Figure 15, it could be possible to study the layout proposed.

This configuration layout, the most utilized actually to implement the modern process, has reported poor performance regarding the homogeneity of drying: the value of median is low and the standard deviation is high, and the distribution of temperature emphasizes a long tail.

This conclusion is further verified through the visualization of internal flow of drying media, which, thanks to this layout, is able to reach principally the product placed near the inlet, as shown in Figure 16.

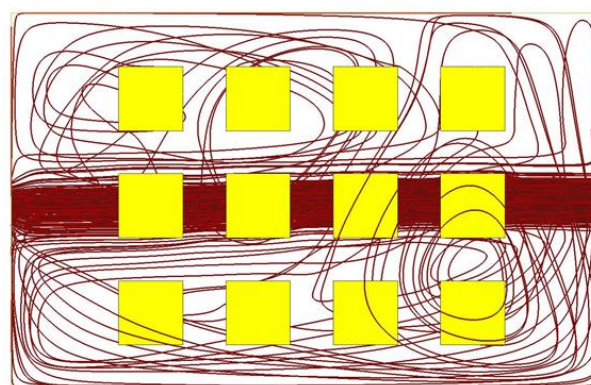


Figure 16 - internal flows

The tool developed can be used to compare the performances of different layouts of drying chambers in order to determine which one ensures the best drying treatment, in terms of homogeneity and effectiveness of the process. This improvement in the dryer design results in a better quality of the product and in a more efficient use of energy resource.

## REFERENCES

- Singh, R.P., Heldman, D.R., 2001. *Introduction of Food Engineering*. Third Edition. London: Academic Press
- Milatovich, L. and Mondelli, G., 1990. *La tecnologia della pasta alimentare*. Pinerolo: Chirioti Editore.
- Migliori, M., De Cindio, B., Gabriele, D., Pollini, C.M., 2004. Science Direct. *Modelling of high quality pasta drying: mathematical model and validation*. 18 October
- Migliori, M., De Cindio, B., Gabriele, D., Pollini, C.M., 2004. Science Direct. *Modelling of high quality pasta drying: quality indices and industrial application*. 22 December
- Waananen, K.M., Okos, M.R., 1994. *Journal of Food Engineering*. *Effect of Porosity on Moisture Diffusion during Drying of Pasta*, 5 December
- Rizzo, R., Romagnoli, G., Vignali, G., (2009). *Food Manufacturing Efficiency Process parameter optimization for the design of a fresh filled pasta pasteurizer*, Vol. 3, N. 1, pp. 25-33
- Yue, P., Rayas-Duarte, P., Elias, E., 1999. *Cereal Chemistry*. *Effect of drying temperature on physicochemical properties of starch isolated from pasta*. 76.

# BAKERY PRODUCTION SCHEDULING OPTIMIZATION – A PSO APPROACH

Hecker F. <sup>(a)</sup>, Hussein W. B. <sup>(b)</sup>, Hussein M. A. <sup>(c)</sup>, Becker T. <sup>(d)</sup>

<sup>(a)</sup>Department for Process Analysis and Cereal Technology  
(FG Prozessanalytik und Getreidetechnologie)  
University of Hohenheim, Garbenstr. 23, 70599 Stuttgart, Germany

<sup>(b), (c), (d)</sup>Department of Life Science Engineering  
(Ingenieurwissenschaften für Lebensmittel und biogene Rohstoffe  
Arbeitsgruppe (Bio-)Prozesstechnik und Prozessanalyse)  
TECHNISCHE UNIVERSITÄT MÜNCHEN,  
Weihenstephaner Steig 20, 85354 Freising, Germany

<sup>(a)</sup>[heckerfl@uni-hohenheim.de](mailto:heckerfl@uni-hohenheim.de), <sup>(b)</sup>[whussein@wzw.tum.de](mailto:whussein@wzw.tum.de), <sup>(c)</sup>[hussein@wzw.tum.de](mailto:hussein@wzw.tum.de), <sup>(d)</sup>[tb@wzw.tum.de](mailto:tb@wzw.tum.de)

## ABSTRACT

Production of bakery goods is strictly time sensitive due to the yeast proofing of doughs. That causes special requirements for production planning and scheduling, which is in bakeries often completely based on the practical experience of the responsible employee instead of mathematical methods. This often leads to sub-optimal performance of companies due to the sometimes “chaotic” scheduling approach. This work presents an approach to use PSO as a highly efficient numerical method to solve complex scheduling problems. In all probability bakeries will be able to increase efficiency by optimizing their production planning with numerical methods like the one presented here. After modelling the production line with a limited range of input data and a pre-processing of the necessary data to match the real process induced requirements, a PSO algorithm handles the optimization task. Results show the high potential of this method to solve the scheduling problem in good computational time.

Keywords: production planning, baking industry, scheduling, particle swarm optimization

## 1. INTRODUCTION

A high number of bakery goods contain yeast (*Saccharomyces cerevisiae*) as a proofing agent. Due to the fact that this form of proofing is a fermentation performed by microorganisms in which sugars are metabolised to CO<sub>2</sub> (among other components), the production of such goods is not highly but strictly time sensitive from the point on where the microorganisms get in contact with water and substrates, as happens in the dough making process. Since it is costly to regulate or slow down the fermentation speed of yeast by

cooling and its sometimes negative influence on the product quality on the one hand and the decrease of product quality (up to the total loss of marketability) due to a too long fermentation process on the other hand, the time dependency of the processing has always to be considered in the production scheduling.

Focusing on the German baking industry the production planning is almost completely based on the practical experience of the responsible employee(s) instead of the usage of mathematical methods like scheduling or optimization theory. Combined with the high diversity of the product range that includes around 100 different products in a common German bakery and the high complexity of the scheduling task induced therein the performance of bakeries is often sub-optimal. The use of numerical methods like particle swarm optimization (PSO) to solve the scheduling task might increase the efficiency of companies by calculating an optimal production plan and by that determine the exact time schedule and capacities of devices needed to reach the production goal. Thus idle times of machines can be reduced or completely erased which leads to a reduction of energy consumption.

The German baking industry consists of approximately 16000 companies that produce about 6.25 billion tons of baked goods with a business volume of almost 12 billion Euros per year and employs over 275000 employees. Thus the increase of companies' efficiency in respect of e. g. energy consumption or staff allocation and working hours comprises a high potential to decrease production costs in this industry.

After its invention by Kennedy and Eberhart (Kennedy and Eberhart 1995) PSO was widely used to tackle numerous scheduling or optimization problems in many different industry branches (Eberhart and Shi 2001; Lian, Gu and Jiao 2008; Liu, Wang, Liu, Qian



and Jin 2010; Pan, Tasgetiren, and Liang 2008; Tasgetiren, Liang, Sevkli and Gencyilmaz 2007; Wang and Yang 2010). As a swarm intelligence algorithm it mimics the behaviour of animal swarms like fish schools or bird flocks and iteratively searches the search space for the optimal solution. Since PSO provides easy implementation, easy modification and the ability to solve high complex scheduling or optimization tasks in short computational time it is an appropriate method to solve the bakery scheduling problem.

## 2. MATERIAL AND METHODS

The developments and investigations presented here were performed on a “lenovo ThinkPad R500” with an “Intel Core 2 Duo” 2.26 GHz processor, 2 GB RAM and Microsoft XP 2002 as system software. The modelling and optimization were made with MATLAB 7.1 (The MathWorks, Inc) and the visualization and validation of the optimization results with ARENA 11.0 (Rockwell Automation, Inc).

## 3. MODELLING AND OPTIMIZATION

Due to the high diversity of products in a German bakery it is not possible to solve the scheduling task with exact methods, at least not in reasonable computational time. Using an exact method to calculate the parameters of all possible schedules would mean to calculate an enormous number of combinations given by the relation in equation (1), where  $n$  is the number of jobs (products) and  $m$  is the number of machines used.

$$\text{number of schedules} = (n!)^m \quad (1)$$

It is obvious that this relation causes an almost unmanageable amount of combinations for the scheduling problem in German bakeries, where the normal range of products is commonly above 100 and the machinery in operation between 10 and 50 (depending on the bakery’s size).

From the scheduling point of view the production in a bakery can be described as a *hybrid flow-shop* according to the common definitions, e. g. in (Pinedo 2008; Ruiz and Vázquez-Rodríguez 2010).

The number of possible schedules can be reduced significantly by considering the scheduling task in a bakery as a *permutation flow-shop* instead of a ‘normal’ *hybrid flow-shop* by adding the constraint that the order in which the jobs  $n$  pass through the production is fixed and does not change between production stages (Pinedo 2008; Lian, Gu and Jiao 2008; Tasgetiren, Liang, Sevkli and Gencyilmaz 2007). Although the real process in a bakery does not fulfil this requirements entirely this model can be used and modified to match with the real production processes, where products can bypass other previously started products. By doing so the number of possible combinations is reduced from  $(n!)^m$  to  $n!$  and each schedule is a permutation of  $n$  (Perez-Gonzalez and Framinan 2010). Each of those permutations is used then to determine a sequence of products on the first production stage of the bakery that is crucial for sequencing the work flow. Still  $n!$  different schedules

may easily lead to optimization problems unsolvable with exact methods in reasonable computational time for high numbers of  $n$ .

Additionally the scheduling in a bakery is subject to a no-wait constraint due to the fact that there are very small tolerances for waiting times of yeast containing doughs/products.

### 3.1. Modelling of the production processes

The modelling of the production site and the compliance of the no-wait constraint are done prior to the actual optimization of the scheduling.

As first step a matrix  $A$  is formed that contains all processing times ( $PT$ ) of the products to be produced. The rows and columns of the matrix represent the products and the production stages respectively, such that e. g.  $a_{2,3}$  would return the processing time of product 2 on stage 3, meaning in this case that the product “Dinkelbrot” requires a dough rest of 30 minutes. The processing times are determined by the recipe and the desired characteristics of the finished product. Basically all products follow the same way through the production on consecutive stages, meaning that a product does not return to an already passed stage. The common progression of production stages in a bakery is shown in Figure 5. Some products do not require processing on a certain stage (e. g. if a product needs no dough rest) and skip it. A zero entry in the matrix  $A$  indicates that the product skips the respective stage and is not processed there.

The information contained in such a matrix is given as an example in Figure 1 (the production line data used for one of the examples presented here).

Product/Process	Ingredients Preparation	Kneading	Dough Rest	Dividing and Forming	Proofing	Baking
Korn Eck	10	9	25	11	30	50
Dinkelbrot	8	9	30	16	45	50
Körnerbrot	10	9	20	20	50	60
Annabrot	5	9	20	16	50	63
Holzucken	5	9	15	21	70	55
Mischbrot 500g	5	8	15	31	60	53
Mischbrot 1kg	5	8	15	26	60	55
Bauernbrot 500g	5	9	20	31	60	55
Bauernbrot 1kg	5	9	20	26	60	60
Fitnessbrot	5	9	20	31	60	50
Kartoffelbrot	8	15	20	25	40	45
Doppelbrot	5	9	20	31	60	40

Figure 1: Processing Times of Products (Line 1)

Based on this matrix the starting times of all products on all stages are calculated for the investigated sequence of products (which represents a particle in the PSO) and form a new matrix  $B$ .

To make sure that no product waiting times appear during the optimization process (and the no-wait constraint is not violated) the calculation of the products’ starting times on the respective stages uses the following steps:

1. The starting time ( $ST$ ) of product 1 (first product of the sequence) on the first stage is “0” as this represents the start of the production.



2. The starting times of product 1 on the following stages  $m$  are simply a summation of the processing times ( $PT$ ) on the previous stages as given in equation (2), where  $j$  is job (or product) 1 and  $m$  is the current stage calculated.

$$ST_{j,m} = \sum_{n=1}^{m-1} PT_{j,n} \quad (2)$$

3. From this first row in matrix  $B$  the starting times for the next product in the sequence are calculated by first picking the adequate starting time on the last stage  $m_{max}$ . The “adequate” time means the  $ST$  that makes sure that no waiting-time will occur and is determined by following equation (3), where  $j = 2, 3, \dots, j_{max}$  and  $m = 1, 2, \dots, m_{max}$ .

$$ST_{j,m_{max}} = \max(ST_{j-1,m} + PT_{j-1,m} + \sum_{n=m}^{m_{max}-1} PT_{j,n}) \quad (3)$$

Since there are often stages in a bakery production that have a practically unlimited capacity (subject to the condition of sufficient dimensioning), like proofing chambers or the dough resting that is often performed by just put the dough aside for a certain time span, only stages with limited capacity are considered for choosing  $ST_{j,m_{max}}$ .

Thus it is made sure that a possible “bottle-neck” in the production line will have the deciding impact on the calculation of the  $ST$ .

4. After  $ST_{j,m_{max}}$  is determined the other starting times for product  $j$  on stages  $m$  are calculated by just subtracting the respective  $PT_{j,m}$  to form the matrix  $B$ . Thus no waiting times for products will appear in the schedule.
5. To assure an optimal scheduling within each investigated product sequence and to allow the skipping of stages or the bypassing of other products each row in  $B$  (which represents a product) is compared to a set of conditions during its calculation process to create the best possible schedule setup.

The progress of the modelling algorithm is shown in the flow chart in Figure 2.

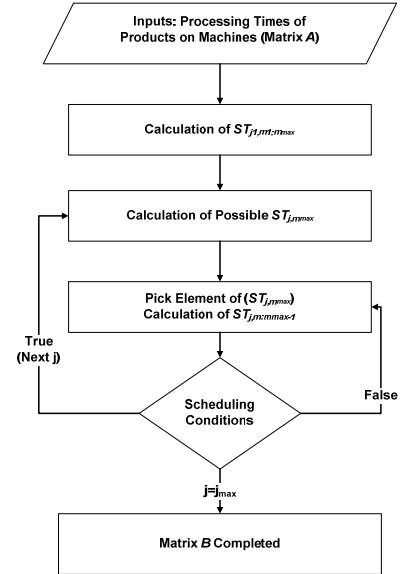


Figure 2: Flow Chart of Modelling Algorithm

### 3.2. Particle swarm optimization

The PSO method is capable of solving scheduling problems with high complexity and easy to implement due to the relative low input data needed. It follows the basic algorithm shown for an example in the flow chart in Figure 3.

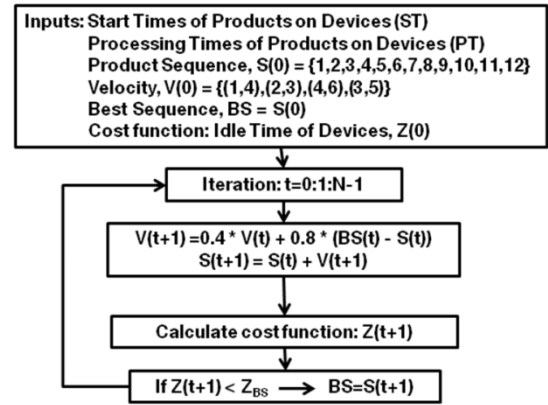


Figure 3: Basic PSO Algorithm

The cost or objective functions to be evaluated could be e. g. the idle time of machines (as presented in the example), the minimum total idle time of machines, utilization of machines, the makespan or other objectives of economical interest.

During the iterations of the algorithm the particles are “flying” through the search space and due to the frequent update and comparison of the best sequence so far and each particle’s current value of the cost function, move to the optimal solution of the given optimization problem. Figure 4 shows this behaviour of the algorithm by illustrating the “flight” of a particle towards the optimal solution (in this case the minimum of a cost function).

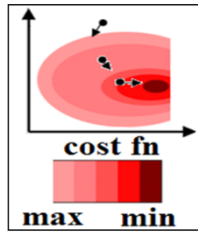


Figure 4: Particle “Flight” through the Search Space

### 3.3. Visualisation and validation of PSO output

The simulation software ARENA 11.0 (Rockwell Enterprises) was used as a visualisation tool for the production line investigated. A production line model was built out of predefined modules (setup shown in Figure 5) in the software and the relevant production sequences were simulated.

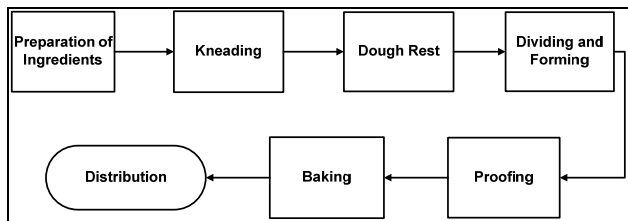


Figure 5: Production Line Model

Thus it was possible to further investigate the flow of each product through the production line and to validate the performance of the numerical generated best production sequence.

### 3.4. Example optimizations

Two bakery production lines were modelled and optimized as example applications. The first line, taken from a medium sized baking company that runs 53 chain stores, produces 12 different kinds of bread using six different stages/machines. The second line is taken from a small bakery workshop with just two chain stores that produces 10 different bread products on a comparable machinery setup (but with other dimensioning). The 12 products in example one and the 10 products in example two give a total of 479.001.600 ( $n!=12!$ ) and 3.628.800 ( $n!=10!$ ) possible different sequences respectively if the production line is handled as a *permutation flow-shop*.

Thus it would be a time consuming procedure to solve these setups by calculating each possible schedule to find the optimal one. So the aforementioned PSO method proposed here was used to find the best schedule.

The processing times of the products and the used machines of example one are given in Figure 1 and for example two in Figure 6. The general line setup for both is shown as a model in Figure 5. Since the time sensitivity of dough or bakery products ends with the baking process the investigations of the production lines end with this process step and the picking of products for distribution was not added to this analysis.

ARENA as a simulation software provides the possibility to validate the optimization output by defining the ARENA model parameters according to the optimization results. Also it is possible to plot the utilization of each process module and thus can be used to visualize e. g. utilization gaps or the different utilization of machines according to the investigated product sequence.

Product/Process	Ingredients Preparation	Kneading	Dough Rest	Dividing and Forming	Proofing	Baking
Baguette	10	5	20	15	45	30
Dinkelbrot	5	7	15	20	50	60
Mischbrot	7	10	30	30	60	55
Roggenbrot	10	8	10	25	45	60
Kosakenbrot	8	9	25	15	55	60
Weißbrot	5	15	45	30	60	35
Körnerbrot	8	10	20	30	50	45
Fladenbrot	7	6	15	35	60	30
Genetztes Brot	9	7	35	20	45	55
Bauernbrot	5	5	10	16	50	45

Figure 6: Processing Times of Products (Line 2)

### 3.5. Optimization results

For both example lines the minimization of the total idle time was the objective function. The idle time of a machine means the time a machine that is free and able to process a product has to wait for a product that has not yet finished its processing on a previous machine. In many cases this idle time means a waste of energy because the machine stays in a stand-by mode. In the case of bakery devices the idle time of an oven is the most important factor regarding potential energy savings because an oven has to be heated to remain in stand-by. But besides the oven stage every idle time of other devices means a waste of energy.

The capacity limited stages on both lines are stages one (preparation of ingredients), two (kneading), four (dividing and forming) and six (baking). All those stages can only handle one product batch at a time. The two other stages in the production line (dough rest and proofing) are practically unlimited for the investigated scenario due to their ability to handle several product batches at the same time.

As initial sequences on both investigated production line examples the real process sequences were taken as used in the two companies (shown in Figure 1 and 6 respectively). First the total idle times for both lines were calculated as reference values. After that the optimization procedure was used to find the sequence holding the minimal total idle time, which is the optimal sequence in the analyzed case. The results for both examples are shown in Table 1.

Table 1: Results of Optimization for both Production Line Examples

Production Line Example	Total Idle Time Initial [min]	Total Idle Time Optimal [min]	Shift Length Initial [min]	Shift Length Optimal [min]	Reduction of Total Idle Time [min]	Reduction of Shift Length [min]
1	1611	1519	727	722	92	5
2	1175	1063	570	570	112	0

The mean computational time (500 iterations) for the optimization of the two investigated production lines was 1.065 s and 0.878 s respectively.

Although both analyzed production lines are used in this (initial) setup for a long time and have repeatedly been enhanced empirically, there was an improvement possible coming from the optimization approach.

On production line one the total idle time could be reduced by 92 minutes (or 5.7%) by changing the product sequence. Furthermore the total runtime of the line could be reduced by 5 minutes. Focusing on the oven's utilization, the gap present in the former utilization could be closed (see Figure 7).

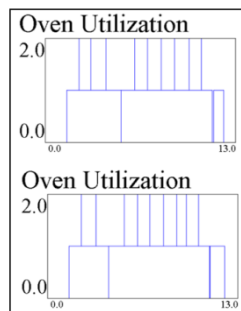


Figure 7: ARENA Plot of Oven Utilization on Production Line One before (top) and after Optimization (below)

The utilization of the other capacity limited stages is illustrated in Figure 8. Although the plots seem only slightly different the changes in the sequence result in the aforementioned reduction of the total idle time.

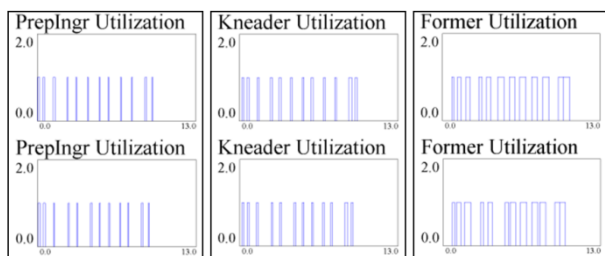


Figure 8: ARENA Plot of Utilization of Stages One, Two and Four on Production Line One before (top) and after Optimization (below)

On production line two the total idle time of the used machines could also be reduced significantly by 112 minutes (or 9.5%), although the total runtime of the production line did not change. Since this line was scheduled with special respect to the oven utilization there were no gaps in its utilization by running the initial product sequence. This status was preserved after optimization (see Figure 9).

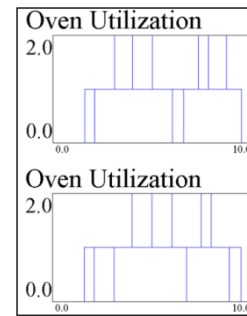


Figure 9: ARENA Plot of Oven Utilization on Production Line Two before (top) and after Optimization (below)

On this line the changes in machine utilization mainly appear in the previous stages (as shown in Figure 10).

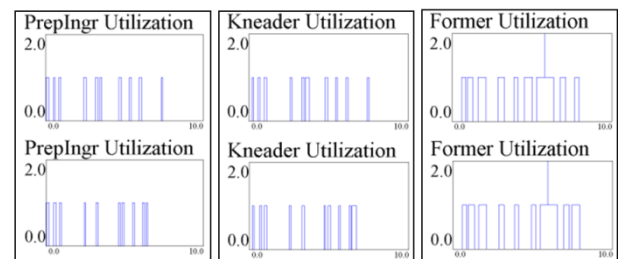


Figure 10: ARENA Plot of Utilization of Stages One, Two and Four on Production Line Two before (top) and after Optimization (below)

#### 4. CONCLUSION

The results obtained in the optimization of the two presented bakery production lines show the high potential that the use of numerical methods in general and of PSO in special has to improve the efficiency of baking companies. Especially the optimization of these two production lines that are used for several years in the presented setup and thought of as already optimal underlines the advantages process analysis with numerical methods can provide.

This potential is even higher if the presented PSO method is used as a decision support in the planning of new setups because it provides a fast and reliable possibility to find optimal schedules.

Since the implementation of PSO is easy and the modelling of production environments can be modified to match special product requirements, this method is suitable not only for bakery production but also for other time sensitive products as can often be found in other food industry branches.

#### REFERENCES

- Eberhart, RC, Shi, Y, 2001. Particle Swarm Optimization: Developments, Applications and Resources. *Proceedings of the IEEE Conference on Evolutionary Computation*, 81-86.
- Kennedy, J, Eberhart, RC, 1995. Particle swarm optimization. *IEEE International Conference on Neural Networks-Conference Proceedings*, 1942-1948.

- Lian, Z, Gu, X, Jiao, B, 2008. A novel particle swarm optimization algorithm for permutation flow-shop scheduling to minimize makespan. *Chaos, Solitons and Fractals* 35: 851-861.
- Liu, B, Wang, L, Liu, Y, Qian, B, Jin, YH, 2010. An effective hybrid particle swarm optimization for batch scheduling of polypropylene processes. *Computers and Chemical Engineering* 34: 518-528.
- Pan, QK, Tasgetiren, MF, Liang, YC, 2008. A discrete particle swarm optimization algorithm for the no-wait flowshop scheduling problem. *Computers & Operations Research* 35: 2807-2839.
- Perez-Gonzalez, P, Framinan, JM, 2010. Setting a common due data in a constrained flowshop: A variable neighbourhood search approach. *Computers & Operations Research* 37:1740-1748.
- Pinedo, M L, 2008. *Scheduling - Theory, Algorithms, and Systems*. 3<sup>rd</sup> edition. New York, NY, USA: Springer.
- Ruiz, R, Vazquez-Rodriguez JA, 2010. The hybrid flow shop scheduling problem. *European Journal of Operational Research* 205: 1-18.
- Tasgetiren, MF, Liang, Y, Sevkli, M, Gencyilmaz, G, 2007. A particle swarm optimization algorithm for makespan and total flowtime minimization in the permutation flowshop sequencing problem. *European Journal of Operational Research* 177: 1930-1947.
- Wang, Y, Yang, Y, 2010. Particle swarm with equilibrium strategy of selection for multi-objective optimization. *European Journal of Operational Research* 200: 187-197.

#### AUTHORS BIOGRAPHIES

**Florian Hecker** born 1979, finished his apprenticeship as a baker in 2002 and afterwards studied Food Technology at the University of Hohenheim (Stuttgart), where he received his Engineer's Degree in 2008. In May 2008 he started working as a researcher and PhD student in the Department of Process Analysis and Cereal Technology, University of Hohenheim, Germany. His research interests are in production optimization, scheduling theory and image processing based quality evaluation systems.

**Walid B. Hussein** born 1978, received his B.Sc. and M.Sc. with Honour Grade in Aerospace Engineering from Cairo University, Egypt in 2000 and 2005, respectively. During this period, he worked as a research assistant and teacher in the Egyptian Civil Aviation Academy & German University in Cairo. His research interests are in the area of signal and image processing, chemometrics and production optimization, with particular consideration of bioacoustics and pattern recognition. Between 2007 and 2009, He worked as a

researcher and PhD student in the Department of Life Science, University of Hohenheim, Germany. Afterwards, he moved as a PhD student to the Centre of Life Science Engineering, Technische Universität München, Germany.

**Dr.-Ing. Mohamed A. Hussein** born 1977, received his B.Sc. and M.Sc. with Honour Grade from Aerospace Engineering, Cairo University, Egypt in 2000 and 2003, respectively. From 2000-2006 he worked as a research assistant and teacher in the Aerospace Department, Cairo University & German University in Cairo. From 2006 to 2008 he worked as a Scientific employee and PhD Student in the University of Hohenheim-Stuttgart, primarily focused on Numerical Modelling, Finite methods, Numerics in Signal and Image Processing. Since 2008 he became the Group Leader of Process Analysis at University of Hohenheim. Moved in 2009 to Technische Universität München, where he received his PhD in 2010 on the thesis entitled "On the Theoretical and Numerical Development of lattice Boltzmann Models for Biotechnology and its Applications" and held the group leader of Bio-Process Analysis Technology & Process Analysis (PAT).

**Prof.-Dr. Thomas Becker** born 1965, studied Technology and Biotechnology of Food at the Technische Universität München. He received his PhD in 1995: "Development of a computerized enzyme integrated flow injection system and its use in the biotechnological process and quality control." From 1996 to 2004 he worked as Deputy Head of Department and assistant lecturer at the Institute of Fluid Mechanics and Process Automation of Technische Universität München. His habilitation took place in 2002: "Management of Bioprocesses by means of modeling and cognitive tools." In the period from 2004 to 2009 he worked as a university professor at the University of Hohenheim (Stuttgart) - until 2005 as head of the Department Process Analysis and Grain Technology, following as head of the Department of Process Analysis and Cereal Technology. In 2009, he became the professor (Ordinarius) of Beverage Technology Department at Technische Universität München.

# SIMULATION BASED DESIGN OPTIMIZATION FRAMEWORK FOR A GAIT PATTERN GENERATION OF A SMALL BIPED ROBOT WITH TIPTOE MECHANISM.

Krissana Nerakae<sup>(a)</sup>, Hiroshi Hasegawa<sup>(b)</sup>

<sup>(a)</sup>Functional Control Systems – Graduate School of Engineering and Science  
Shibaura Institute of Technology, Japan

<sup>(b)</sup>Department of Systems Engineering and Science  
Graduate School of Engineering and Science  
Shibaura Institute of Technology, Japan

<sup>(a)</sup>[m710505@shibaura-it.ac.jp](mailto:m710505@shibaura-it.ac.jp), <sup>(b)</sup>[h-hase@shibaura-it.ac.jp](mailto:h-hase@shibaura-it.ac.jp)

## ABSTRACT

The research of biped robot is still far from proposing a solution which generates a level of flexibility and reliability gait pattern that would enable practical walking on the variety of rough ground human's negotiation with ease on a regular basis.

To solve its issue, we learn from the mammal, belong to the plantigrade, on the natural world. The plantigrade, i.e., human, monkey or bear walk by adapting to a ground from a toe up to heel. From this adaptive walk, its flexibility and stability are excellent although this gait is not suitable for a quick walk.

In this study, to get flexibility and reliability of gait pattern, we introduce a tiptoe mechanism to a small biped robot through inspiration from its adaptive walk, and develop the Simulation Based Design Optimization framework (SBDO).

In this paper, the SBDO framework by approximated optimization process using Adaptive Plan system with Genetic Algorithm (APGA) is proposed. Moreover, to validate an ability of this framework, a small biped robot with tiptoe mechanism is simulated on flat ground at first trial. We discuss this validation result for estimating the SBDO framework.

**Keywords:** Simulation Based Design Optimization (SBDO), Tiptoe Mechanism, Adaptive Plan system with Genetic Algorithm (APGA).

## 1. INTRODUCTION

The research of biped robot is still far from proposing a solution which generates a level of flexibility and reliability gait pattern that would enable practical walking on the variety of rough ground human's negotiate with ease on a regular basis.

To solve its issue, we learn from the mammal, belong to the plantigrade, on the natural world. The plantigrade, i.e., human, monkey or bear walk by adapting to a ground from a toe up to heel. From this adaptive walk, its flexibility and stability are excellent although this gait is not suitable for a quick walk, to get

flexibility and reliability of gait pattern, we introduce a tiptoe mechanism to a small biped robot through inspiration from its adaptive walk.

Until recently, the studies of gait analysis for walking biped robot are incessant. Zhe Tang et al. have proposed a optimization for humanoid walking based on Genetic Algorithm (GA) base optimization for humanoid walking (Zhe et al. 2006). Lingyun Hu et al. have presented bipeds gait optimization using spline function based on probability model (Lingyun et al. 2006). These studies are about gait optimization of biped robot. Furthermore, a natural human walking was proposed in tiptoe mechanism for biped robot. Y. Xiang et al. have presented optimization based dynamic human walking prediction (Xiang et al. 2007) which were studied an optimization-based approach for simulation the motion of a digital human model. A model has 55 degrees of freedom which included tiptoe joints. Nandha Handharu et al. have proposed gait pattern generation with knee stretch motion for biped robot using toe and heel joints (Nandha et al. 2008). Cheol Ki Ahn et al. have proposed development of a biped robot with toes to improve gait pattern (Cheol et al. 2003), the gait pattern of the robot with toes was compared to the robot without toes by 3D graphical simulation. Shuuji Kajita et al. have proposed zero of moment point (ZPM) based running pattern generation for a biped robot equipped with toe spring (Shuuji et al. 2007). Abovementioned research was based on flat plate.

In addition, there are some studies which related to framework for biped robot locomotion. S. Ali A. Moosavian et al. have proposed the introduction of a cartesian approach for gate planning and control of biped robots and implementation on various slopes (S. Ali A. Moosavian et al. 2007). Naoya Ito and Hasegawa Hiroshi have presented a robust optimization uncertain factors of environment for simple gait of biped robot (Naoya and Hasegawa 2007), to optimize the gait for biped robot by using Simulated Annealing (SA). The robust optimization considered random values as floor



of fiction and restitution. Yu Zheng et al. have proposed a walking pattern generator for biped robots on uneven terrains (Yu et al. 2010), these approach were more general and applicable to uneven trains as compared with prior research methods based on the ZMP criterion. Above mentioned were used without toes mechanism model.

On the other hand, The Simulation Based Design (SBD) is powerful technology for improving the competitive power of products. Especially optimal design is an important technology for SBD. Therefore, authors have proposed Simulation Based Design Optimization (SBDO) in automotive design area (Hasegawa et al. 2007).

In this paper, to design and develop a tiptoe mechanism of a small biped robot, the SBDO framework by approximating optimization process using Adaptive Plan system with Genetic Algorithm (APGA) is proposed. Moreover, to validate an ability of this framework, a small biped robot with tiptoe mechanism is simulated on flat ground at first trial.

## 2. METHODOLOGY FOR GAIT OPTMIZATION

### 2.1. Simulation Model

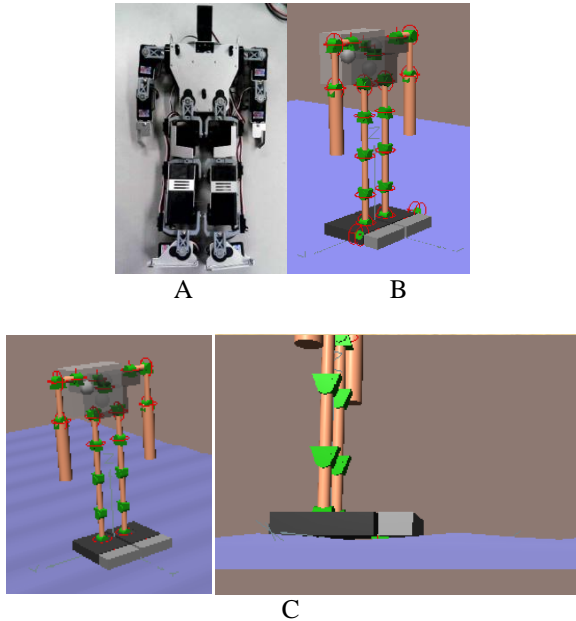


Figure 1: Model A Is KHR-2HV Robot  
Model B is simulation model on flat floor  
Model C is simulation model on rough floor

The optimization in this paper simulated on flat plate as friction constant by basing on KHR-2HV model. The robot is show in Figure. 1. It has 10 RC-Servo motor under the hip. The simulation uses same degree of freedom on these joints.

Several researches of toe joint utilization in bipedal locomotion have been proposed such as, 1) passive tiptoe joints in order to achieve stable foot lifting, 2)

tiptoe joints that are both active and passive control both actively and passively for less energy consumption walking and 3) active tiptoe joints for stepping up stairs. In this study, to get flexibility and reliability of gait pattern, we introduce a passive tiptoe mechanism to a small biped robot is shows in Figure. 2.

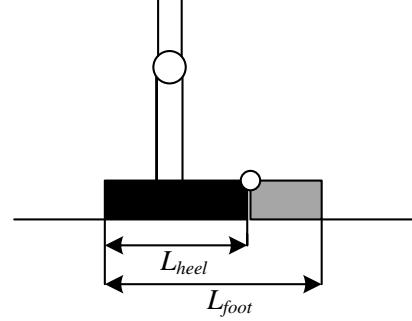


Figure 2: Model Of Robot Foot

The lengths are shown in Figure. 2,  $L_{foot}$  is length of foot defined to  $92[mm]$  and  $L_{heel}$  is length of heel define to  $75[mm]$ . The foot-heel length ratio obtained from Eq. 1 is 1.227, the range of ratio between from 1.196 to 1.426 (Chockalingam and Ashford 2007).

$$L_{foot-heel} = \frac{L_{foot}}{L_{heel}} \quad (1)$$

### 2.2. Simulation Based Design (SBD)

#### 2.2.1. Definition of the gait function

This paper is assumed the robot walks based on the gait function. Therefore, the function is defined based on a human gait pattern that focused attention on the walk cycle. To express this periodic cycle, the function to generate the gait is defined as follows:

$$\theta_i(t) = a_i + b_i \cos(\omega t) + c_i \sin(\omega t) + d_i \cos(2\omega t) + e_i \sin(2\omega t) \quad (2)$$

Where  $t$  is time,  $\omega$  is angular velocity,  $i$  is number of each joint  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$  are coefficients to generating the gait for various wave. The gait for biped robot is changed by operating these coefficients.

#### 2.2.2. Adaptation to simulation

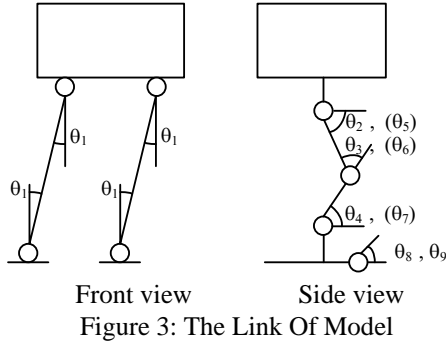
A sampling time for the function to generate the gait is quarter a gait cycle. The generated angle data is allocated the joint for position control value. A joint will move as a constant velocity between control points. In this simulation, 1 cycle of walking is defined 1.2 seconds. Thus, angular velocity is give as follows:

$$\omega = \frac{2\pi}{1.2} \quad (3)$$



3 cycle of walking time is 3.6 seconds. And the total time is 4.8 seconds taking 1.2 seconds in order to check after walking stability. In this simulation, 1 step takes 0.2 seconds, thus the number of total step was 240 steps. For example, a gait pattern of a joint angle which is made by the gait function.

The position of the joint is shown in Figure. 3 In addition, because of the servomotor of the joint of the biped robot, its joint can be rotated by 60 degree every 0.14 seconds. The rotate directions for each joint are shows in Table. 1. Gait Functions of model are defined as follows:



$$\theta_1 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ \theta_1(t) & \text{if } 0 < t \leq 3.3 \end{cases} \quad (4)$$

$$\theta_2 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ 0 & \text{if } t = 0.3 \\ 30 & \text{if } t = 3.3 \\ \theta_2(t) & \text{if } t = 0 < t < 3.3 \end{cases} \quad (5)$$

$$\theta_3 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ 0 & \text{if } t = 0.3 \\ 60 & \text{if } t = 3.3 \\ \theta_3(t) & \text{if } t = 0 < t < 3.3 \end{cases} \quad (6)$$

$$\theta_4 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ 0 & \text{if } t = 0.3 \\ 30 & \text{if } t = 3.3 \\ \theta_4(t) & \text{if } t = 0 < t < 3.3 \end{cases} \quad (7)$$

$$\theta_5 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ 30 & \text{if } t = 0.3 \\ 0 & \text{if } t = 3.3 \\ \theta_2(t + 0.6) & \text{if } t = 0 < t < 3.3 \end{cases} \quad (8)$$

$$\theta_6 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ 60 & \text{if } t = 0.3 \\ 0 & \text{if } t = 3.3 \\ \theta_3(t + 0.6) & \text{if } t = 0 < t < 3.3 \end{cases} \quad (9)$$

$$\theta_7 = \begin{cases} 0 & \text{if } t = 0 \text{ or } t > 3.3 \\ 30 & \text{if } t = 0.3 \\ 0 & \text{if } t = 3.3 \\ \theta_4(t + 0.6) & \text{if } t = 0 < t < 3.3 \end{cases} \quad (10)$$

$$\theta_8 = \begin{cases} 0 & \text{if } t = 0 \\ 0 < \theta_8 < 30 & \text{if } 0 < t' \end{cases} \quad (11)$$

$$\theta_9 = \begin{cases} 0 & \text{if } t = 0 \\ 0 < \theta_9 < 30 & \text{if } 0 < t' \end{cases} \quad (12)$$

In addition, minimum rotation angle use 0 degree in this simulation. Eq. (4) - (8) defines the behavior of lifting right leg of the robot to stop its movement after 3.3 seconds. Eq. (7) - (10) define the behavior of lifting left leg of the robot to start walking while 0 to 0.3 second.

Table 1: Parameter And Rotative Direction

Parameter	Leg	Joint	Rotative Direction
$\theta_1$	Both	Hip and Ankle	Side-to-Side
$\theta_2$	Right	Hip	Backward-and-Forward
$\theta_3$	Right	Knee	Backward-and-Forward
$\theta_4$	Right	Ankle	Backward-and-Forward
$\theta_5$	Left	Hip	Backward-and-Forward
$\theta_6$	Left	Knee	Backward-and-Forward
$\theta_7$	Left	Ankle	Backward-and-Forward
$\theta_8$	Right	Tiptoe	Backward-and-Forward
$\theta_9$	Left	Tiptoe	Backward-and-Forward

Knee joints do not rotate to backward direction from standing. Thus, these joint are stricter rotating to minus angle as follows:

$$\theta_3 = \begin{cases} 0 & \text{if } \theta_3 < 0 \\ \theta_3 & \text{if } 0 \leq \theta_3 \end{cases} \quad (13)$$

$$\theta_6 = \begin{cases} 0 & \text{if } \theta_6 < 0 \\ \theta_6 & \text{if } 0 \leq \theta_6 \end{cases} \quad (14)$$

The horizontal surface is applied for the ground surface of the simulation. Moreover, for the ground surface, Friction and restitution coefficients are defined as 0.3 and 0.0, respectively.

### 2.2.3. Formulation of the optimization

In the determinate optimization by using APGA, design variable vectors, an objective function, a penalty function and constraints are defined as shown from Eq. (15) to Eq. (21).

Design variable vector (DVs):

$$X_i = [a_i, b_i, c_i, d_i, e_i]; \quad (i = 1, 2, 3, 4) \quad (15)$$

$$X_{all} = [X_1, X_2, X_3, X_4],$$

*Objective function:*

$$F = -Y_d + \gamma P \rightarrow \text{Min.} \quad (16)$$

*Penalty function:*

$$P = \sum_{j=1}^3 \max(g_j, 0) + h_1 \quad (17)$$

*Constraint functions:*

$$g_1 = \begin{cases} 30.0 - X_d \leq 0 & \text{if } X_d > 0 \\ 30.0 + X_d \leq 0 & \text{Otherwise} \end{cases}, \quad (18)$$

$$g_2 = \begin{cases} 5.0 - R_d \leq 0 & \text{if } R_d > 0 \\ 5.0 + R_d \leq 0 & \text{Otherwise} \end{cases}, \quad (19)$$

$$g_3 = 200.0 - Z_h \leq 0 \quad (20)$$

$$h_1 = 240 - N_s = 0 \quad (21)$$

The objective function is minimized. In Eq. (16)  $Y_d$  denote the distance between centers of the biped robot model as shown in Figure. 4. The penalty coefficient is the value of  $\gamma = 1.0$ . The penalty function includes four constraint functions. In Eq. (18)  $g_1$  and  $X_d$  are the distances at the side under  $\pm 30$  [mm]. In Eq. (19)  $g_2$  and  $R_d$  is the angle to the direction under  $\pm 5$  degree. In Eq. (20)  $g_3$  and  $Z_h$  are the heights from the ground to the hip part. It is over 200[mm] to check to slipping at the end of the simulation. In Eq. (21)  $N_s$  is the number of steps should be 240 to indicate the success of the simulation

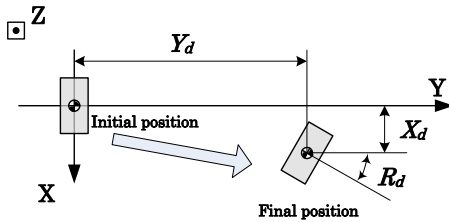


Figure 4: Overview Of The Simulation

Figure. 5 shows the approximated optimization process by following steps.

1. Initial design is initialized by specifying the simple Analysis.
2. For making a response surface model (RSM),  $X_{all}$  to generate the angle of each joint is defined. Moreover, random sampling is performed to get results of simulation.
3. Random sampling and then simulated to obtain the results for making the RSM.
4. Using APGA for optimization, the design variables were optimized by APGA based on RSM.
5. The design variables from APGA in step 4. are used to verify the result by simulation.

6. Verification process is conducted to check the convergence of the solutions. If the convergence is achieved, the optimal design will be stopped. Otherwise, the repetition of this process from step 3. will be carried out.

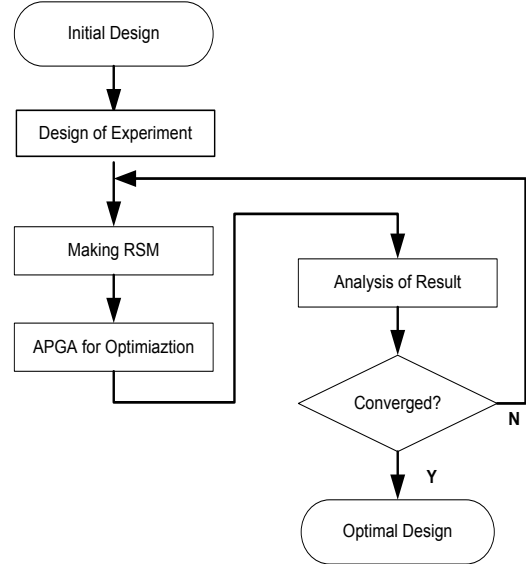


Figure 5: Approximated Optimization Process

#### 2.2.4. Response Surface Model (RSM)

In the optimization process of gait pattern for the robot is used a lot of calculation times. A large portion of this cost can be avoided using RSM by approximating the more cost analysis. Therefore, various optimization problems with this cost are applied by this method. In the paper used Response Surface Model (RSM) of 3<sup>rd</sup> orders (Cubic polynomial) to approximate response of an actual analysis code. A number of exact analyses using the simulation code(s) have to be performed initially to construct a model with a set of analyzed design point can be used. The 3<sup>rd</sup> order (Cubic) model is represented by a polynomial of the follows:

$$\begin{aligned} \tilde{F}(x) = & a_0 + \sum_{i=1}^N b_i x_i + \sum_{i=1}^N c_{ii} x_i^2 \sum_{ij (i < j)} c_{ij} x_i x_j \\ & + \sum_{i=1}^N d_i x_i^3 \end{aligned} \quad (22)$$

Where:

$N$  is number of model inputs or number of variable

$X_i$  is the set of model inputs

$a, b, c, d, e$  are the polynomial coefficients

The number of sampling for initialization equaled to the number of polynomial coefficients which for a cubic polynomial is:

$$\frac{(N+1)(N+2)}{2} + N \quad (23)$$

The number of sampling is calculated using the equation Eq. (23) and  $N = 20$  (number of design variables), and is obtained as 215 of data for the calculation of all polynomial coefficients.

### 3. ADAPTIVE PLAN SYSTEM WITH GENETIC ALGORITHM (APGA)

APGA has been proposed to solve the multi-peak optimization problems with multi dimensions. The proposed algorithm is a stochastic global search heuristics in which EAs based approaches are combined with local search method. APGA uses the Adaptive Plan (AP) to control its optimization process for the local search process. APGA differs in handing design variable vectors (DVs) from genetic algorithm (GA) encode DVs into genes, and handle them through GA operator. However, APGA encodes control variable vectors (CVs) of AP, which searches to local minima, into its genes. CVs decide on global behavior of AP, and DVs are handled by AP in the optimization process of APGA. APGA has been confirmed to improve the calculation cost and the stability of convergence towards the optimal solution.

#### 3.1. Formulation of the simple APGA

APGA is developed for overcoming the difficulty in controlling switching, choosing and steering a combination of global and local search method. On the other hand, Natural and artificial system adapt the behavior of themselves to the changing global and social environments over generations. These systems have been defined as an Adaptive System (AS) by Honlland (Honlland 1992), and AS has an adaptive plan (AP) which decides on its behavior through response to environment (ENV).

APGA is introduced its concept to new evolution algorithm (EA) strategy for multi-peak optimization problems. Hence, this study considers a global search method as AS, a local search method as AP and an optimization space of DVs as ENV, respectively. The conceptual strategy of APGA is show in Figure. 6. APGA differs in handing DVs from general EAs based on GAs to overcome these difficult problems. EAs generally encode DVs into genes of chromosome, and them through GA operator. However, APGA separates DVs of global search and local search methods completely. It encodes control variable vector (CVs) of AP, which searches to local minima into its genes on AS. CVs decide on a global behavior of AP, and DVs are handled by AP in the optimization process of APGA. The generation process of DVs is show in Figure. 7. This process generates a new DVs  $X_{t+1}$  from current search point  $X_t$  according to the following formula:

$$X_{t+1} = X_t + AP(C_t, R_t), \quad (24)$$

Where  $AP()$ ,  $C$ ,  $R$  and  $t$  denote a function of  $AP$ , CVs, response value (RVs) and generation, respectively. AS is controlled by the behavior of  $AP()$

via feedback loop of fitness value  $f$  or from a function of problem (ENV) during the global search process. Moreover,  $C$  can be renewed by estimating  $f$  by suing the GA operators within this process, and their trends are believed to make optimal behaviors like a cooling temperature of Simulate Annealing (SA).

#### 3.2. Adaptive plan (AP): Sensitivity plan

In this paper, the plan introduces a DV generation formula using a sensitivity analysis, which is effective in the convex function problem as a heuristic rule, because a multi-peak problem is combined of convex functions. This plan uses the following equation.

$$AP(C_t, R_t) = -Scale \cdot SP \cdot sign(\nabla R_t) \quad (25)$$

$$SP = 2C_t - 1, 0.0 \leq c_{i,j} \leq 1.0, \quad (26)$$

Where scale and  $\nabla R$  are denoted the scale factor and sensitivity of the RVs, respectively. A step size  $SP$  is defined by CVs for controlling a global behavior to prevent it from falling into the local optimum.  $C = [c_{ij}, \dots, c_{ip}]$  is used by Eq. (26) so that it can change the direction to improve or worsen the objective function, and  $C$  is encoded into a chromosome. In addition,  $i$ ,  $j$ , and  $p$  are the individual number, design variable number, and its size.

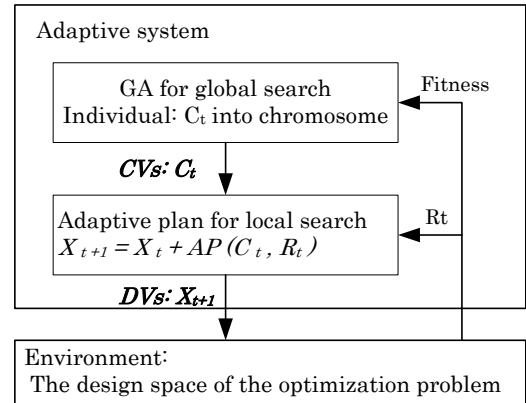


Figure 6: Conceptual Strategy Of APGA

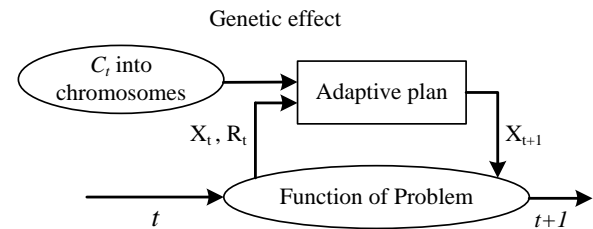


Figure 7: Generation process of DVs

### 3.3. Ingenuities for DVs and CVs

#### 3.3.1. Handing of significant figures

In the optimal design of the product design, dimension of products can be mainly dealt with as DVs. These are always assigned dimensional accuracies on a

mechanical drawing. Therefore, a value of DVs is done well to use a number of significant figures of assigned dimensional accuracy in its drawing in the optimal process. In APGA, a number of significant figures of DVs are defined, and DVs truncated to it within optimal process.

### 3.3.2. Handling of DVs's out of range

DVs are renewed by AP, and when their values exceed the range of them, returns by Eq. (27) into the range of them.

$$\begin{cases} \text{if } X_t < X^{LB} & \text{then } X_i = X^{LB} \\ \text{if } X_t > X^{UB} & \text{then } X_i = X^{UB} \end{cases} \quad (27)$$

### 3.3.3. Coding into chromosome for CVs

This 10 bit string with two values 0 and 1 represents a real value of CVs by using procedure of Figure. 8. In addition, Figure. 8. shows DVs and CVs of two dimensional cases.

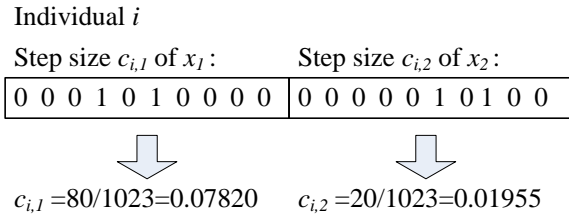


Figure 8: Encoding Into Genes

## 3.4. GA operators

### 3.4.1. Selection

Selection is performed using tournament strategy to keep a diverseness of individuals at early generations. The tournament size of 2 is used.

### 3.4.2. Elite strategy

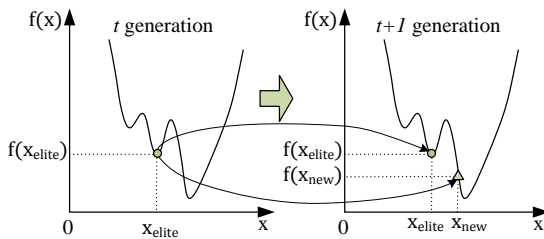


Figure 8: Elite Strategy

An elite strategy, where the best individual survives in the next generation, is adopted during each generation process. It is necessary to assume that the best individual, i.e., as for the elite individual, generates two global behaviors of AP by updating DVs with AP, not GA. Therefore, its strategy replicates the best individual to two elite individuals, and keeps them to the next generation. As shown in Figure. 8, DVs of one of them ( $\blacktriangle$  Symbol) is renewed by AP, and its CVs which are

coded into chromosome are not changed by GA operators. Another one ( $\bullet$  Symbol) means that both DVs and CVs are not renewed, and are kept to next generation as an elite individual at the same search point.

### 3.4.3. Crossover and mutation

In order to pick up the values of each CV, a single point crossover is used for string of each CV. This can be considered to be a uniform crossover for the string of chromosome as show in Figure. 9. The mutation is adapted for string of each CV, and its method reverses the 1 bit in its string.

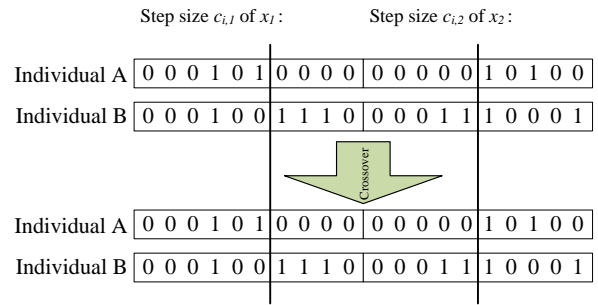


Figure 9: Example Of Crossover

### 3.4.4. Recombination of gene

At following conditions, the genetic information on chromosome of individual is recombined by a uniform random function.

- One fitness value occupies 80% of the fitness of all individuals.
- One chromosome occupies 80% of the population.

If this manipulation is applied for general GAs, improved chromosome which DVs have been encoded into is broken down. However, in APGA, genetic information is only CVs to make a decision of a behavior of AP. Therefore, to prevent from falling into a local optima, and to get out from the condition converged with a local optimal search process can recalculate by recombination gene of CVs into chromosome. This strategy is believed to make behavior like a reannealing of SA.

## 4. RESULTS OF THE SIMULAION

### 4.1. Results of the simulation on flat plate

The simulation model of small biped robot with passive tiptoe mechanism simulated by using APGA method, it can walk on flat ground (FG). The results are shown in Table 2. We set the number of iterations to 500. Maximum distance is  $116[mm]$ , distance side ( $X_d$ ) is  $24.8[mm]$  in the case of 24 time.

In the rough ground (RG), the results are not satisfactory because robots cannot walk or can walk but falls down. As a result, we cannot make RSM for using

APGA method. The rough ground will be study in the further. However, the trajectory of the robot's Center of mass (CoM) that can walk shown in Figure. 10.

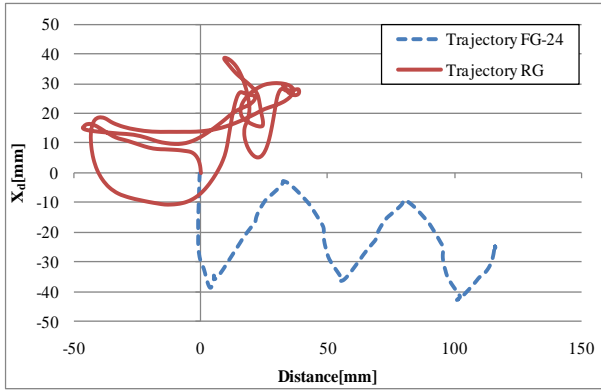


Figure 10: Center Of Mass (CoM) Trajectories Of The Robot Simulation.

The trajectory of the robot's CoM is compared FG-24 with RG as shown in Figure. 10. In RG, trajectory is small and walking is awkward. The other side, the trajectory of FG-24 is larger and similar to human walking trajectory (Gait Analysis Based on Joint Moment 1997). Therefore, we expect this biped robot will be able to walk in rough ground similar to a human walking.

Table 2: The Simulation Results

No. Iteration	X	Y	$R_d$	Z	$N_s$
4	-22.7	120.2	7.0	244.0	240.0
7	-28.8	114.9	9.1	244.0	240.0
11	-27.9	119.4	10.6	244.0	240.0
14	-24.5	119.1	5.9	244.0	240.0
16	-20.6	119.5	5.4	244.0	240.0
18	-22.7	121.1	6.1	244.0	240.0
20	-22.4	122.5	7.1	244.0	240.0
22	-24.9	119.0	5.9	244.0	240.0
23	-24.9	119.0	5.9	244.0	240.0
<b>24</b>	<b>-24.8</b>	<b>116.0</b>	<b>4.8</b>	<b>244.0</b>	<b>240.0</b>
25	19.5	121.8	7.2	244.0	240.0

Waveforms of the gait functions assigned to joints are compared FG4, FG24 and RG as shown in Figure. 11-14. The widely of waveform of less changed, and in  $\theta_1(t)$  (hip and ankle joints, rotating side to side) and  $\theta_4(t)$  that similar to cosine function.

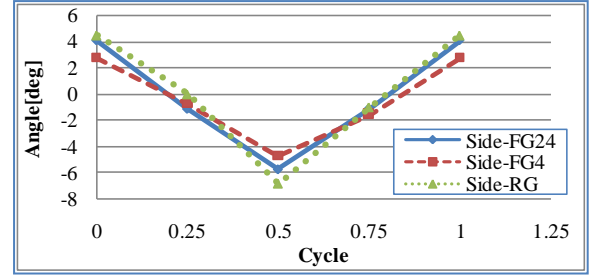


Figure 11: A Cycle Of Gait Function  $\theta_1(t)$  @ (Hip And Ankle, Side To Side)

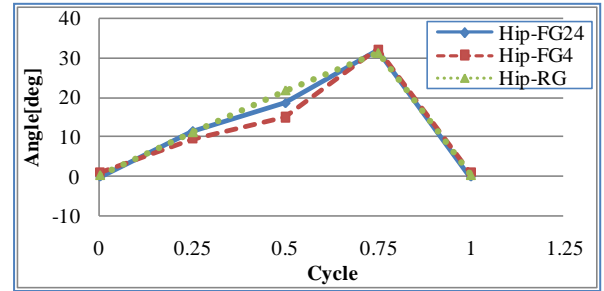


Figure 12: A Cycle Of Gait Function  $\theta_2(t)$  @ (Right Hip)

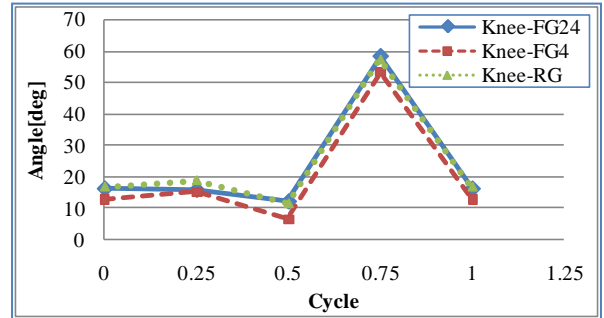


Figure 13: A Cycle Of Gait Function  $\theta_3(t)$  @ (Right Knee)

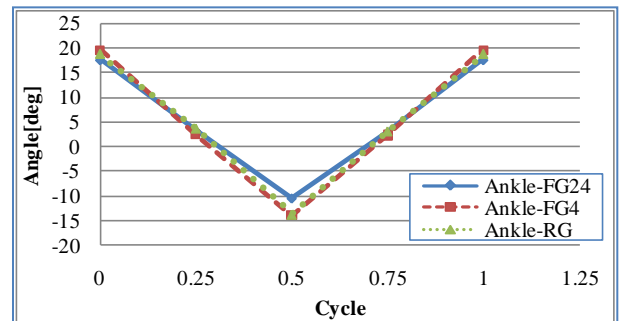


Figure 14: A Cycle Of Gait Function  $\theta_4(t)$  @ (Right Ankle)

## 5. CONCLUSION

We discussed a SBDO framework by approximated optimization process using APGA. The simulation of robot with tiptoe can walk and obtained a good gait pattern on flat ground. The RSM is created it can reduce scope of the problems, solves problems quickly and effectively. However, it is difficult to design and make the RSM to cover all problems. This is seen as a simulation on rough ground. This is an important problem which must be resolved and studied further.

The APGA is another tool to ultimate effectiveness to find the optimal solution. In this paper, the gait patterns generated by APGA converge fast and accurately. Regularly, it depends on the accuracy in the design of the experiment.

As our next target, we will resolve in simulation framework such as on rough ground. Finally, real small biped robot walking will be test as an experiment.

## ACKNOWLEDGMENTS

This work was supported by Shibaura Institute of Technology Japan Scholarship. I would like to extend my heartfelt thanks to the Shibaura Institute of Technology, Professor. Hiroshi Hasegawa and everyone for their suggest and help in making this paper.

## REFERENCES

- Zhe, T., Zengqi, S., and Changji Z., 2006, GA Based Optimization for Humanoid Walking, *ICGST-ARAS journal*, 5(2), pp.1-10, June
- Lingyun, H., Changjiu, Z., and Zengqi S., 2006, Biped Gait Optimization Using Spline Function Based Probability Model, *Proceeding of the 2006 IEEE International Conference on Robotics and Automation*, pp. 830-835, Orlando, Florida
- Xiang, Y., Chung, H.J, Mathai. A., Rahmatalla, S., Kim, J., Marler, T., Beck, S., Yang, J., Arora, J.S., Abdel-malek, K., 2007, Optimization-based Dynamic Human Walking Prediction, *SAE Human Modeling for Design and Engineering Conference*, June 12–14, Seattle, Washington
- Nandha, H., Jungwon, Y., Gabsoon, K., 2008, Gait Pattern Generation with Knee Stretch Motion for Biped Robot using Toe and Heel Joints, *International Conference on Humanoid Robots*, pp. 265-270, Daejeon, Korea
- Cheol. K.A., Min, Cheol, L., and Seok, J.G., 2003, Development of a biped robot with toes to improve gait pattern Advanced Intelligent Mechatronics, *IEEE/ASME International Conference*, pp. 729-734
- Shuuji Kajita, Kenji Kaneko, Mitsuhasu Morisawa, Shinichiro Nakaoka and Hirohisa Hirukawa, 2007, ZMP-based Biped Running Enhanced by Toe Springs, *IEEE International Conference on Robotics and Automation Roma*, pp. 3963-3969, Italy
- Moosavian, S.A.A., Alghooneh, M. Takhmar, A., 2007, Introducing a Cartesian Approach for Gate Planning and Control of Biped Robots and Implementing on Various Slopes, *Proceedings of the 2007 7th IEEE-RAS International Conference on Humanoid Robots*, pp. 545-550, Pittsburgh, United States of America.
- Naoya, I., Hiroshi, H., 2007, Robust optimization under uncertain factors of environment for simple gait of biped robots, *Proceedings of the 6th EUROSIM Congress on Modeling and Simulation*, Ljubljana, Slovenia
- Yu, Z., Ming, C.L., Dinesh, Manocha, Albertus, Hengrawan, Adiwahono., Chee-Meng, C., A Walking Pattern Generator for Biped Robots on Uneven Terrains, 2010, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 18-22, Taipei, Taiwan
- Holland, J., 1992, Adaptation in Natural and Artificial Systems. *The University of Michigan 1975*, MIT Press
- Hiroshi, H., 2007, Adaptive Plan System with Genetic Algorithm based on Synthesis of Local and Global Search Method for Multi-Peak Optimization Problems, *Proceedings of the 6th EUROSIM Congress on Modeling and Simulation*, Ljubljana, Slovenia
- Chockalingam, N., Ashford, RL., 2007, Selected foot length ratios in a non-pathological sample. *Revista Internacional de Ciencias Podologicas*, 1(2), pp. 25-30
- Masayuki, S., Naoya, I., and Hiroshi, H., 2009, The Simulation of Tiptoe mechanism for Biped Robot, *Asia Simulation Conference 2009 (JSST 2009)*, October 7-9, Shiga, Japan
- iSIGHT is a trademark of Engineous Software Inc., 2004, *iSIGHT version 9.0*, pp. 321-327, United States of America.
- The Clinical Gait Analysis Forum of Japan.1997, *Gait Analysis Based on Joint Moment*. Ishiyaku publisher, pp. 19-22, Japan

## AUTHORS BIOGRAPHY

**Krissana Nerakae** received her BE and ME from Suranaree University of Technology, Thailand, in 2005 and 2009, She study in Graduate School of Mechanical Engineering in Shibaura Institute of Technology. Her research interests are optimization methods and robotics technology.

**Hiroshi Hasegawa** received his B.E. (1992) and M.E. (1994) from Shibaura Institute of Technology, Japan. He received PhD (1998) in Mechanical Engineering from Tokyo Institute of Technology, Japan. He has been working at Shibaura Institute of Technology, and currently is a Professor. He is member of JSME, ASME, JSST, JSCS and JSDE. His research interests include Creativity of Design, CAX (Computer Aided eXploration) and Biped Robot.



# RFID-BASED REAL-TIME DECISION SUPPORT IN SUPPLY CHAINS

Tobias Hegmanns<sup>(a)</sup>, Michael Toth<sup>(b)</sup>

<sup>(a)(b)</sup> Fraunhofer-Institute for Material Flow and Logistics, J.-v.-Fraunhofer Str. 2-4, 44227 Dortmund, Germany

<sup>(a)</sup> [tobias.hegmanns@iml.fraunhofer.de](mailto:tobias.hegmanns@iml.fraunhofer.de), <sup>(b)</sup> [michael.toth@iml.fraunhofer.de](mailto:michael.toth@iml.fraunhofer.de),

## ABSTRACT

For an effective global order management, the concepts of RFID and company-wide information brokerage offer promising technological background.

This paper will introduce the research project RAN (RFID-based Automotive Network) which develops a new way of collecting and sharing RFID based information in supply networks. This paves the way for new IT-systems, so called Logistic Assistant Systems (LAS), which use this new real-time process information for improving decision support. This article describes a simulation-based decision-support system based on real-time RFID event information from the supply chain.

**Keywords:** Supply chain management, supply chain simulation, simulation-based decision support systems, capable-to-promise planning, RFID

## 1. INTRODUCTION

Considering new challenges in global markets, companies are forced to plan their supply chains relating to more flexibility, effectiveness and cost reduction. High-value-added branches, like the automotive industry, are characterized by complex product configurations resulting in thousands of possible parts variants. Furthermore a high percentage of value creation is done by global supply networks with high order lead times, dynamic and risky transport relations and high transfer stocks.

Today one mayor problem in automotive supply networks is the lack of transparency, especially concerning in-transit stocks and the material flow. Often simple questions like “Where is my part no. 12345 now?”, “How long does it take until shipment 67890 will be available in my warehouse?” or “How many parts of type A will be available next week?” cannot be answered. The problem is a very heterogeneous IT-infrastructure and insufficient communication between the supply chain partners. Furthermore simple basic information about the current status and geographic position of an object in a supply network is often nonexistent or unavailable for partners or logistic service providers.

The consequence is that shortage situations or delays in supply are often not detected. Either costly extra transports or expedited freight are initiated to

avoid possible bottleneck situations or high inventory buffers are held. Both results in increased costs.

For an effective global order management, the concepts of RFID and company-wide information brokerage offer promising technological background. RFID and other AutoID technologies enable capturing a real-time picture of the supply chain operations. Each movement of goods, each operational step is documented by a RFID read point event. Thus, the current as-is state of the system is accessible at any time and the system state is updated with each object passing through a read point.

However, up to now companies apply RFID solutions mainly in closed-loop systems. A standardized industry-wide exchange of RFID/Auto-ID information between manufacturers, logistic service providers and suppliers has not been realized so far (Lange 2011). Some individual solutions exist, but the technology can not yet exploit its full potential for the management of supply chains due to missing standards and collaborative solutions. The automotive industry is a prominent example for this situation. Especially first tier suppliers suffer from this situation since they have to deal with multiple OEM-specific solutions.

This paper will introduce the objectives of the research project RFID-based Automotive Network (RAN) which develops a new way of collecting and sharing RFID based information in supply networks. This paves the way for new IT-systems, so called Logistic Assistant Systems (LAS), which use this new real-time process information for improving decision support. This article describes a simulation-based decision-support system based on real-time RFID event information from the supply chain.

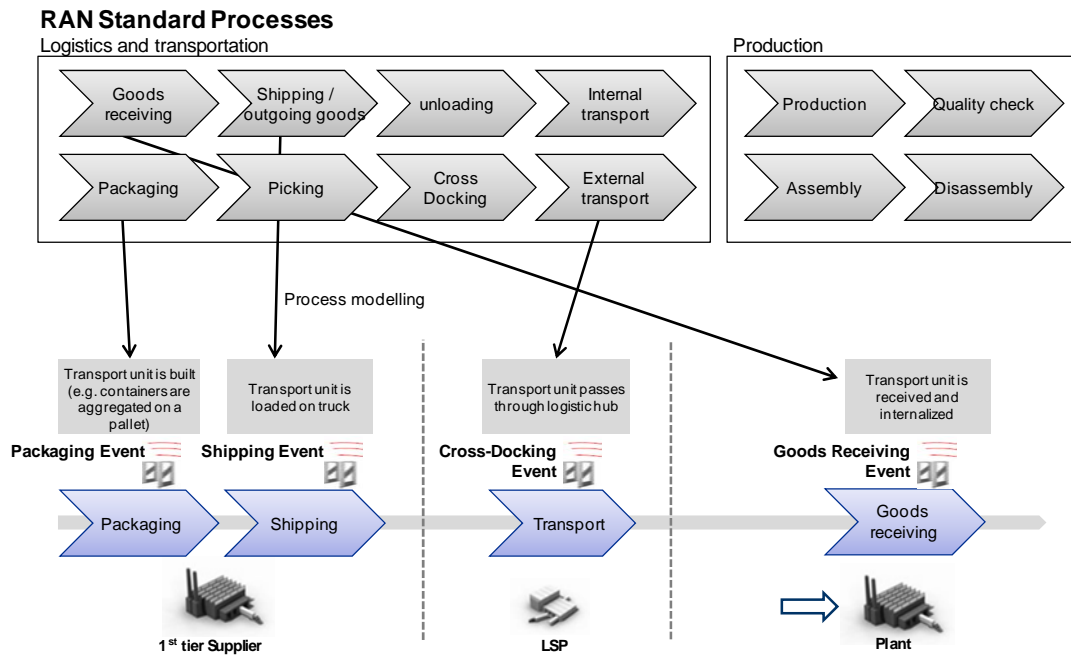


Figure 1: Example of process modeling using the set of RAN standard processes

## 2. RAN – RFID-BASED AUTOMOTIVE NETWORK

The research project “RAN – RFID-based Automotive Network” funded by the Federal Ministry for Economics and Technology (BMWi) has the objective to increase the transparency of information exchange in production and logistic networks of the automotive industry (FMET 2010) through the application of RFID. The consortium is composed of many partners involved in the production and logistics chain: Automotive OEMs (e.g. Daimler, BMW, Opel), suppliers (e.g. Bosch, Kaiper, Rehau) and logistic service providers (e.g. BLG) as well as RFID software providers and system integrators (e.g. IBM, SAP, Siemens) and leading research institutes (BIBA, IWB, FZI and IML).

The RAN project intends to standardize the technological infrastructure, the event data and informational systems so that event information can be shared across company borders easily. The core elements of the RAN approach to achieve this are the following:

**RFID-based RAN standard processes:** So called RAN standard processes give a reference which RFID-events typical material flow processes can provide. These RFID-events represent the kind of information relevant to supply chain planning and event management systems. For typical supply chain operations, such as goods receiving process, outgoing goods process, transport, production, assembly and other common material flow processes, events are defined. Across several use cases with various automotive OEM, suppliers and logistic service providers it was analyzed which informational content these events must contain to support planning and event management in different supply chain scenarios, such as

monitoring and control of KANBAN-call-offs for material, track & trace of reusable containers, available-to-promise planning in global supply chains, monitoring and control of vehicle distribution operations.

In Figure 1 the set of RAN standard processes is depicted. From this set of generic material flow processes the real supply chain process and its RFID deployment can be constructed. Supply chain partners agreeing on using the automotive RAN standard can rely on the availability of the defined RFID event information from a partner’s process.

**Data and event standardization:** each event used in the RAN project scope is exactly defined. The target is a common industry-wide understanding of the process events for the users of the RAN process and event standard.

The EPCglobal EPCIS 1.0 (EPCglobal 2007) standard is adopted to specify the RFID events. Where necessary, extensions to the existing EPCIS standard are made to meet the needs of the automotive industry. The EPCIS standard allows encoding the relevant information about the object identified and the circumstances of the capturing:

- What? Unique identifier of the object or the or the transaction
- When? Time log of the event
- Where? Read point plus business location
- Why? Business step (e.g. goods arriving, goods receiving) + Disposition (e.g. container empty, damaged etc.)

For the identification of logistic or production objects the Electronic Product Code EPC as well as the

ISO-standard is applied. Both standards ensure that information sharing between suppliers, customers and logistic service providers works reliably and the unique reference of a part or object to its manufacturer or owner is possible. Table 1 depicts the identification standards used for typical logistic items in the RAN project.

Table 1: Identification standards used in RAN

Object type	ISO framework	
	EPC	ISO norm
Freight containers	ISO 17363	
		15459-2
Returnable transport items	ISO 17364	
	GRAI	15459-5
Transport units	ISO 17365	
	SSCC	15459-1
Product packaging	ISO 17366	
	SGTIN/ GPIK	15459-4
Product tagging	ISO 17367	
	SGTIN/ GPIK	15459-4
Locations	SGLN	
Fixed Assets	GIAI	

**RFID read point infrastructure and RFID-technology:** read point hardware serves to read RFID tag information from logistic objects, e.g. incoming / outgoing objects, objects entering or leaving a process; edge- and middleware captures the object event and transfers it to connected applications or storage repositories. The goal is to define standard settings for typical automotive applications.

### Standardized inter-operable IT infrastructure:

Building on the standardized processes and data structures a so called InfoBroker infrastructure serves as a basis for effective exchange of object-related event data between companies. Distributed event repositories constitute the core of the InfoBroker (see Figure 2). Each repository is assigned to a defined organizational domain. It stores the event information generated within this domain in standardized form. Using query and subscription services companies can exchange event information between repositories and other periphery systems. Information exchange is controlled by rules for security and privacy of data. Each domain owner must authorize external partner to receive or retrieve information from his repository.

### 3. THE BENEFIT OF IMPROVED TRANSPARENCY

Transparency about the system's current state is clearly valuable for the management of logistic systems. But the visibility of logistic objects through read point events itself, the tracking and tracing of objects on their path through a logistic system, is only the bottom of the potential benefit an improved informational transparency can yield. The possible benefits new IT-systems can achieve on the basis of process event data is manifold. The functionalities of these new IT-systems reach from the simple localization of logistics objects over event management and altering systems up to complex decision support systems. In order to systematize these IT-functionalities, a classification scheme of supply chain IT-systems building on real-time event information from the process level is introduced below. The classification scheme is shown in Figure 3. The complexity of the IT-system's functionality increases from level 1 to 6:

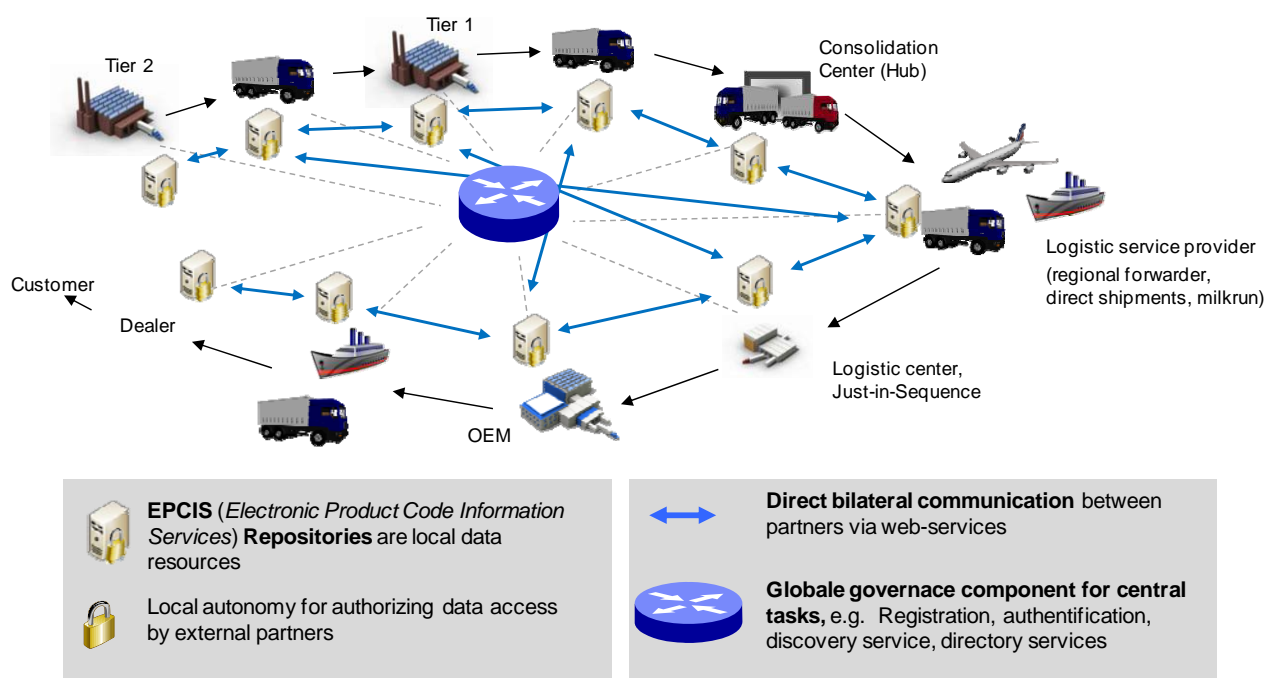


Figure 2: The RAN Infobroker concept

**Level 1 system:** representation of the actual current state; track and trace functionalities; locate-objects functionalities; reports on simple logistic key performance indicators (e.g. lead times, transit times; inventory)

**Level 2 system:** integration of to-be-state information like ETA, order and shipment number, advice note or production data

**Level 3 system:** comparison target vs. actual state; identification of deviations; visualization of deviations target vs. actual; monitoring functionalities

**Level 4 system:** identification of consequences; event management functionalities; detection of process deviations; integration of rules for alerting-functionalities; forward projection of the current state (simulation); analysis of the impact of process deviation on targets or desired future states; identification of critical process deviations

**Level 5 system:** evaluation of decision alternatives; configuration of solution measures and testing; application of decision support methods: simulation, optimization, cost-benefit calculation etc.

**Level 6 system:** integrated control of the supply chain; execution of processes; automatic problem solving processes; decisions or solution measures are instantly generated and initiated via established links to operative systems.

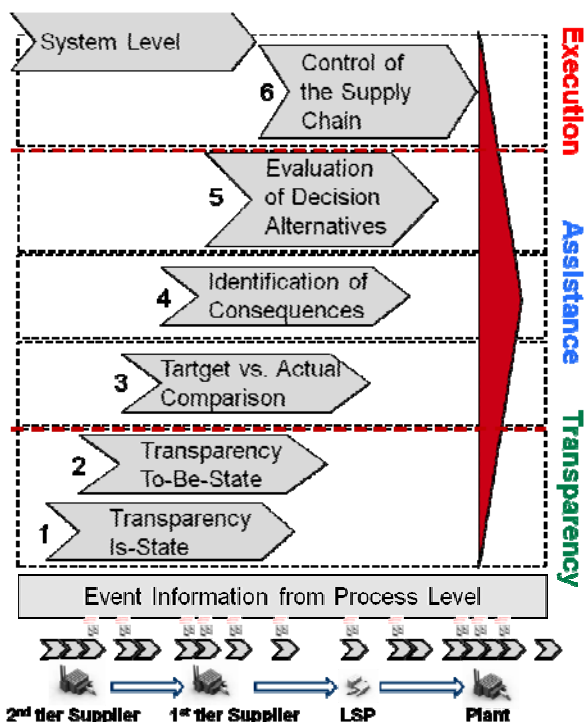


Figure 3: Levels of complexity of IT-applications based on real-time process event information

Within the RAN project IT-systems on all levels are being developed. This article will introduce the concept for a level 3 to 5 logistic IT-system based on real-time process event information.

#### 4. LOGISTIC ASSISTANCE SYSTEMS (LAS)

Logistic Assistance Systems (LAS) are designed to assist logistic experts in planning and execution by offering transparency about all relevant supply chain information and integrating specific decision support methods and planning approaches into one combined approach. Besides that the idea of LAS is to provide a simple planning and software system based on the RAN infrastructure, which can be adapted to new supply chains or planning situations with low effort. It can be easily integrated into a company's organization. Therefore LAS integrate events from the RAN infrastructure and planning data to consolidate all relevant information along the supply chain. Furthermore, LAS comprise monitoring and event management functionality to allow for effective supply chain planning and execution applying current and high quality data. Thus, the LAS concept rests upon extended collaboration, consistent supply chain information, process transparency and planning functionality.

The following sections will show the application of the LAS approach in a RAN use case dealing with monitoring and available-to-promise planning in global supply chains.

Consequently we divide this LAS concept into three blocks: Data acquisition and transparency, decision support and collaborative planning. These blocks allow dynamic and collaborative supply chain planning and will be introduced in the following sections.

**Data Acquisition and Transparency:** LAS need to consolidate all relevant information for a specified planning task. To achieve data acquisition and data transparency LAS integrate functionalities of SCM concepts like SCMo. SCMo is a collaborative multi-level SCM concept founded on software support for processing information between network partners (Odette 2009). The basic functionality is the exchange of production demands and inventory levels among business partners in a supply network to gain extended transparency and avoid time lags in information flow. Recent developments in SCMo applications integrate Supply Chain KPI Frameworks (Key Performance Indicators) and the assortment of graphical tools (e. g. predefined charts, cockpits or dashboards) for information presentation (Bäck and Gössler 2006). Exemplary KPIs that are applied in SCMo applications are forecasting accuracy and days of inventory (Hellingrath et al. 2008).

Consequently, SCMo systems provide functionalities for monitoring the current status of a supply chain (level 1 and 2). Nevertheless, they lack methods for forecasting and planning, i.e. optimization



and simulation. On the operational level the concept of Supply Chain Event Management (SCEM) aims to support the execution of agreed plans by automatic identification of unacceptable deviations and suggestion of alternative solutions (level 3 to 5). Therefore a SCEM system supports online data acquisition via tracking & tracing. It raises alerts if there are significant deviations between plan and current data. LAS need to combine functionalities of both SCM concepts but stick to a simple approach: This system concept focuses on information transparency for decision makers without implementing a complex system requiring organizational changes. Therefore LAS provide standardized interfaces to the RAN InfoBroker Architecture as well as tracking & tracing devices (e.g. XML specifications). This functionality is connected with the material flow level and extracts all relevant events including event interpretation and data aggregation. The key factor here is, that all relevant, task oriented information is collected and presented to the decision maker and critical deviations are directly illustrated on a dash board. The user interface is designed clear and simple; the technological background of this functionality is designed flexible in order to be prepared for dynamically changing environments and supply chain structures. Figure 4 shows typical data sources, which are relevant to enable sufficient information transparency in global supply networks.

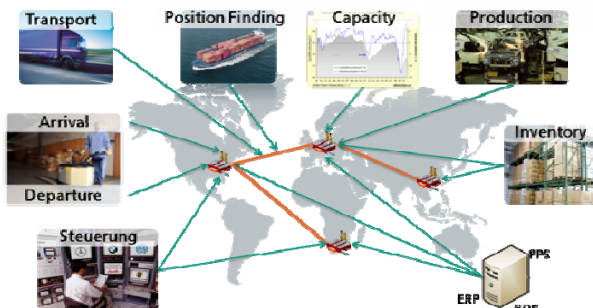


Figure 4: Forms of Data Acquisition and Consolidation

**Decision Support (DSP):** Decision Support (DSP) for LAS means that depending on the given logistic task the software system supports the human decision process in all its steps: decision preparation, option selection, decision execution and control. Depending on the focused planning task itself, LAS integrate different types of decision support concepts and systems: LAS realizations typically integrate a simulation based concept to enable available-to-promise (ATP) and capable-to-promise (CTP) planning in complex supply networks. Therefore, the DSP module comprises a simulation component, which allows scenario-based analysis of different demand, inventory and capacity situations within the supply network. Based on the current transport and inventory situation (warehouse and transport stock based on online data from the RAN

InfoBroker) and a precise model of the supply network, LAS allow the simulation of future system's behavior.

The simulation model is initialized with the current state situation and enables the simulation of existing transports and material flows to the final warehouses or production sites. In combination with given demands it is possible to calculate deviations between the future availability of material and the future demand. Furthermore the feasibility of a change of plan (demand change or later estimated time of arrival of a ship) can be checked against reliably and thus be optimized ("what-if-analysis"). Finally detected deviations of transports due to ETA and geographical position of a ship can be integrated into the simulation and possible delays as well as subsequent problems can be detected in time.

This so-called dynamic ATP allows for exact calculation if future demand could be fulfilled by then available inventory (see figure 5). This supports the user to control the chosen decision execution and its impact on the subsequent process (Toth and Wagenitz 2009).

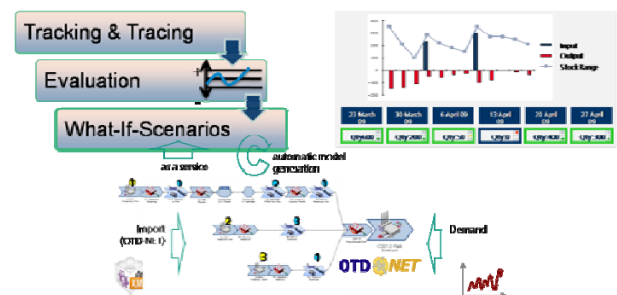


Figure 5: Simulation based ATP functionality and stock range calculation

The general idea of the DSP module is to offer flexible planning support by integrating different functionalities, i.e. simulation or optimization, as services into the architecture. Especially service oriented architecture and easy configuration of LAS is a current research topic in this context.

**Collaborative planning:** For collaborative planning LAS combine data from all relevant network partners. Thus, information from all supply chain tiers may be used for information processing and decision support. Supporting the planning process, manual changes in data need to be tracked and documented by users. In addition to that it is necessary to integrate individual views by a dedicated role-based user management. This way proprietary information can be revealed to selected user groups, network partners or organizational departments. To support an enhanced communication process for collaborative planning workflows and message systems must be applied (Bockholt et al. 2009, Deiseroth et al. 2008).

To sum up: By providing decision support and collaborative planning profound decisions can be made.

But only if relevant and current supply chain information is given. With RAN and RFID-based LAS an early identification of possible bottleneck situations, supply shortages or surplus stocks is possible which yields cost and service benefits. Decision makers may intervene in an early stage to ward off cost-intensive additional processes. Especially for global order management LAS provide company-spanning and consequently holistic collaboration to stay competitive within rapid changing markets.

## 5. CONCLUSION

This paper explains the motivation and objectives of the research project “RAN RFID-based automotive network” (FMET 2010), which aims at improving the transparency in logistic networks through real-time process event information. Furthermore Logistic Assistance Systems (LAS) are introduced to show, how this highly actual information can be applied to improve supply network planning and execution in different levels. Finally, the LAS approach was described within a successful implementation of a global supply chain use case.

## 6. REFERENCES

- Bäck, S.; Gössler, S., 2006. *SCM-Kompetenz-Management – Focus: Planungs- und Dispositionssysteme*. In: E.-N. Corinna, *Ausbildung in der Logistik*, Deutscher Universitäts-Verlag, 1st edition, pp. 155-179, Wiesbaden
- Bockholt, F.; Raabe, W.; Toth, M., 2009. *Logistic Assistance Systems for Decision Support in Collaborative Supply Chain Planning and Control, HMS Conference (Harbour, Maritime & Multimodal Logistics Modelling and Simulation)*, September 23-25, 2009, Tenerife, Spain, 2009.
- Deiseroth, J.; Weibels, D.; Toth, M.; Wagenitz, A., 2008. *Simulationsbasiertes Assistenzsystem für die Disposition von globalen Lieferketten*. In: Rabe, M., *Advances in Simulation for Production and Logistic Applications*, Fraunhofer IRB-Verlag, pp. 41-50.
- EPCglobal, 2007. *EPC Information Services (EPCIS) Version 1.0 Specification*, ratified April 12th 2007, [http://www.gs1.org/gsm/kc/epcglobal/epcis/epcis\\_1\\_0-standard-20070412.pdf](http://www.gs1.org/gsm/kc/epcglobal/epcis/epcis_1_0-standard-20070412.pdf) [accessed 2011/07/01]
- EPCglobal, 2010. *EPC Tag Data Standard Version 1.5*, ratified on August 18th 2010, [http://www.gs1.org/gsm/kc/epcglobal/tds/tds\\_1\\_5-standard-20100818.pdf](http://www.gs1.org/gsm/kc/epcglobal/tds/tds_1_5-standard-20100818.pdf) [accessed 2011/07/01]
- Federal Ministry of Economics and Technology, *RAN Rfid-based Automotive Network*, ‘<http://www.autonomik.de/en/218.php>’, [accessed 2011/07/01]
- Hellingrath, B.; Toth, M.; Hegmanns, T. Maaß, J.C., 2008. *Prozesse in Logistiknetzwerken. Supply Chain Management*. In: D. Arnold, H. Isermann, A. Kuhn, H. Tempelmeier and K. Furmans, ed. *Handbuch Logistik*, Springer-Verlag, 3rd edition, pp. 458-486.
- ISO International Organization for Standardization, *ISO 17363:2007 Supply chain applications of RFID -- Freight containers, ISO 17364:2009 Supply chain applications of RFID -- Returnable transport items (RTIs), ISO 17365:2009 Supply chain applications of RFID -- Transport units, ISO 17366:2009 Supply chain applications of RFID -- Product packaging, ISO 17367:2009 Supply chain applications of RFID -- Product tagging*, [http://www.iso.org/iso/iso\\_catalogue.htm](http://www.iso.org/iso/iso_catalogue.htm) [accessed 2011/07/01]
- ISO International Organization for Standardization, *ISO/IEC 15459-5:2007 Information technology -- Unique identifiers Part 1 - 5*, [http://www.iso.org/iso/iso\\_catalogue.htm](http://www.iso.org/iso/iso_catalogue.htm) [accessed 2011/07/01]
- Lange, B., 2011. *Automobilindustrie wagt gemeinsame RFID-Wege*, *VDI Nachrichten*, Nr. 13, 01.04.2011
- Odette International 2009. *Supply Chain Monitoring Version 1.0*, <http://www.odette.org/> [accessed 17 June 2009].
- Toth, M.; Wagenitz, A., 2009. *Neue Wege für die effektive Planung logistischer Netzwerke. Dynamische Verfügbarkeitsplanung mit Hilfe von Assistenzsystemen*, *Industrie Management*, vol. 25 (2), pp. 55-58.

## AUTHORS BIOGRAPHY

**Dr.-Ing. Tobias Hegmanns**, since 2006 he works as a researcher at the Fraunhofer-Institute for Material Flow and Logistics in Dortmund, Germany. He studied mechanical engineering at the TU Dortmund and holds a Masters Degree in Industrial Engineering from the Georgia Institute of Technology in Atlanta, USA. His research interests are planning and control methods for logistic networks, design of logistic processes and logistic IT-systems.

**Dr.-Ing. Michael Toth** is scientist and head of the Supply Chain Engineering department at Fraunhofer Institute for Material Flow and Logistics IML, one of the biggest logistic research institutes worldwide. He has been working for the IML since 2003 and is responsible for industry and research projects, especially in the areas automotive, supply chain management and logistic assistance systems LAS.



# FEATURE SELECTION FOR DATASETS OF WINE FERMENTATIONS

Antonio Mucherino<sup>(a)</sup>, Alejandra Urtubia<sup>(b)</sup>

<sup>(a)</sup> CERFACS, Toulouse, France

<sup>(b)</sup> Universidad Tecnica Federico Santa Maria, Valparaiso, Chile

<sup>(a)</sup> [mucherino@cerfacs.fr](mailto:mucherino@cerfacs.fr), <sup>(b)</sup> [alejandra.urtubia@usm.cl](mailto:alejandra.urtubia@usm.cl)

## ABSTRACT

The fermentation is the most important process in the production of wine. Problematic fermentations can cause losses to wine makers, because such fermentations could be too slow to provide the final product, or they may even become stagnant. An efficient prediction of problematic fermentations at the first stages of the process is therefore of great interest. The aim of this paper is two-fold. We apply a supervised biclustering technique to a dataset of wine fermentations with the aim of selecting and discovering the features that are responsible for the problematic fermentations. We also exploit the selected features for predicting the quality of new fermentations, and we propose a new strategy for validating the obtained classifications.

Keywords: wine fermentations, feature selection, classification, combinatorial optimization

## 1. INTRODUCTION

Wine is widely produced around the world. Industrial production of wine is an important business in many countries. For this reason, the study of the fermentation process, which is able to transform grape juice into the alcoholic beverage, is of increasing interest in the field of agriculture. Problematic fermentations, indeed, may cause losses to industries. If a fermentation process is slower than usual, for example, the final product is produced in a longer time. Moreover, in the worst case, when the fermentation process gets stuck, a part of the production could be completely spoiled.

Data mining is a field of operations research that analyzes large databases with the aim of acquiring novel knowledge (Mucherino, Papajorgji, Pardalos, 2009). In recent years, data mining techniques have specifically been applied to agricultural problems in order to find important information about the problem under study. In the case of wine fermentations, a database of compound measurements, taken at different times during the fermentation process, can be exploited for extracting information that can help the prediction of problematic fermentations. This prediction could allow wine makers to interfere with the problematic

fermentation processes in order to guaranteeing a good fermentation.

We present in this paper some analysis regarding wine fermentations. We consider a dataset obtained from a winery in Chile's Maipo Valley, which is the result of 24 measurements of industrial vinifications of Cabernet sauvignon (Urtubia, Perez-Correa, Meurens, 2004). By using a supervised biclustering technique, we select the compounds that are responsible for the problematic fermentations, and we also attempt a prediction of other unknown fermentations. In the case in which the fermentations are only divided in two classes (normal and problematic fermentations), we show that correct predictions can be performed by using the new classification strategy we present.

The rest of the paper is organized as follows. In Section 2, we discuss the problem of predicting wine fermentations, and we give more details about the considered dataset containing fermentations. In Section 3, we briefly present the data mining technique that we employ for performing wine predictions. In particular, we consider a supervised biclustering technique that is able to solve two problems at the same time (Busygin, Prokopyev, Pardalos, 2005). We can select the compounds measured during the fermentations that actually allow discriminating between normal and problematic fermentations, and we can also exploit the acquired knowledge for performing predictions on new fermentations. The basic strategy, as well as an improved strategy that we propose for performing the predictions, are both described in Section 4. Finally, Section 5 presents some experiments, and Section 6 concludes the paper.

## 2. PREDICTING WINE FERMENTATIONS

The problem of predicting wine fermentations is very important for wine makers. Some problems occurring during the fermentation process of wine can indeed impact the quality and the productivity of the final product (Mucherino, Papajorgji, Pardalos, 2009). Predicting how good the fermentation process is going to be may help enologists who can then take suitable steps to make corrections when necessary and to ensure that the fermentation process concludes smoothly and

successfully. Classic problematic fermentations are slow fermentations, and fermentations that can get stuck at a certain point of the process.

We consider in this paper a dataset of wine fermentations that have been obtained from 24 measurements of industrial vinifications of Cabernet sauvignon (Urtubia, Perez-Correa, Meurens, 2004), in a winery in Chile's Maipo Valley. The data are related to the harvest of 2002. The levels of 29 compounds are analyzed. Among them, sugars are analyzed, such as glucose and fructose, and also organic acids, such as the lactic and citric acids, and nitrogen sources, such as alanine, arginine, leucine, etc. The whole set of data consists of approximately 22000 data points. In this work, the used compounds are actually 30, because we added one for representing the total sugar, which consists of the sum of the glucose and fructose levels.

Measurements are sorted in a sequence from the first to the last. For each fermentation, measurements occurring approximately at the same time take the same position in this sequence. To this purpose, 15 temporal windows comprising 10 hours are defined, to which a subset of measurements are associated. When data in the temporal window were not available for some fermentation, they have been obtained by exploiting measurements close in time and by employing linear regression techniques. Finally, the considered set of data contains 24 fermentations described by  $15 \times 30 = 450$  features: the first class contains *normal* fermentations (9 in total), whereas the second class contains *problematic* fermentations (15 in total). Preliminary studies on this dataset have been presented in (Mucherino, Urtubia, 2010).

In order to verify the quality of the performed predictions, this dataset has been divided into training and testing set. The training set contains 16 randomly chosen fermentations: 6 normal fermentations and 10 problematic fermentations. The testing set is smaller in size and it contains 8 fermentations in total: 3 normal and 5 problematic fermentations.

In this work, we consider the training set for selecting the features that allow for performing correct classifications of the fermentations. To this aim, we search *consistent biclusterings* of the training set, which are able to associate subgroups of features to subgroups of samples of the dataset (each sample represents one fermentation). In order to obtain a consistent biclustering of the training set, some features are removed from the set. This is done by solving a combinatorial optimization problem. Details about this procedure are given in Section 3.

Once a consistent biclustering is found from a training set, the corresponding relationship between samples and features can be exploited for classification purposes. Given a testing set related to the same

problem, the classification of its samples, by definition, is supposed to be not known. However, the classification of its features is known, because it is exactly the same of the training set, and this information can therefore be exploited for reconstructing the classification of the samples of the testing set. More details are given in Section 4.

### 3. FEATURE SELECTION BY CONSISTENT BICLUSTERING

Let  $A = \{a_{ij}\}$  be a matrix representing a dataset. The matrix  $A$  contains  $n$  samples (column by column) and  $m$  features (row by row). A *bicluster* is a submatrix of  $A$ , that is able to group together a subset of samples (a class  $S_i$ ) and a subset of features (a class  $F_j$ ) of the dataset. A biclustering

$$\mathbf{B} = \{(S_1, F_1), (S_2, F_2), \dots, (S_k, F_k)\}$$

is a partition of  $A$  in disjoint biclusters that covers  $A$  (Busygina, Prokopyev, Pardalos, 2005).

Biclusterings of datasets are usually searched by unsupervised techniques, where it is supposed that no information about the data is available. The interested reader can read the survey by Madeira and Oliveira (2004) for a wider discussion on the topic. In our approach, instead, it is supposed that the set of data  $A$  is a training set, i.e.  $A$  is a set for which the classification of its samples is already known. The corresponding biclustering is therefore computed by employing a supervised technique, which is able to select the important features of the set. This information is then exploited for classifying samples having no known classification.

Let us suppose that  $A$  is a training set. Therefore, the classification of its samples in  $k$  classes is known. From this classification, it is possible to construct a classification of its features in  $k$  classes. The basic idea is to assign each feature to the class where it is mostly expressed (in other words, where it has higher value), in average, in the corresponding class of samples. Note that the same procedure can be inverted and it can be used for finding a classification of the samples of  $A$  from a known classification of its features. The reader can refer to Mucherino and Cafieri (2010) for more details about this supervised technique.

By combining the two classifications, the one for the samples of  $A$ , and the other one for the features of  $A$ , a biclustering can be defined for the matrix  $A$ . As already remarked, the supervised procedure mentioned above can construct classifications for the samples from classifications for the features, and vice versa. If the biclustering remains unchanged when the supervised procedure is applied, then it is said to be *consistent*. In other words, the biclustering is consistent if the classification of the samples (the features) suffices for

correctly reconstructing the biclustering (Busygin, Prokopyev, Pardalos, 2005).

In real life applications, however, training sets usually do not allow for consistent biclusterings. This is due to the fact that some features used for representing the samples could actually not be adequate for their representation. This is a common problem in data mining, because, at the time features are chosen for representing the samples, the knowledge about the problem is very limited. For this reason, in order to allow for consistent biclusterings, a subset of features must be removed from the training set. During this process, the number of rejected features must be as small as possible, in order to preserve the information contained in the dataset. This is done by solving a combinatorial optimization problem, where selected features are maximized while constraints ensure that the corresponding biclustering is consistent. The reader can find the formal definition of this optimization problem in (Busygin, Prokopyev, Pardalos, 2005).

In order to overcome issues related to noisy data and experimental errors, the concepts of  $\alpha$ -consistent and  $\beta$ -consistent biclustering have been introduced in (Nahapatyan, Busygin, Pardalos, 2008). In order to find  $\alpha$ -consistent or  $\beta$ -consistent biclusterings, the formulation of the optimization problem is slightly modified, and it allows for selecting the features that are relevant for the problem and, at the same time, the ones that should not be sensitive to noise and experimental errors.

Solving these three optimization problems (one for finding consistent biclusterings of training sets, one for  $\alpha$ -consistent biclusterings and another for  $\beta$ -consistent biclusterings) is NP-hard (Kundakcioglu and Pardalos, 2009). In this work, we find approximations of the solutions to these problems by employing the VNS-based heuristic proposed in (Mucherino, Cafieri, 2010).

#### 4. PERFORMING CLASSIFICATIONS WITH THE SELECTED FEATURES

We briefly described in the previous section a technique for selecting the features of  $A$  by finding a consistent biclustering of the training set. By exploiting this acquired knowledge, supervised classifications of the samples that do not belong to  $A$  can then be performed.

Let  $B$  be a dataset which is not a training set and that it is related to the same classification problem as the set  $A$ . No information regarding the classification of the samples in  $B$  is therefore available, but  $B$  contains the same features of  $A$  and a classification of these features is known because a biclustering for  $A$  is available. By using the supervised procedure, then, a classification for the samples of  $B$  can be found from the known classification of its features. Since the biclustering of  $A$  is consistent, the procedure is able to

find the correct classification for the samples in  $A$ , and therefore it should be able to do most likely the same for the samples in  $B$  (Busygin, Prokopyev, Pardalos, 2005).

The main issue concerning this technique is related to the fact that the biclustering associated to  $B$  is most likely not consistent. When this is the case, there are probably samples in  $B$  that are misclassified. In practice, samples in  $B$  may contain some additional information that  $A$  did not contain, causing in this way the loss of consistency.

We propose therefore a new strategy that is able to manage the case in which the biclustering associated to  $B$  is not consistent. In order to verify which samples in  $B$  might be misclassified, we build and check the consistency of the biclusterings associated to a set of matrices  $A_1, A_2, \dots, A_b$ , where  $b$  is the number of samples in  $B$  and each  $A_j$  is defined by adding the  $j^{\text{th}}$  sample to  $A$ . If the biclustering associated to  $A_j$  is consistent, then the  $j^{\text{th}}$  sample is probably well classified by the supervised procedure.

If the biclustering associated to  $A_j$  is instead not consistent, the obtained prediction could be not correct. In general, we cannot immediately declare a certain prediction wrong if the biclustering associated to  $A_j$  is not consistent. However, we can try to verify the quality of these classifications.

Let us suppose that the prediction for the  $j^{\text{th}}$  sample suggests that it belongs to class 1 and that  $A_j$  is not consistent. We can therefore try to make this biclustering consistent by removing some other features that were not unselected during the solution of the optimization problem. To this aim, a similar optimization problem can be solved, where the matrix  $A_j$  is considered and previously unselected features are directly discarded. It is important to note that, in the definition of the problem, a classification for the  $j^{\text{th}}$  sample must be provided.

One possibility is to give to the  $j^{\text{th}}$  sample the same classification that the procedure suggested. In this case, however, since the given classification for the  $j^{\text{th}}$  sample is exactly what the prediction gives, the biclustering may become easily consistent. This does not ensure that the  $j^{\text{th}}$  sample actually belongs to class 1.

The second possibility is to give to this sample the opposite classification (say class 2). We can then solve the optimization problem for removing some further features. Solutions to this problem can give clues about the actual classification for the sample.

If many features are removed from  $A_j$  in order to make its corresponding biclustering consistent, then the  $j^{\text{th}}$  sample probably belongs to class 1. The fact that many features are discarded indicates that assigning the sample to the other class goes against the original information in  $A$ . When instead few more features are

removed from  $A_j$  when solving this optimization problem, then the  $j^{th}$  sample probably belongs to class 2.

It follows naturally that, in case of 2-class classification problems, this strategy can be employed for attempting to perform completely exact classifications. We will show some experiments in next section where good-quality predictions have been performed by using this strategy.

## 5. COMPUTATIONAL EXPERIMENTS

We present in this section some experiments related to the dataset of wine fermentations presented in Section 2 and that contains 24 fermentations of Cabernet sauvignon from a winery in Chile's Maipo Valley (Urtubia, Perez-Correa, Meurens, 2004). We select the features that are relevant for discriminating between normal and problematic fermentations by finding consistent biclusterings of the training set. This is done by solving a combinatorial optimization with a VNS-based heuristic recently proposed in (Mucherino, Cafieri, 2010). Then, the classifications of the samples of the testing set are performed by using the obtained biclustering. We also consider the strategy we presented in Section 4 for performing predictions without any misclassifications.

Table 1 shows some experiments in which the combinatorial optimization problem has been solved in order to find  $\alpha$ -consistent biclusterings of  $A$ .  $f(x)$  is the objective function of this optimization problem, which is a counter a selected features, that must be maximized. As expected,  $f(x)$  decreases when the parameter  $\alpha$  increases, because features subject to a noise or to an error that is larger than  $\alpha$  are supposed to be removed from the set.  $err$  is the number of misclassifications on the testing set, when its samples are classified accordingly to the found  $\alpha$ -consistent biclusterings. The biclustering with the largest  $\alpha$  value is able to predict correctly 4 out of 8 samples. The corresponding biclustering related to the testing set is in fact not consistent.

Table 1:  $\alpha$ -consistent biclusterings obtained from our database of wine fermentations

$\alpha$	$f(x)$	$err$
0.00	448	5
1.00	445	5
1.50	442	4
1.70	440	4
2.00	438	4
2.10	431	4
2.20	431	4
2.30	402	4

In Table 2 we instead report some experiments that we performed for finding  $\beta$ -consistent biclusterings of  $A$ . As before,  $f(x)$  becomes smaller and smaller when larger values for the parameter  $\beta$  are considered. In these experiments, we are able to discard many features that are potentially wrong or noisy, and we are able to define a  $\beta$ -consistent biclustering formed by 141 features only, that are supposed to be the most relevant for classification purposes. However, even in this case, not all samples of the testing set are correctly classified: 3 out of 8 are assigned to the wrong class.

For each of the 8 samples of the testing set, we also constructed the 8 matrices  $A_j$ , obtained by adding a column representing the  $j^{th}$  sample of the testing set to the training set  $A$ . Only the 141 features selected in the  $\beta$ -consistent biclusterings with  $\beta = 1.80$  are considered.

As expected, the 5 matrices  $A_j$  that are related to the 5 samples that are correctly classified are all consistent. As a consequence, in case the classification of the samples of the testing set is not known, this information on the consistency of  $A_j$  can be exploited for validating the obtained predictions. The other 3 matrices are instead not consistent. We therefore assigned to the corresponding samples a classification which is the opposite of the one the supervised technique is able to provide. For two of the three matrices  $A_j$ , only one further feature has been removed in order to have the consistency of the biclustering associated to  $A_j$ . For one matrix, two features have been removed. For this reason, the actual classification of these three samples is probably the opposite of the one the procedure suggests. We verified the correctness of these results by removing these new features from the original training set  $A$  and by checking the consistency of the corresponding biclustering. All predictions in these experiments are correct.

Table 2:  $\beta$ -consistent biclusterings obtained from our database of wine fermentations

$\beta$	$f(x)$	$err$
1.00	448	5
1.20	397	5
1.30	340	5
1.40	281	4
1.50	262	3
1.60	211	3
1.70	147	3
1.80	141	3

## 6. CONCLUSIONS

We presented an analysis on a dataset of wine fermentations related to the harvest of 2002 of a Chilean winery. We considered a supervised biclustering technique that is based on the idea of constructing

consistent biclusterings in order to select the important features of a training set, and to exploit the obtained information for performing classifications for the samples of the testing set. We proposed in this paper a strategy for validating the results of the predictions, which may allow performing predictions with no mistakes on testing sets.

Future works will be mainly performed in the following two directions. First, larger datasets of wine fermentations need to be considered for obtaining better results. The fact that the considered testing set contains information which is not included in the training set suggests that it does not contain all necessary information for a correct definition of the biclusterings. Since industrial data are usually difficult to obtain, one possibility is to produce these data in laboratory, where small quantities of wine are fermented into a controlled environment. Moreover, we also plan to work on the formalization of the strategy that we proposed in this paper for validating the obtained classifications.

## REFERENCES

- Busygin, S., Prokopyev, O.A., Pardalos, P.M., 2005. *Feature Selection for Consistent Biclustering via Fractional 0-1 Programming*, Journal of Combinatorial Optimization 10, 7-21.
- Kundakcioglu, O.E., Pardalos, P.M., 2009. The Complexity of Feature Selection for Consistent Biclustering. In: S. Butenko, P.M. Pardalos, W.A. Chaovalitwongse (Eds.), *Clustering Challenges in Biological Networks*, World Scientific Publishing, 2009.
- Madeira, S.C., Oliveira, A.L., 2004. *Biclustering Algorithms for Biological Data Analysis: a Survey*, IEEE Transactions on Computational Biology and Bioinformatics 1 (1), 24-44.
- Mucherino, A., Cafieri, S., 2010. *A New Heuristic for Feature Selection by Consistent Biclustering*, arXiv e-print, arXiv:1003.3279v1.
- Mucherino, A., Papajorgji, P., Pardalos, P.M., 2009. *Data Mining in Agriculture*, Springer Optimization and Its Applications.
- Mucherino, A., Urtubia, A., 2010. *Consistent Biclustering and Applications to Agriculture*, IbaI Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10), Berlin, Germany, 105-113.
- Nahapatyan, A., Busygin, S., Pardalos, P.M., 2008. An Improved Heuristic for Consistent Biclustering Problems. In: R.P. Mondaini and P.M. Pardalos (Eds.), *Mathematical Modelling of Biosystems*, Applied Optimization 102, 185-198.
- Urtubia, A., Perez-Correa, J.R., Soto, A. Pszczolkowski, P., 2007. *Using Data Mining Techniques to Predict Industrial Wine Problem Fermentations*, Food Control 18, 1512-1517.

# DEMAND-SUPPLY INTERACTION AND INVENTORY BUILDUP STRATEGIES FOR SHORT LIFE CYCLE PRODUCTS

Rong Pan<sup>(a)</sup>, Adriano O. Solis<sup>(b)</sup>, Bixler Paul<sup>(c)</sup>

<sup>(a)</sup> School of Computing, Informatics, and Decision Systems Engineering, Arizona State University,  
Tempe, AZ 85287, USA

<sup>(b)</sup> Management Science Area, School of Administrative Studies, York University, Toronto, Ontario M3J 1P3, Canada

<sup>(c)</sup> Global Fulfillment, Dell, Inc., Austin, TX 78753, USA

<sup>(a)</sup> [rong.pan@asu.edu](mailto:rong.pan@asu.edu), <sup>(b)</sup> [asolis@yorku.ca](mailto:asolis@yorku.ca), <sup>(c)</sup> [bixler\\_paul@dell.com](mailto:bixler_paul@dell.com)

## ABSTRACT

Revenues and profits from short life-cycle products will depend upon careful formulation and execution of production plans in response to demands in the marketplace. This research contributes to the development of the production and inventory buildup strategies for short life-cycle products under different demand-supply scenarios. A modified Bass diffusion model is used to characterize the product demand pattern with consideration of demand-sales interaction. We develop cost models based on production costs, inventory carrying costs, backlog costs, and cost of lost sales for a number of different production scenarios. The optimal production rate can be obtained by minimizing the total cost. We also investigate the benefit of an initial buildup of inventory before the product's sales period starts.

## 1. INTRODUCTION

In today's highly competitive, global marketplace, supply chains must respond quickly to changes in customer requirements (Chopra and Meindl 2010). The impact of a product's life cycle on supply chain design and performance is of significant interest due to the uncertainty and risk involved (Kurawarwala and Matsuo 1996). Short life-cycle products, e.g., personal computers and other consumer electronics, are generally characterized by constant innovation. Revenues and profits generated from new products would hinge on the careful formulation and implementation of production plans in response to demand in the marketplace. To achieve cost minimization/profit maximization, production planning requires accurate demand forecasting, detailed analysis of demand signals, and careful consideration of the dynamics of demand-sales interactions.

Short life-cycles are typically encountered in two kinds of products: innovative products such as electronic goods and fashion goods which have a seasonal demand. Short life-cycle products have distinct characteristics such as capricious demand patterns, high rate of obsolescence, risky capacity

decisions, and high levels of uncertainty at all levels of operations. Judging a customer's desire to own or buy is often highly unpredictable. Adding to the forecasting uncertainties are the rapid market diffusion in an Internet-connected world and severe competition which brings in newer technologies and speed up the rate of obsolescence. The growth stage in the short life-cycle poses many interesting questions. Due to "panic" growth, management invariably faces the issue of production capacity expansion. In a case study of the *Tamagotchi*<sup>TM</sup> 'virtual pet' toy, Higuchi and Troutt (2004) discussed how an electronic toy company went from boom to bust in a total period of 24 months due to an imprudent expansion decision.

The current study delves into the mechanics of a short life-cycle product's diffusion patterns and their implications on supply chain design. We seek to develop generalized cost models to better explain the demand-supply interaction that occurs in short life-cycle products and its impact on production capacity planning. Integrated demand-supply interaction and production graphs are utilized to visualize the times and volumes of inventory, backlogs, and other components that generate costs during product's entire life cycle. The optimal production capacity planning responding to a particular demand pattern can be found after we build the cost models to analyze each demand-sales interaction circumstance.

The remainder of the paper is organized as follows. In Section 2, a demand-sales model based upon the well-known Bass diffusion model (Bass 1969) is introduced and its interaction with production rate will be discussed. The cost minimization/profit maximization models based on inventory, backlog, lost sales, and production quantities for the different production scenarios over the product's life cycle will be derived in Section 3, followed by a discussion of the strategy of inventory build-up at the beginning of product's life cycle in Section 4. Finally, we conclude and remark on our findings in Section 5.



## 2. BASS DIFFUSION MODEL

The Bass diffusion model (Bass 1969) posits that the instantaneous rate of adoption of a new product by the population of potential adopters at any time period is subject to two means of communications: mass-media (external) and word-of-mouth (internal). The external communication influences ‘innovators’, while the internal communication describes the interaction between innovators and ‘imitators’.

Let  $p$ , the “coefficient of innovation,” and  $q$ , the “coefficient of imitation,” represent the extent of external and internal communication levels, respectively. Let  $m$  be the size of the target population and let  $D(t)$  be the cumulative number of adopters of a new product by time  $t$ . Under the assumption that  $D(t)$  is a continuous function with  $D(0) = 0$ , then

$$\frac{dD(t)}{dt} = \left[ p + q \frac{D(t)}{m} \right] [m - D(t)], \quad t \geq 0. \quad (1)$$

That is, the growth rate of  $D(t)$  at time  $t$  is equal to the product of  $m - D(t)$ , the size of the remaining population, and  $\left[ p + q \frac{D(t)}{m} \right]$ , the instantaneous

adoption rate of an individual in the remaining population. Initial purchases of the product are made by both ‘innovators’ and ‘imitators’. Innovators are not influenced in the timing of their initial purchase by the number of people who have already bought the product, while imitators are influenced by the number of previous buyers. Imitators ‘learn’, in some sense, from those who have already bought. The solution to the differential equation (1) is given by

$$D(t) = \left[ 1 - e^{-(p+q)t} \right] / \left[ 1 + (q/p) e^{-(p+q)t} \right]. \quad (2)$$

This function has been found to provide an excellent empirical fit for the timing of initial purchase for a wide range of consumer durables (Niu 2004). The derivative of cumulative demand is then given by

$$d(t) = mp(p+q)^2 e^{-(p+q)t} / \left[ p + q e^{-(p+q)t} \right]^2, \quad (3)$$

which gives instantaneous demand at time  $t$ . Figure 1 is a sample graph of both instantaneous demand and cumulative demand, as specified by equations (2) and (3) respectively, with parameter values  $m = 3000$ ,  $p = 0.03$ , and  $q = 0.4$ .

The original Bass model does not consider the production capacity issue, an important aspect in supply chain design (Ho, Sergei, and Terwiesch 2002). In a supply chain, there always exists a maximal production rate as defined by the capacity of the plant, which can adversely affect the product diffusion rate (Jain, Mahajan, and Muller 1991). An important modification to the Bass model was developed by Kumar and Swaminathan (2003), where they propose that, in light of the rapid growth of demand, it is possible that a large

number of consumers may attempt to buy the product but will be unsuccessful due to supply constraints. It is unreasonable to assume that these customers would continue to spread word about the product. The word-of-mouth effect then is better represented as being proportional to the cumulative sales  $S(t)$  up to time  $t$ , instead of the cumulative demand  $D(t)$ . Then the Bass model can be modified to

$$\frac{dD(t)}{dt} = \left[ p + q \frac{S(t)}{m} \right] [m - D(t)]. \quad (4)$$

Figure 2 illustrates the modification in the cumulative demand curve.

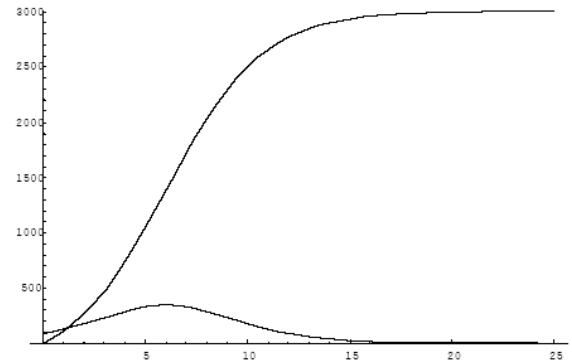


Figure 1: Instantaneous and Cumulative Demand Curves with  $m = 3000$ ,  $p = 0.03$ , and  $q = 0.4$

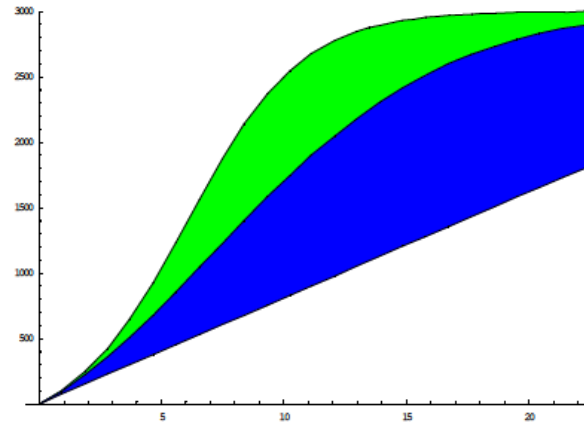


Figure 2: Change of Demand Curve due to Supply Constraint

The key feature of this modified Bass model is that the future demand depends not only on past demand but also on past realized sales (Kumar and Swaminathan 2003). The quantity of sales at a given time is determined by the demand as well as the production volume.

To evaluate the interaction between sales and demand, Ho, Savin, and Terwiesch (2002) looked into the Bass model with supply constraints and found that a

myopic sales plan is always optimal given an objective of profit maximization rather than minimization of lost sales. Angelus and Porteus (2002) developed a model to better understand the interaction between capacity and production management. They derived an optimal simultaneous capacity and production plan for a short life-cycle, make-to-stock good under stochastic demand.

There are four basic scenarios where the production plan is matched against the demand profile:

- *Scenario 1.* Production plan is barely satisfying the demand generated from the beginning of the product's life cycle.
- *Scenario 2.* Production rate is satisfying some demand (till a period of time when the demand overtakes the production) with inventory but total demand cannot be met by total production quantity.
- *Scenario 3.* Production completely satisfies the total market capacity and is terminated when the total production quantity reaches the total demand quantity.
- *Scenario 4.* Production rate is so high that it reaches market capacity much ahead of actual complete market consumption, that it carries inventory to satisfy the demand.

Based on the above scenarios, we develop various cost functions in order to establish the optimal production plan for a given set of parameters which govern the demand profile.

### 3. PRODUCTION CAPACITY OPTIMIZATION

It is assumed that a product's life cycle begins at the time when sales occur and it ends when the sales reach close to the total market size. The production may start much earlier than the beginning of product's life cycle to build up inventory for the anticipated rapid-growing sales. Combining the Bass demand curve (or its modification) and production curve we formulate cost functions over the entire product lifetime and seek to minimize the total cost. The various costs that are considered here include inventory carrying costs, backlog costs, production costs and cost of lost sales. In addition, we associate a discount factor to make the model more realistic. The discount factor is used to analyze the present value of money over the period of the product's life cycle.

#### 3.1. Deterministic Demand-Production Model Development

We start from the simplest demand-production scenario where the demand is driven by a Bass diffusion model and the production rate is kept constant. There are several assumptions we make to best explain the conditions within which the cost models are built:

1. *Loss of sales occurs only at the end of the life cycle.* It is assumed that the demand fulfillment

is executed using a first-in, first-out (FIFO) policy and loss of sales are attributed to those customers who are in the queue to be satisfied but, due to the end of production and non-availability of the product, the manufacturer would have to refuse the product from those customers.

2. *Demand-sales interaction does not occur at any instant of the product lifetime.* Here, the demand curve is unaffected by the production rate; this assumption will be removed later.
3. *No inventory buildup.* It is assumed that the sales and production would start simultaneously at the beginning of product's life cycle. This implies that there will be no inventory at the beginning of the product's sales process. Again, we will consider the inventory buildup strategy later.
4. *Company employs a make-to-stock production policy.* With a constant production rate, the residual inventory at any time will be held to satisfy the anticipated future demand.
5. *Monopolistic market state.* As clearly mentioned in Bass (1969), we consider that the product is in a monopolistic market condition where the manufacturer has no competitor or competing product in the same category. In a competitive market environment, the parameters which govern the adoption rates are many more than the ones explained here in this paper and they are also susceptible to qualitative measures which govern the success of the product (e.g., value, design, form, fit, function, etc.).

The modified Bass model which we use in this paper has three parameters which govern the shape of the demand pattern. The cumulative demand is as specified in (4). Production rates govern the extent of sales-demand interaction. Holding the given set of parameters  $m$ ,  $p$ , and  $q$  constant in the demand function, higher production would typically entail more inventory which may fully satisfy market demand. On the contrary, however, if we have a production plan which cannot fulfill total demand of the market, then after a partial fulfillment of the demand, the sales would typically follow the output of the production.

In Figure 3 both demand and production curves are plotted. The demand curve asymptotically approaches  $m$ , and is arbitrarily cut off at 99.9% of this total market potential, in effect signaling the end of the product's life cycle. For a given demand function (based upon values of  $m$ ,  $p$ , and  $q$ ) there are three special production lines:

- Line  $y_1$  - This line passes through the origin and its slope is given by

$$b_1 = mp. \quad (5)$$

- Line  $y_2$  – The production line intersects the demand curve at the end of the product's life cycle. The production here totally satisfies the demand of the market with no inventory and no backlogs at the end of the life cycle. The slope of  $y_2$  is given by

$$b_2 = 0.999m/T. \quad (6)$$

- Line  $y_3$  – This line is tangent to the cumulative demand curve. The slope of the line is

$$b_3 = mp(p+q)^2 e^{-(p+q)T} / \left[ p + qe^{-(p+q)T} \right]^2. \quad (7)$$

There are four regions,  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ , separated by these special lines:

- Region  $R_1$  – Production plan never satisfies the demand generated from the beginning of the product's life cycle (*Case 1*).
- Region  $R_2$  – Production rate is satisfying some demand with inventory (until a period of time when the demand overtakes the production) but total demand cannot be met by total production quantity (*Case 2*).
- Region  $R_3$  – Production completely satisfies the total market capacity and production is terminated when the total production quantity reaches the total demand quantity (*Case 3*).
- Region  $R_4$  – Production rate is so high that it reaches market capacity much ahead of actual complete market consumption, that it carries inventory to satisfy the demand (*Case 4*).

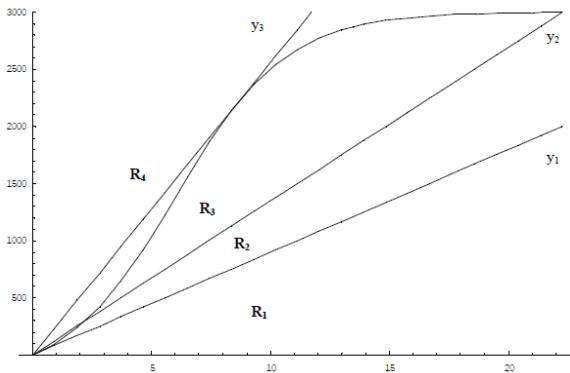


Figure 3: Cumulative Demand Curve Shown with the Different Production Plans

### 3.1.1. Cost Functions of Region $R_1$

In this case (Figure 4), the production rate is anywhere between zero and  $b_1$ . Total production can never fully satisfy cumulative demand at any point in time. All the unmet demand is backlogged. At the end of the product's life cycle, there would be some demand which would be unmet due to the limited production and be eventually lost. The cut-off time at the threshold

limit of 99.9 % of the market capacity is the solution of the equation

$$0.999\left(1 + \frac{q}{p}e^{-(p+q)T}\right) = 1 - e^{-(p+q)T}. \quad (8)$$

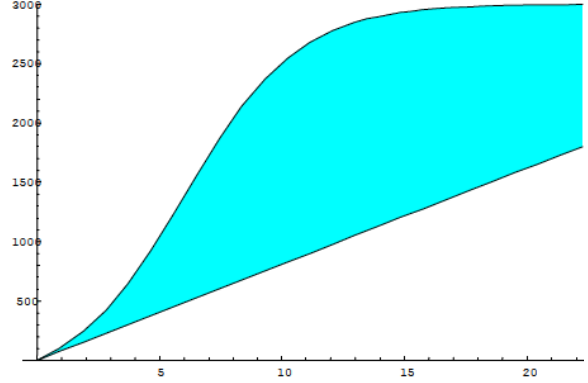


Figure 4: Case 1 – Production Never Meets Total Demand with  $m = 3000$ ,  $p = 0.03$  and  $q = 0.4$

Both production volume and total sales volume at time  $t$  are given by the production curve  $y = bt$ , where  $b$  is the production rate. There is no inventory built up in this scenario. The backlog volume at time  $t$  is the difference between demand curve and production curve, i.e.,  $\left[ m\left(1 - e^{-(p+q)t}\right) / \left(1 + \frac{q}{p}e^{-(p+q)t}\right) - bt \right]$ . To convert the cost generated at time  $t$  to the value as at the beginning of the product's life cycle, the discount rate can be written as  $(1 + \gamma)^{-t}$ , where  $\gamma$  denotes the rate of return. Since we are dealing with short life-cycle products,  $\gamma$  should be very small and we use the approximation  $(1 + \gamma)^{-t} \approx (1 - \gamma)^t \approx e^{-\gamma t}$ .

Thus, the backlog cost can be evaluated by the area between two curves,

$$CL = \int_0^T \omega \left[ m\left(1 - e^{-(p+q)t}\right) / \left(1 + \frac{q}{p}e^{-(p+q)t}\right) - bt \right] e^{-\gamma t} dt. \quad (9)$$

where  $\omega$  is the unit backlog cost.

The volume of lost sales over the product's entire life cycle is the difference between cumulative demand and production curves at the production termination time. Hence, lost sales cost is

$$CS = \chi \left[ m\left(1 - e^{-(p+q)T}\right) / \left(1 + \frac{q}{p}e^{-(p+q)T}\right) - bT \right] (1 - \gamma)^T, \quad (10)$$

where  $\chi$  represents the unit lost sales cost.

We assume the production cost is proportional to the total production volume, with

$$CP = \int_0^T \alpha b t e^{-\gamma t} dt \quad (11)$$

where  $\alpha$  is the unit production cost.

Therefore, in this case the total cost is the sum of total production cost, cost of lost sales and the total backlog costs, i.e.,  $TC = CP + CL + CS$ .

### 3.1.2. Cost Functions of Region $R_2$

In this case, the production rate  $b$  is higher than  $b_1$  but less than  $b_2$ . Any generic line that has a slope between these two special lines intersects the demand curve at a point  $t_0$ , which differentiates the inventory and backlog that the production plan creates. The point  $t_0$  can be determined using

$$t_0 = \log[-(mp + bTq)/(bT - m)p]/(p + q). \quad (12)$$

A small inventory is developed at the early stage of product's life cycle and the backlog will occur once the production curve falls below the demand curve. The cumulative inventory is represented by the area between production curve and demand curve, when production is larger than demand. The total inventory cost is

$$CI = \int_0^{t_0} h \left\{ bt - \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) \right] \right\} e^{-\gamma t} dt, \quad (13)$$

where  $h$  is the unit inventory carrying cost. The total backlog cost is

$$CL = \int_{t_0}^T \omega \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) - bt \right] e^{-\gamma t} dt. \quad (14)$$

The cost of lost sales would be as in (10).

### 3.1.3. Cost Functions of Region $R_3$

In this case, the production rate  $b$  is between  $b_2$  and  $b_3$ . The production is able to produce enough to sustain till the end of the product's life cycle with little backlog, and the production stops when the total quantity reaches  $m$ . There are no lost sales registered. The production line intersects the demand curve at two points and terminates at a third point before the end of product's life cycle, as illustrated in Figure 5. We denote the three points as  $t_1$ ,  $t_2$ , and  $t_p$ . There are three regions  $I_1$ ,  $I_2$ , and  $I_3$  where inventory is carried. Carrying costs of those inventories are determined separately to constitute total inventory cost:

$$CI_1 = h \int_0^{t_1} \left\{ bt - \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) \right] \right\} e^{-\gamma t} dt, \quad (15)$$

$$CI_2 = h \int_{t_1}^{t_p} \left\{ bt - \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) \right] \right\} e^{-\gamma t} dt, \quad (16)$$

$$CI_3 = h \int_{t_p}^T \left\{ 0.999m - \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) \right] \right\} e^{-\gamma t} dt. \quad (17)$$

The backlog cost is given by

$$CL = \int_{t_2}^{t_3} \omega \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) - bt \right] e^{-\gamma t} dt. \quad (18)$$

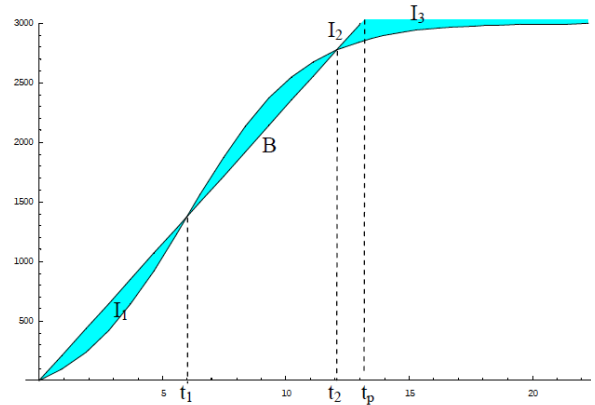


Figure 5: Case 3 – Inventory and Backlog

### 3.1.4. Cost Functions of Region $R_4$

In this case, as shown in Figure 6, production is always enough to cover demand until the end of the product's life cycle. There would be no backlog and lost sales costs, but a large inventory cost is associated with this case.

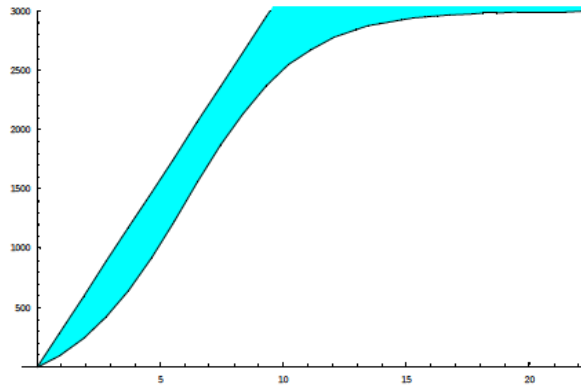


Figure 6: Case 4 – More than Enough Production

The inventory costs associated with this case comes from two regions. It can be determined by

$$CI = h \int_0^{t_p} \left\{ bt - \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) \right] \right\} e^{-\gamma t} dt + h \int_{t_p}^T \left\{ 0.999m - \left[ m(1 - e^{-(p+q)t}) / \left( 1 + \frac{q}{p} e^{-(p+q)t} \right) \right] \right\} e^{-\gamma t} dt. \quad (19)$$

So far, we have developed the cost functions under different production plans. One can find the optimal production rate by minimizing the total cost over the product's life cycle subject to any specified inventory or backlog constraints.

### 3.2. Analysis Based on Demand-Supply Interaction

In practical situations, it is unrealistic to have a demand model which is not supported by the production plan and still expect the demand to follow the same growth pattern. With the modified Bass model, the word-of-mouth effect is constrained by the quantity of sales at a given time. For example, in the first case where the production is so low that it will never satisfy the demand, Figure 7 gives the potential demand curve and the actual demand curve after considering the demand-sales interaction. Here, the sales is limited by production and the actual demand is determined by solving equation (4) with  $S(t) = b$  and the initial condition  $D(0) = 0$ ; hence,

$$D(t) = m(1 - e^{-pt - \frac{bqt^2}{2m}}). \quad (20)$$

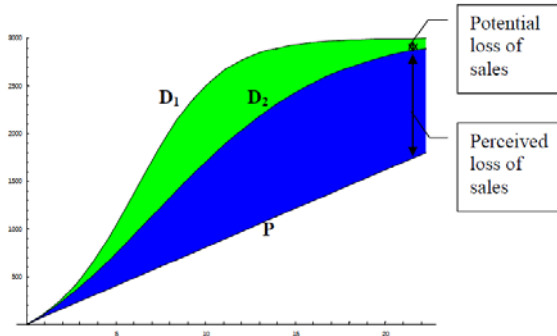


Figure 7: Case 1 – Demand-Sales Interaction

In this case, we have two demand curves: an assumed demand curve ( $D_1$ ), which is the total potential of the market, and the actual demand curve which is the outcome of the demand-sales interaction due to the production plan ( $D_2$ ). We define the perceived loss of sales as the total number of unfulfilled demand, which is the difference between actual demand and the corresponding production plan at the end of product's life cycle. We also realize that there is a potential loss of sales, which is the difference between the total potential of the market and the actual total demand. We call it potential because, for various reasons, there are potential customers in the market who have never heard about the product. Again the cost function for these losses of sales can be easily derived.

We developed cost functions for the cases as discussed previously. The optimal production rate can be sought by minimizing the total cost over the product's life cycle.

### 3.3. Analysis Based on Initial Inventory Buildup Strategy

Due to the capacity constraints and expectations of a higher than normal demand, firms build up enough inventories prior to the launch of the product. This helps prepare the firms for the steep increase in demand once the product is introduced into the market. There are two critical issues which need to be addressed in this strategy: First, how early should we start the inventory build-up (time)? And second, how much should we build-up for sales (quantity)? They can be answered after optimizing the total profit generated over the whole life cycle.

### 4. ILLUSTRATIVE NUMERICAL EXAMPLES

The examples presented in this section pertain to Case 3, in which the production plan completely satisfies the total market capacity and production is terminated when the total production quantity reaches the total demand quantity.

Figure 8 shows the discounted profits versus production rate plots.

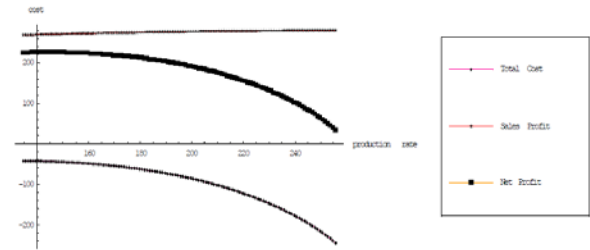


Figure 8: Case 3 – Discount Graph with  $m = 3000$ ,  $p = 0.03$  and  $q = 0.4$

The various costs and sales revenue are each graphed, in Figures 9-13, as a function of production rate and inventory buildup.

We are then able to establish appropriate production plans for a given set of values of the parameters  $m$ ,  $p$ , and  $q$ . We could also analyze different production plans based on the cost functions associated with each of the cases to establish the optimum strategy (in terms of production rates) to respond to the demand in the market.

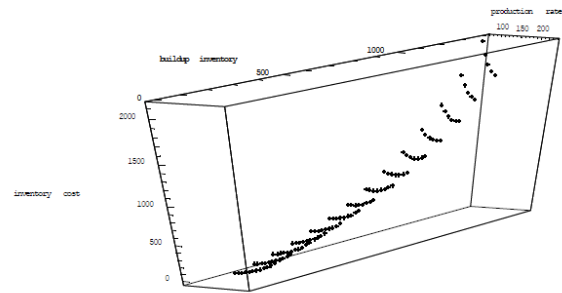


Figure 9: Case 3 – 3D Plot for Inventory Cost, Buildup Inventory, and Production Rate

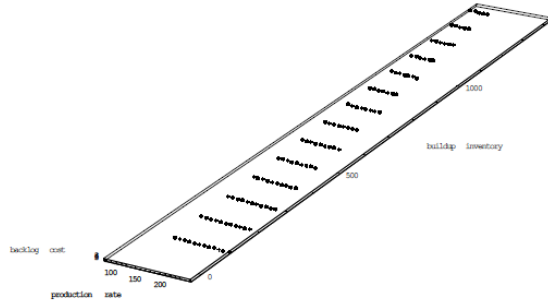


Figure 10: Case 3 – 3D Plot for Backlog Cost, Buildup Inventory, and Production Rate

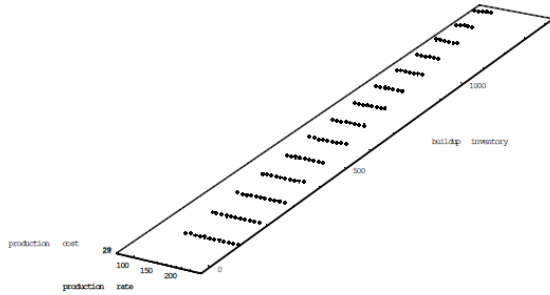


Figure 11: Case 3 – 3D Plot for Production Cost, Buildup Inventory, and Production Rate

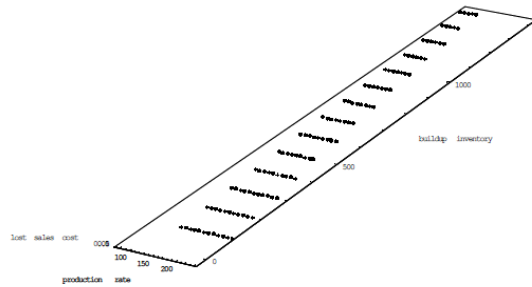


Figure 12: Case 3 – 3D Plot for Lost Sales Cost, Buildup Inventory, and Production Rate

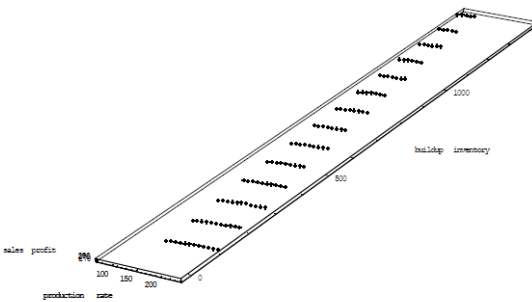


Figure 13: Case 3 – 3D Plot for Sales Revenue, Buildup Inventory, and Production Rate

## 5. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we develop deterministic models to study the effect of production capacity on the demand and sales behaviors in products with short life-cycles. Cost

functions over the entire product's life cycle are built and they can be utilized for finding the optimal production rate at the beginning of the life cycle. We also study the effects of an inventory build-up strategy to satisfy the sharp increase of demand at the growing stage of life cycle.

The contribution of this research is different from previously published papers (Ho, Savin, and Terwiesch 2002; Kumar and Swaminathan 2003). Our approach is more graphical in nature and the cost models associated with the different scenarios developed in Section 3 aid in developing an inventory buildup strategy. From several numerical/simulation studies (for brevity, those studies are not included in this report), we draw some observations.

### 5.1. Effect of Parameters $m, p, q$ on Production Plan

We note that changing  $m$  (size of the target market) does not have much of an influence on the shape of the demand curve. However, the shape of the demand curve can drastically vary with a change in  $p$  (the coefficient of innovation) or  $q$  (the coefficient of imitation). As  $p$  increases, the curve tends to grow sharply which indicates that the product reaches a major percentage of the market capacity in a short period of time. This typically entails a scenario in the market where consumers are trying the new product without the word-of-mouth communication (this is typically applicable when there is plenty of advertisement in the news/entertainment media).

An increase in  $q$  (coefficient of imitation) would tend to boost the sales by generating consumers who are waiting and watching the product from the time it was introduced into the market.

### 5.2. Inventory Buildup Strategy

In the inventory buildup strategy, it is very important to establish the start time of production which would decide the total inventory quantity prior to the beginning of the sales period. Another important factor is the cost coefficient associated with the holding of the inventory at the beginning of the product's life cycle. As seen in Section 3, we have developed cost functions for the different scenarios presented.

### 5.3. Future Research

In this paper, we arbitrarily chose the end of the product's life cycle to be at the time demand reaches 99.9% of total market size. The production period is determined accordingly. However, further research could be done by finding the optimal cut-off point for a specific production plan. Moreover, only a constant production rate is discussed in this paper. Clearly choosing a proper time for production expansion or contraction to better match the demand curve will provide a better payoff. It would be very useful for management to have an interactive graphical tool, which will show cost components under different production plans. Finally, demand-supply interaction modeling with stochastic demands is yet to be



researched and it will be considered in our future research.

## REFERENCES

- Angelus, A. and Porteus, E., 2002. Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand. *Management Science*, 48 (3), 399-413.
- Bass, F.M., 1969. A new product growth for model consumer durables. *Management Science*, 15 (5), 215-227.
- Chopra, S., and Meindl, P., 2010. *Supply Chain Management: Strategy, Planning, and Operation*, 4<sup>th</sup> edition, Upper Saddle River, New Jersey, USA: Prentice Hall.
- Higuchi, T. and Troutt, M.D., 2004. Dynamic simulation of the supply chain for a short life cycle product—Lessons from the Tamagotchi case. *Computers and Operations Research*, 31 (7), 1097-1114.
- Ho, T-H., Savin, S., and Terwiesch C., 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science*, 48 (2), 187-206.
- Jain, D., Mahajan, V., and Muller, E., 1991. Innovation diffusion in the presence of supply restrictions, *Marketing Science*, 10 (1), 83-90.
- Kumar, S. and Swaminathan, J.M., 2003. Diffusion of innovations under supply constraints, *Operations Research*, 51 (6), 866-879.
- Kurawarwala, A.A. and Matsuo, H., 1996. Forecasting and inventory management of short life-cycle products. *Operations Research*, 44 (1), 131-150.
- Liu, B., Chen, J., Liu, S., and Zhang, R., 2006. Supply-chain coordination with combined contract for a short-life-cycle product, *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 36 (1), 53-61.
- Meade, N. and Islam, T., 2006. Modelling and forecasting the diffusion of innovation—A 25-year review, *International Journal of Forecasting*, 22 (3), 519-545.
- Niu, S-C., 2002. A stochastic formulation of the Bass model of new-product diffusion. *Mathematical Problems in Engineering*, 8 (3), 249-263.
- Silver, E.A., Pyke, D.F., and Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*. New York, USA: John Wiley & Sons.
- Weng, Z.K., 1999. The power of coordinated decisions for short-life-cycle products in a manufacturing and distribution supply chain, *IIE Transactions*, 31 (11), 1037-1049.
- Weng, Z.K. and McClurg, T., 2003. Coordinated ordering decisions for short life cycle products with uncertainty in delivery time and demand, *European Journal of Operational Research*, 151 (1), 12-24.
- Weng, Z.K. and Parlar, M., 2005. Managing build-to-order short life-cycle products: benefits of pre-

season price incentives with standardization, *Journal of Operations Management*, 23 (5), 482-495.

## AUTHORS' BIOGRAPHIES

**Rong Pan** is an Associate Professor of Industrial Engineering in the Ira A. Fulton School of Engineering at Arizona State University, USA. He received his PhD degree in Industrial Engineering from Pennsylvania State University after completing a BS in Materials Science and Engineering from Shanghai Jiao Tong University and an MS in Industrial Engineering from Florida A&M University. He was previously Assistant Professor of Industrial Engineering at the University of Texas at El Paso. He has published in *International Journal of Production Research*, *Journal of Applied Statistics*, *Journal of Quality Technology*, *Quality and Reliability Engineering International*, and other scholarly publications.

**Adriano O. Solis** is an Associate Professor of Logistics Management and Management Science at York University, Canada. After receiving BS, MS and MBA degrees from the University of the Philippines, he joined the Philippine operations of Philips Electronics where he became a Vice-President and Division Manager. He later received a PhD degree in Management Science from the University of Alabama. He was previously Associate Professor of Operations and Supply Chain Management at the University of Texas at El Paso. He has published in *European Journal of Operational Research*, *International Journal of Simulation and Process Modelling*, *Information Systems Management*, *International Journal of Production Economics*, *Computers & Operations Research*, and *Journal of the Operational Research Society*, among others.

**Bixler Paul** is currently Senior Global Program Manager responsible for strategy and program management for retail fulfillment globally at Dell, Inc. He received a Bachelor of Engineering degree in Industrial Engineering, graduating with First Class with distinction honors from Bangalore University in India, and an MS in Industrial Engineering from the University of Texas at El Paso.

# Evaluation of the performance of solar air collector by using Bond Graph approach

H. Oueslati <sup>(1,2\*)</sup>, S. Ben Mabrouk <sup>(1)</sup>, A. Mami <sup>(2)</sup>

<sup>(1)</sup> Laboratoire d'Energétique et Procédés Thermiques (LEPT), Centre de Recherches et des Technologies de l'Energie (CRTE n) ; BP: 95, Hammam-Life 2050. Tunisie.

<sup>(2)</sup> Laboratoire Analyse et Commande des Systèmes, École Nationale d'Ingénieurs de Tunis, BP 37, 1002 Tunis, Tunisie.

E-mail addresses: [Houeslati@gmail.com](mailto:Houeslati@gmail.com) (H. Oueslati), [Salah.Benmabrouk@crtcn.rnrt.tn](mailto:Salah.Benmabrouk@crtcn.rnrt.tn) (S. Ben Mabrouk), [abdelkader.mami@fst.rnu.tn](mailto:abdelkader.mami@fst.rnu.tn) (A. Mami).

## Abstract

In this paper, the modelling and thermal performance results of the solar air collector have been presented. This system is conceived to heat the ambient air in order to be used in many domains like drying processes and heating buildings. The pseudo-bond graph methodology was used in modelling this system. Such methodology was very suitable for this thermodynamic process since it allows good management of the non-linearity present in the system.

The simulation of the global model with bond graph software [20-sim] allows us to analyze the temperature variations and the efficiency of the solar collector and compared with experiment results.

*Keywords:* Solar air collector; Temperature; Collector efficiency; Pseudo bond graph modelling.

## 1. Introduction

It exists not only several manners to supply of the solar energy, but equally different methods to captivate the solar energy proceeding from an incident radiation.

The solar collector absorbs the solar radiation and transforms him in heat transmitted to a fluid (air) who will be used for else applications (Gao W, Lin W, Liu T and Xia C 2007; Grag H.P and Kumar R 2000; Duffie J.A and Beckman W.A 1991; Dickson W.C and Chermisinoff P.N 1980).

In this paper we analysed thermal performances of the flat plate solar collector based on the bond graph approach. This technique was developed as an appropriate tool for modelling all types of the engineering systems. It is based on the topologic representation of different mechanisms regarding exchange, storage and energy dissipation in any thermodynamic system (Karnopp D and Rosenberg R. 1983). For thermal and chemical engineering, the classic bond graph behaviours introduce complex variables likes entropy and chemical potential, which do not present simple conservation laws. Therefore, we used the pseudo-bond graph methodology where the product effort and flow is not a power. Some research

works have been published in this area (Karnopp D, Margolis D and Rosenberg R 1990; Thoma J. and Ould Bouamama B 2000; Cellier F. E 1991).

## 2. System description

The solar air collector is used as a generator of hot air. It is a flat-plate collector with simple glazing and classical absorber (metallic undulate plate carrying chicaneries). The air passes through the flat-plate collector between the absorber plate and the glass cover as shown in Figure 1.

The solar air collector is sensitive of two external parameters: solar radiation and wind velocity. The recovered heat proceeds essentially from thermal convection exchanges between the air, the absorber plate and the glass cover.

The studied device admits following assumptions:

- The solar air collector is placed in north-southern direction with inclination angle 37° and exposed to the solar radiation throughout the day.
- The air flow is unidirectional and mass flow is considered as constant along the channel of the solar air collector.
- Radiation heat transfer between all the absorber plate and the insulation, the insulation and the soil are assumed negligible.
- The system studied is considered with lumped parameters, that has an evenly distributed temperature.

Different heats transfers took in consideration are well described in Figure 1.

## 3. BOND GRAPH MODELLING IN PROCESS ENGINEERING

The dynamic behavior of the thermo fluid processes is generally described by the non linear differential equations.

Their formulation and resolution by the classic methods are limited (Hadim H and Vafai K 1999). With the bond graph methodology, these equations are associated to the storage and the dissipation of energy.

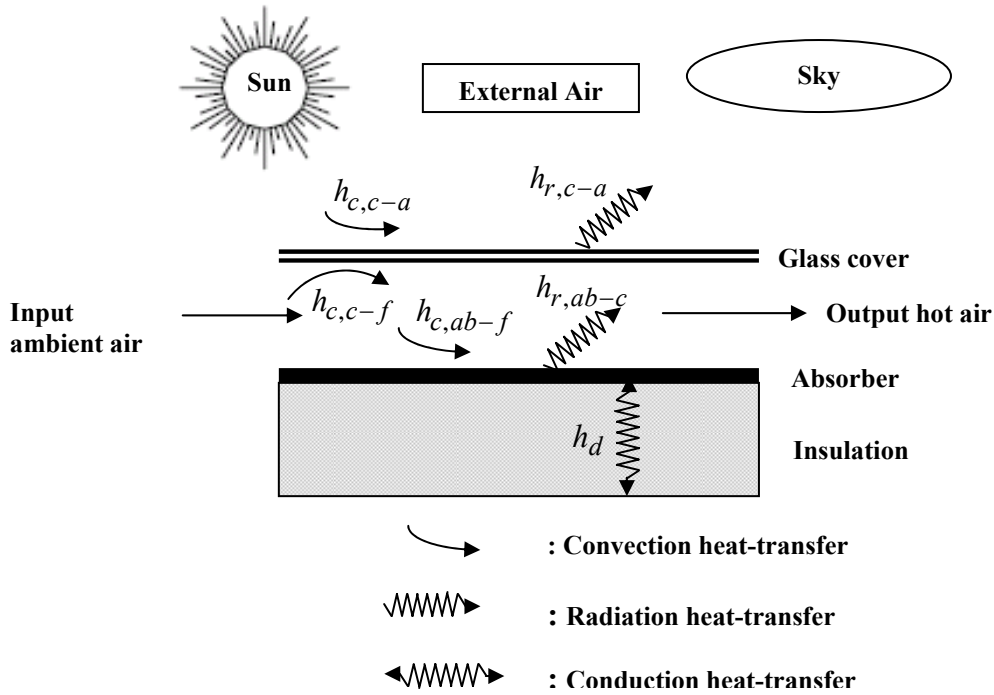


Figure 1: Schematic sketch of cross-sectional view along the flow direction of the solar collector.

Therefore, the bond graph tools permit by its graphic description to explicit the power exchanges in the system, as the energy storage and the thermal dissipation.

The bond graph method was developed as an appropriate tool for modelling all types of the engineering systems. It is based on the topologic representation of different mechanisms regarding exchange, storage and energy dissipation in any thermodynamic system (Karnopp D and Rosenberg R 1983).

With this methodology, the parameter energy is represented through two kinds of general variables: the power variables (effort and flow) and the energy variables (momentum and displacement). In thermo fluid process, thermal and hydraulic energies are coupled (Thoma J and Ould Bouamama B. 2000; Ould Bouamama B 2003). Their coupling can be represented by small rings around the bond as shown in Figure 2.



Figure 2: Bond graph representation:  
e: (effort) - f: ( flow)

### 3.1. Word pseudo bond graph

The technological level of modelling can be represented by word pseudo bond graph model. In this step, one splits the entire system into simple subsystem as shown on Figure 3 where at the entry and the release of each subsystem we have already used the liaison variables corresponding, and this will be based on an energizing description of the process.

### 3.2. Pseudo bond graph model

If the words of the word pseudo-bond graph shown in Figure 3 are replaced by the corresponding elements, we obtain the pseudo-bond graph shown in Figure 4.

This model can be represented by two sub-models, the red corresponds to the thermal subsystem and the blue to the hydraulic subsystem.

The thermal sub-model has temperature ( $T$ ) as one effort variable, and heat flow ( $\dot{Q}$ ), enthalpy flow ( $\dot{H}$ ) and solar flow ( $I$ ) as many flow variables. On the other hand the hydraulic sub-model has pressure ( $P$ ) as effort variable, and mass flow ( $\dot{m}$ ) as flow variable.

The pseudo bond graph model contains seven elements: capacitive elements (**C**), resistive elements (**R**), effort sources (**Se**), flow sources (**Sf**), modulated or controlled flow source (**MSf**), **0** – junctions and **1** – junctions.

**C** and **R** elements are passive elements because they convert the supplied energy into stored or dissipated energy. **Se**, **Sf** and **MSf** elements are active elements because they supply power to the system and **0**- and **1**-junctions are junction elements that serve to connect **C**,

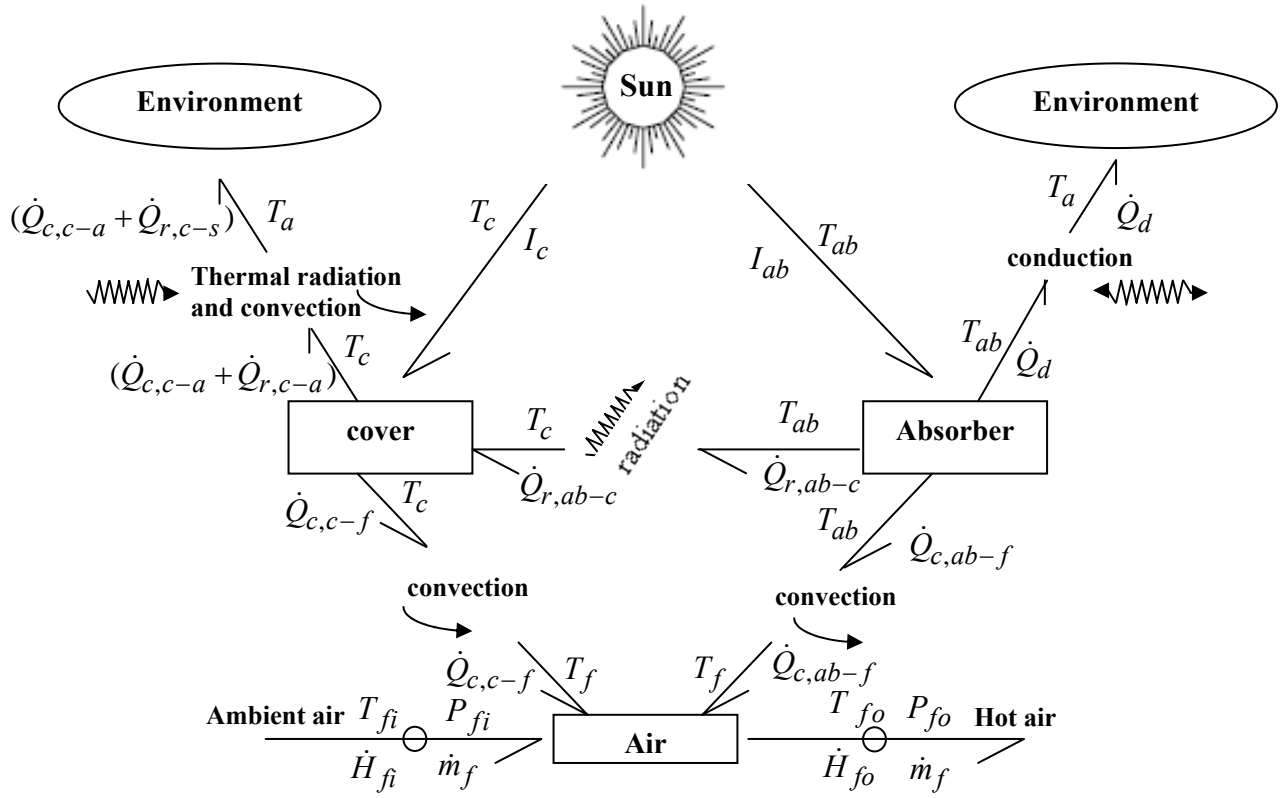


Figure 3: Word bond graph model of the solar air collector.

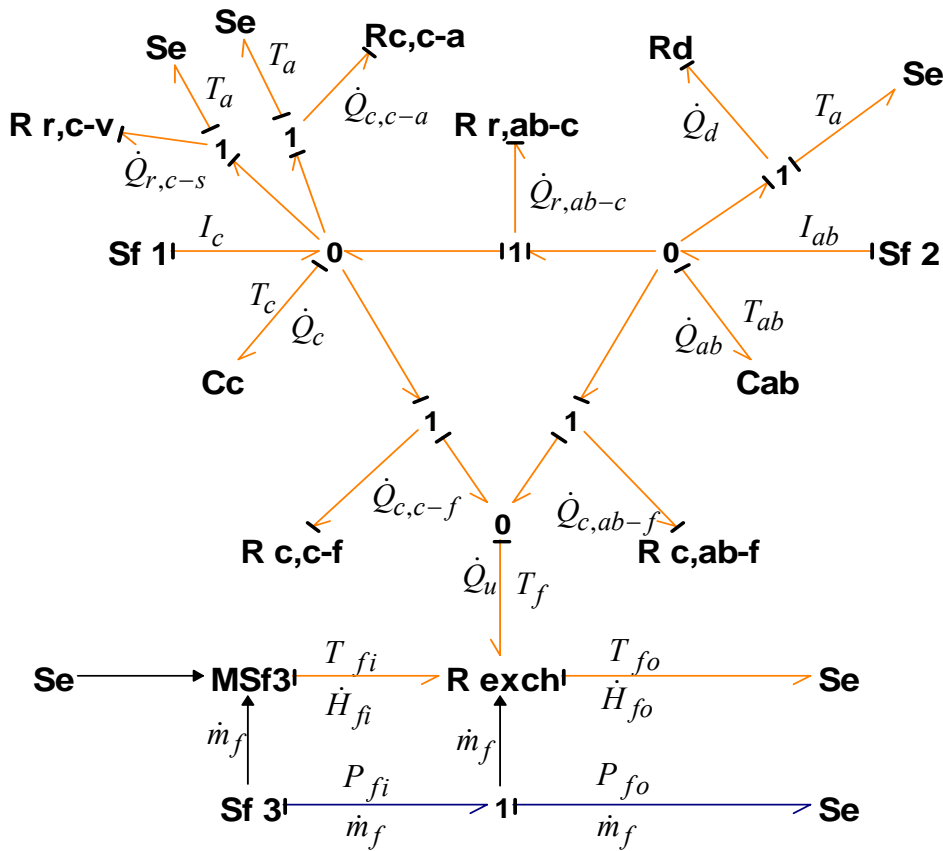


Figure 4: Pseudo bond graph model of the solar air collector.

**R**, **Se**, **Sf** and **MSf** and constitute the junction structure of the Pseudo Bond Graph model.

### 3.2. Bond graph elements and deduced mathematical equations

#### 3.3.1. Flow sources

**Sf1** and **Sf2** are used as input to the thermal sub-system **Sf1** represent the solar flow absorbed by the cover and the corresponding flow is:

$$I_c = \alpha_c G A_c \quad (1)$$

where  $\alpha_c$  is the absorptivity of solar radiation by the glass cover,  $G$  is the solar radiation rate incident on the glass cover ( $\text{W/m}^2$ ) and  $A_c$  is the cover area ( $\text{m}^2$ ).

**Sf2** represent the solar flow absorbed by the absorber and the corresponding flow is:

$$I_{ab} = 0.96 \tau_c \alpha_{ab} G A_{ab} \quad (2)$$

where the factor “0.96” represents the averaged transmittance-absorptance product,  $\alpha_{ab}$  is the solar-radiation absorptive of the absorber plate,  $\tau_c$  is the solar-radiation emissivity of the glass cover and  $A_{ab}$  is the absorber plate area ( $\text{m}^2$ ).

**Sf3** is used as input to the hydraulic subsystem, it represents the air mass flow entering the collector. The corresponding flow is  $\dot{m}_f$  ( $\text{kg/s}$ ).

**MSf3** is a modulated or controlled flow source. It represents the air enthalpy flow entering the collector in the thermal subsystem. This way the enthalpy input of the air can be calculated as the specific enthalpy associated to the input temperature multiplied by the input mass flow:

$$\dot{H}_{fi} = \dot{m}_f H_{fi} = \dot{m}_f C_{pf} T_{fi} \quad (3)$$

Where  $C_{pf}$  ( $\text{J/kg}^\circ\text{C}$ ) is the specific heat of the fluid (air) and  $T_{fi}$  the fluid temperature at the inlet of collector.

#### 3.3.2. Effort sources

**Se** is only the effort source used in this pseudo bond graph model and it represents the ambient temperature  $T_a$ .

#### 3.3.3. C-fields

The C- fields describe the storage energy phenomena and determine the effort variable or the flow according to the fixed causality using this formulation:

$$\begin{cases} e = \varphi_c^{-1}(\int f dt) \\ f = \frac{d}{dt} \varphi_c(e) \end{cases} \quad (4)$$

**Cc** represents the energy flow accumulation on the glass cover. The temperature of the glass cover is given by:

$$T_c = \frac{1}{C_c} \int \dot{Q}_c dt \quad (5)$$

Where  $\dot{Q}_c$  is the thermal heat flow accumulated on the glass cover and  $C_c$  is thermal capacity of glass cover.

$$C_c = \rho_c V_c C_{pc} \quad (6)$$

With  $\rho_c$  ( $\text{kg/m}^3$ ) the density,  $V_c$  ( $\text{m}^3$ ) the volume and  $C_{pc}$  ( $\text{J/kg}^\circ\text{C}$ ) the specific heat of the glass cover.

**Cab** represents the energy flow accumulation on the absorber plate; the temperature of the absorber plate is

$$\text{given by: } T_{ab} = \frac{1}{C_{ab}} \int \dot{Q}_{ab} dt \quad (7)$$

Where  $\dot{Q}_{ab}$  is the thermal heat flow accumulated on absorber and  $C_{ab}$  is the thermal capacity of the absorber plate ( $\text{J}^\circ\text{C}$ ).

$$C_{ab} = \rho_{ab} V_{ab} C_{pab} \quad (8)$$

With  $\rho_{ab}$  ( $\text{kg/m}^3$ ) the density,  $V_{ab}$  ( $\text{m}^3$ ) the volume and  $C_{pab}$  ( $\text{J/kg}^\circ\text{C}$ ) the specific heat of the absorber plate.

#### 3.3.4. R-fields

In our case, R-fields illustrate the heat transfer phenomena in thermal process; the effort or flow variable are determined with taking causalities into account and using this formulation:

$$\begin{cases} e = \varphi_R(f) \\ f = \varphi_R^{-1}(e) \end{cases} \quad (9)$$

Besides the thermal resistances  $R$  are equal to the inverse of heat transfer coefficient  $h$  ( $R=1/h$ )

**Rr,c-s** field models the radiation heat transfer phenomenon between the glass cover and the sky, the radiative heat flow is given by:

$$\dot{Q}_{r,c-v} = \frac{1}{R_{r,c-s}} (T_c - T_a) A_c = h_{r,c-s} (T_c - T_a) A_c \quad (10)$$

$A_c$  is the glass cover area ( $\text{m}^2$ ) and  $h_{r,c-s}$  ( $\text{W/m}^2^\circ\text{C}$ )

is the radiation heat-transfer coefficient from the glass cover to sky, the referred to the ambient air temperature  $T_a$  may be obtained as follows (Zhai X.Q, Dai Y.J and Wang R.Z 2005):

$$h_{r,c-s} = \sigma \varepsilon_c (T_c + T_s) (T_c^2 + T_s^2) \frac{(T_c - T_s)}{(T_c - T_a)} \quad (11)$$

where  $\sigma = 5.67 \times 10^{-8} \text{ w/m}^2\text{k}^4$  is the Stefan-Boltzmann constant,  $\varepsilon_c$  is the emissivity of thermal radiation of the glass cover and the sky temperature  $T_s$

is estimated by the formulation given by Swinbank

$$(Swinbank WC 1963) \text{ as: } T_s = 0.0552T_a^{1.5} \quad (12)$$

**Rr,ab-c** field models the radiation heat transfer phenomenon between the absorber plate and the glass cover. The radiative heat flow is given by:

$$\dot{Q}_{r,ab-c} = \frac{1}{R_{r,ab-c}} (T_{ab} - T_c) A_{ab} = h_{r,ab-c} (T_{ab} - T_c) A_{ab} \quad (13)$$

$A_{ab}$  is the absorber plate area ( $m^2$ ) and  $h_{r,ab-c}$  ( $W/m^2\text{°C}$ ) is the radiation heat-transfer coefficients between the glass cover and the absorber plate is predicted by:

$$h_{r,ab-c} = \frac{\sigma(T_{ab}^2 + T_c^2)(T_{ab} + T_c)}{\frac{1}{\varepsilon_{ab}} + \frac{1}{\varepsilon_c} - 1} \quad (14)$$

where  $\varepsilon_{ab}$  is the emissivity of thermal radiation of the absorber plate.

**Rd** field models the conduction heat-transfer phenomena across the insulation; the thermal loss is given by:

$$\dot{Q}_d = \frac{1}{R_d} (T_{ab} - T_a) A_{ab} = h_d (T_{ab} - T_a) A_{ab} \quad (15)$$

$h_d$  ( $W/m^2\text{°C}$ ) is the conductive heat-transfer coefficient across the insulation and estimated by:

$$h_d = \frac{k_i}{d_i} \quad (16)$$

Where  $k_i$  ( $W/m\text{°C}$ ) is the thermal conductivity of the insulation and  $d_i$  (m) is the average mean thickness of the insulation.

**Rc,c-a** field models the convection heat-transfer phenomena between the glass cover and the wind, the convective heat flow is given by:

$$\dot{Q}_{c,c-a} = \frac{1}{R_{c,c-a}} (T_c - T_a) A_c = h_{c,c-a} (T_c - T_a) A_c \quad (17)$$

$h_{c,c-a}$  ( $W/m^2\text{°C}$ ) is the convective heat-transfer coefficient from the glass cover due to the wind and is recommended by McAdams (McAdams W.H 1954) as:

$$h_{c,c-a} = 5.7 + 3.8V_a \quad (18)$$

Where  $V_a$  (m/s) is the wind velocity of the ambient air.

**Rc,c-f** field models the convection heat-transfer phenomena between the cover and the fluid, the convective heat flow is given by:

$$\dot{Q}_{c,c-f} = \frac{1}{R_{c,c-f}} (T_c - T_f) A_c = h_{c,c-f} (T_c - T_f) A_c \quad (19)$$

$h_{c,c-f}$  ( $W/m^2\text{°C}$ ) is the convective heat-transfer coefficients for the fluid moving on the glass cover and

on the absorbing plate, are calculated by:

$$h_{c,c-f} = \frac{Nu k_f}{D_h} \quad (20)$$

Where  $k_f$  is the thermal conductivity of air,

$D_h = 2WH_g / (W + H_g)$  (in m) is the hydraulic diameter of the air flow channel formed by the glass cover and the absorbing plate,  $W$  (m) is the collector width,  $H_g$  (m) is the mean gap-thickness between the glass cover and the absorbing plate as sketched in Figure 1, and  $Nu$  is the Nusselt number for the convection in the air flow channel.

In free convection, heat transfer takes place through the fluid motion induced by temperature gradients. In such cases,  $Nu$  may be expressed as a function of the Grashof and Prandtl numbers  $Gr$  and  $Pr$  (Monteith J. L 1973) :

$$Nu = C(Gr Pr)^n \quad (21)$$

$$\text{Where } Gr = \frac{g\beta\Delta TL^3}{\nu^2}, Pr=0.7 \quad (22)$$

In which  $g$  the gravitational acceleration ( $m^2/s$ ),  $\beta$  the thermal expansion coefficient ( $1/\text{°C}$ ),  $\Delta T$  the temperature difference ( $\text{°C}$ ),  $L$  length of the air channel (m) and  $\nu$  the kinematic viscosity of fluid ( $m^2/s$ ).

$C$  and  $n$  are constants that depend on the geometry and flow type and have the following value

$C=0.54$  and  $n=1/4$ , for laminar flow

$C=0.14$  and  $n=1/3$ , for turbulent flow

**Rc,ab-f** field models the convection heat-transfer phenomena between the absorber plate and the fluid, the convective heat flow is given by:

$$\dot{Q}_{c,ab-f} = \frac{1}{R_{c,ab-f}} (T_{ab} - T_f) A_c = h_{c,ab-f} (T_{ab} - T_f) A_c \quad (23)$$

$h_{c,ab-f}$  ( $W/m^2\text{°C}$ ) is the convective heat-transfer coefficients for the fluid moving on the glass cover and on the absorbing plate.

$$\text{Where } h_{c,ab-f} = h_{c,c-f} = \frac{Nu k_f}{D_h} \quad (24)$$

**Rexch** field is a multi-port element as shown in Figure 4. it is used to model the heat exchange phenomena between the absorber plate, the glass cover and the fluid.

In the heat exchange process in pipes, we find a continuity of the mass flow, but a variation of temperature and thermal enthalpy. The multi-port element has the following constitutive equations:

$$\dot{Q}_u = \dot{H}_{fo} - \dot{H}_{fi} = \dot{m}_f C_{pf} (T_{fo} - T_{fi}) \quad (25)$$

Where  $\dot{Q}_u$  ( $W/m^2$ ) is the useful energy gain which heats the air in the channel from the inlet temperature



$T_{fi}$  to the outlet temperature  $T_{fo}$ , resulting in the mean

$$\text{air-temperature } T_f = \frac{T_{fi} + T_{fo}}{2} \quad (26)$$

$$\text{So } T_{fo} = 2T_f - T_{fi} \quad (27)$$

Replacing (25) in (24) we obtain:

$$T_f = \frac{\dot{Q}_u}{2\dot{m}_f C_{pf}} + T_{fi} \quad (28)$$

The efficiency of solar heat gain of the heater is

$$\eta = \frac{\dot{Q}_u}{AG} = \frac{2C_{pf}\dot{m}_f(T_f - T_{fi})}{AG} \quad (29)$$

3.3.5 1-junctions correspond to the equality of flows

$$(\sum_i e_i = 0)$$

3.3.6 0-junctions correspond to the equality of effort

$$(\sum_i f_i = 0) \text{ and represent energy flow balances}$$

- The energy balance equation for the cover is written as:

$$\dot{Q}_c = I_c + \dot{Q}_{r,ab-c} - \dot{Q}_{c,c-f} - \dot{Q}_{c,c-a} - \dot{Q}_{r,c-s} \quad (30)$$

Where  $\dot{Q}_c$  is the heat flow accumulated on the cover,

$I_c$  solar flow absorbed by the cover,  $\dot{Q}_{r,ab-c}$  the

radiation heat flow between the absorber and the cover,

$\dot{Q}_{c,c-f}$  the convection heat flow between the cover

and the fluid,  $\dot{Q}_{c,c-a}$  the convection heat flow

between the cover and the environment and  $\dot{Q}_{r,c-s}$  is

the radiation heat flow between the cover and the sky.

- The energy balance equation for the absorber plate is written as:

$$\dot{Q}_{ab} = I_{ab} - \dot{Q}_{r,ab-c} - \dot{Q}_{c,ab-f} - \dot{Q}_d \quad (31)$$

Where  $\dot{Q}_{ab}$  is the heat flow accumulated on the

absorber plate,  $I_{ab}$  the solar flow absorbed by the

absorber,  $\dot{Q}_{r,ab-c}$  the radiation heat flow between the

absorber and the cover,  $\dot{Q}_{c,ab-f}$  the convection heat

flow between the absorber and the fluid and  $\dot{Q}_d$  is the

conduction heat flow between the absorber and the environment through the insulation.

- The energy balance equation for the air inside the collector is written as:

$$\dot{Q}_u = \dot{Q}_{c,c-f} + \dot{Q}_{c,ab-f} \quad (32)$$

$\dot{Q}_u$  is the useful energy gain,  $\dot{Q}_{c,c-f}$  the convection

heat flow between the cover and the fluid and

$\dot{Q}_{c,ab-f}$  is the convection heat flow between the absorber and the fluid.

## 4. Results and discussion

For the numerical appreciation of the developed model for the solar air collector thermal performance, the calculations have been made by using the system parameters (Table 1) and climatic data for one day of the month of May 2011(05-05-2011).

The software 20-sim was used for all simulations. It is dedicated to the Bond Graph simulation. Its use is rather simple and direct.

In a series of experiments conducted, data were recorded for different operating variables to determine the performance of the solar air collector. This section presents the experimental results of solar air collector and, also, a comparison with the predicted values.

Some scatter values can be seen in the experimental data. Fluctuations in the meteorological variables have contributed to the scatter in the results.

Table. 1: Values of parameters used in the analysis.

Parameters	Values
$A_{ab}$	2.2 (m <sup>2</sup> )
$A_c$	2 (m <sup>2</sup> )
$C_{p_{ab}}$	860 (J/Kg°C)
$C_{p_c}$	840 (J/Kg°C)
$C_{p_f}$	1006 (J/Kg°C)
$H_g$	0.04 (m)
$W$	1 (m)
$L$	2 (m)
$d_i$	0.05 (m)
$k_i$	0.034 (W/m°C)
$k_f$	0.026 (W/m°C)
$\rho_{ab}$	2530 (Kg/m <sup>3</sup> )
$\rho_c$	2700 (Kg/m <sup>3</sup> )
$\rho_f$	1.05 (Kg/m <sup>3</sup> )
$\nu$	$1.88 \times 10^{-5}$ (m <sup>2</sup> /s)
$\alpha_{ab}$	095
$\alpha_c$	0.06
$\varepsilon_{ab}$	0.94
$\varepsilon_c$	0.9
$\tau_c$	0.84

Figure 5 shows the variation of solar radiation and ambient temperature for the day 05-05-2011

Figure 6 shows the variation of the collector inlet and outlet air temperatures; it is clearly that the collector outlet air temperature is very sensitive to the variation of the solar radiation rate.

The influence of the solar radiation rate on the collector air outlet temperature and the efficiency is shown in Figures 7 and 9. It can be seen that for a steady sunny day, the air collector outlet temperature and the efficiency increases linearly with the solar radiation rate.

Figure 8 shows that the great efficiency is attained between 15 and 16 h, when the solar radiation rate is greatest to midday. The solar air collector accumulates

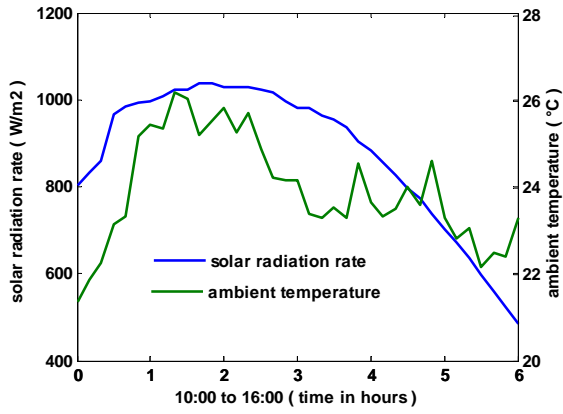


Figure 5: Variation of solar radiation and ambient temperature versus time of the day. (The weather conditions on May 5, 2011).

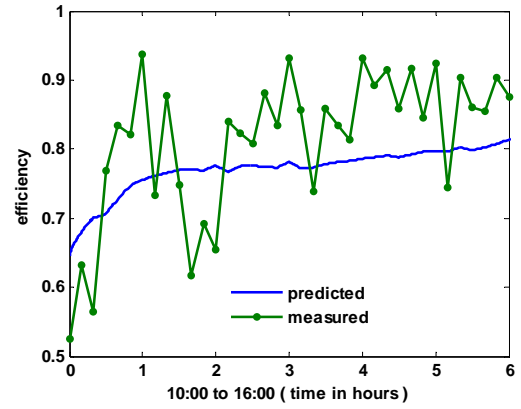


Figure 8: Variation of predicted and measured collector efficiency versus time of the day.

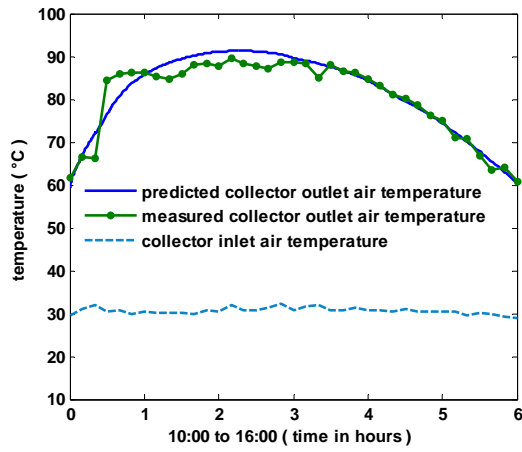


Figure 6: Variation of collector inlet and outlet air temperature versus time of the day.

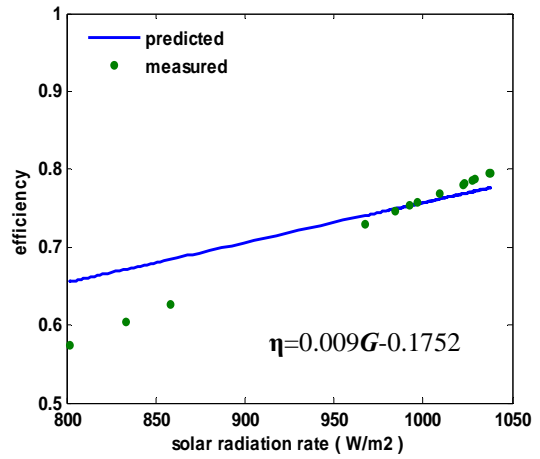


Figure 9: Effect of the solar radiation on the collector efficiency.

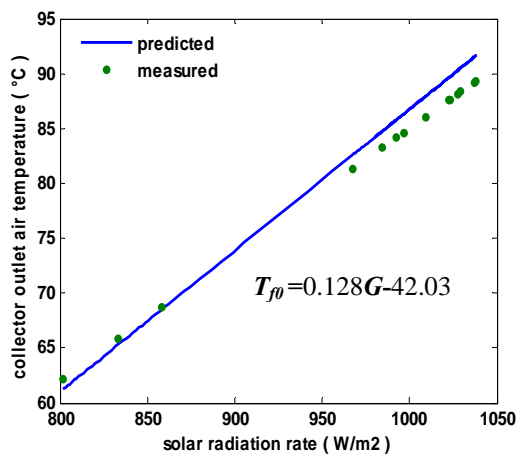


Figure 7: Effect of the solar radiation on the collector outlet air temperature.

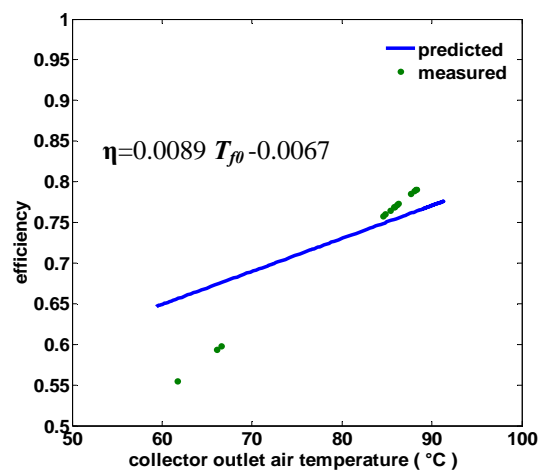


Figure 10: Effect of the collector outlet air temperature on the collector efficiency.

the energy in the morning when the solar radiation rate increases and resituates it the afternoon when the solar radiation rate decreases. The efficiency of the solar air collector depends significantly on the solar radiation and surface geometry of the absorber (Karsli S. 2007). Figure 10 shows the variation of the efficiency with the outlet air temperature, the efficiency increases linearly with the outlet air temperature.

## 5. Conclusion

This paper presents a performance analysis of air heating flat plate solar collector. The results showed that the efficiency depends on the solar radiation and the conception of the solar air collector. The temperature rises varied almost linearly with the incident radiation. In this paper, we have also demonstrated that the thermal performance of a solar air collector can be studied by the bond graph approach.

## ACKNOWLEDGMENTS

This study was supported by the Research Program of Ministry of High Education and Scientific Research in the setting of Tunisian PRF Projects.

## REFERENCES

- Cellier F. E. 1991. *Continuous System modelling*. Springer-Verlag. N.Y: USA.
- Duffie J.A, Beckman W.A. 1991. *Solar engineering of thermal processes*. 2nd ed. New York: Wiley.
- Dickson W.C. Cheremisinoff P.N. 1980. *Solar energy technology handbook*. New York: Marcel Dekker.
- Garg H.P, Kumar R. 2000. Studies on semi-cylindrical solar tunnel dryers: thermal performance of collector. *Applied Thermal Engineering*, (20), 115-131.
- Gao W, Lin W , Liu T, Xia C. 2007. Analytical and experimental studies on the thermal performance of cross-corrugated and flat-plate solar air heaters. *Applied Energy*, (84) 425–441.
- Hadim H, Vafai K. 1999. Overview of Current Computational Studies of Heat Transfer in Porous Media and their Applications: Forced convection and multiphase heat transfer In: *Advances in Numerical Heat Transfer*, vol. 2 (eds. W.J. Minkowycz and E.M. Sparrow), Taylor & Francis, Washington D.C. 291 – 329.
- Karnopp D, Rosenberg R. 1983. *Introduction to Physical Dynamics Systems*. McGraw-Hill Book Company. New York: USA.
- Karnopp D, Margolis D, Rosenberg R. 1990. *System Dynamics: A Unified Approach*. John Wiley & Son Inc. New York: USA.
- Karsli S. 2007. Performance analysis of new-design solar air collectors for drying applications. *Renewable Energy*, (32) 1645 –1660.
- McAdams W.H. 1954. *Heat transmission*. 3rd ed. McGraw-Hill . New York: USA.
- Monteith J. L. 1973. *Principles of Environmental Physics*. Edward. Arnold. New York: USA.
- Ould Bouamama B. 2003. Bond Graph Approach AS .nalysis Tool In Thermofluid Model Library Conception. *Journal Of Franklin Institute*, (340) 1-23.
- Swinbank W.C. 1963. Long-wave radiation from clear skies. *Quart J Royal Meteorol Soc*, (89), 339–48.
- Thoma J. Ould Bouamama B. 2000. *Modelling and Simulation in Thermal and Chemical Engineering. Bond Graph Approach*. Springer. Berlin: Germany.
- Zhai X.Q, Dai Y.J, Wang R.Z. 2005. Comparison of heating and natural ventilation in a solar house induced by two roof solar collectors. *App Therm Eng*, (25), 741–57.

# THE USE OF SIMULATION AS A PEDAGOGICAL TOOL IN CONSTRUCTION EDUCATION

Ronald Ekyalimpa<sup>(a)</sup>, Jangmi Hong<sup>(b)</sup>, Simaan M. AbouRizk<sup>(c)</sup>

<sup>(a)</sup> Graduate Student, Dept. of Civil and Environmental Engineering,  
Univ. of Alberta, Edmonton, Alberta, Canada T6G 2G7.

<sup>(b)</sup> Graduate Student, Dept. of Civil and Environmental Engineering,  
Univ. of Alberta, Edmonton, Alberta, Canada T6G 2G7.

<sup>(c)</sup> Professor, Dept. of Civil and Environmental Engineering, Univ. of  
Alberta, Edmonton, Alberta, Canada T6G 2G7.

<sup>(a)</sup> [rekyalimpa@ualberta.ca](mailto:rekyalimpa@ualberta.ca), <sup>(b)</sup> [jangmi@ualberta.ca](mailto:jangmi@ualberta.ca), <sup>(c)</sup> [abourizk@ualberta.ca](mailto:abourizk@ualberta.ca)

## ABSTRACT

The construction industry's use of simulation has greatly evolved since its introduction in the mid-70s. Aside from its use in day-to-day activities, it has been shown to be a vital tool for the transfer of specialized construction management knowledge and skills to students, which would otherwise have been acquired through a lengthy, risky, and expensive learning process on the jobsite. The construction industry has experienced considerable changes and development with respect to public & client expectations, project size, and complexity, necessitating the delivery of graduates who are already proficient in managing construction projects. This paper presents details on how simulation is used to teach a construction analysis and design course at the University of Alberta, using Symphony.NET, simulation software developed by the third author. The paper introduces Symphony.NET along with the functionality of its various modeling elements and discusses the various aspects of simulation taught in the course. Two case studies which are also covered in the course, a paving and a building operation, are then presented, as well as an exploration of other issues which arise when using a simulation-based approach to teaching a course in construction engineering.

**Keywords:** Simulation, Symphony, Construction engineering and management research and education, simulation-based approach.

## 1. INTRODUCTION

Simulation was introduced to the construction domain when Halpin (1977) first proposed CYCLONE, a simple simulation modeling language. Thereafter, many practitioners and researchers realized and began to explore the potential of simulation in academia and in practical settings in industry.

Simulation has since been used in construction engineering and management education for both instruction and research purposes. It serves as a vital

complementary tool for traditional approaches used for teaching special courses that require the transfer of knowledge and skills to students, which would otherwise have been acquired through an expensive and lengthy learning process on the jobsite.

With the evolution and growth of the construction industry, with respect to project size, complexity, uniqueness and other stringent project requirements, it is becoming more apparent that simulation is an irreplaceable tool with respect to its utilization as a teaching aid in classroom settings. The general public and the construction industry at large expect fresh graduates to possess sound decision making skills and the technical expertise to overcome any challenges that they are likely to face while working on the job site. A number of simulation software packages have emerged over the years through different research initiatives at a number of universities, and several have been used to cope with this ever growing need. Some of the more well-known simulation software packages include CYCLONE (Halpin 1977), DISCO (Huang et al. 1994), ABC (Shi 1999), INSIGHT (Paulson 1978), RESQUE (Chang and Carr 1987), UM-CYCLONE, COOPS (Liu 1991), CIPROS (Odeh 1992), STROBOSCOPE (Martinez and Ioannou 1994), HSM (Sawhney and AbouRizk 1995) and Symphony (AbouRizk and Mohamed 2000).

The Hole School of Construction Engineering at the University of Alberta offers a construction process design and analysis course specifically designed to fill this need in the construction industry. The course is designed to teach students the science of analyzing and designing construction operations. Given that the nature of the course dictates the use of simulation, students are introduced to and taught how to use Symphony to implement practical modeling exercises covered in the course.

The general sequence of instructing the course is as follows: first, students are introduced to the concepts of generating meaningful abstractions of any process, and different aspects of creative modeling, with emphasis

being placed on construction related operations. A typical earth-moving operation is usually used for this purpose. The CYCLONE template is used to introduce students to the concept of building simulation models on computer because of its simplicity. Students are then taught how to process simulation models by hand and then subsequently introduced to modeling using the Symphony General Purpose template. Other key concepts in simulation are taught, such as statistical aspects of simulation (input and output modeling), model verification and validation besides the instruction on the use of Symphony software for process modeling.

Several examples and case studies are used during the course to help students appreciate the merits of adopting simulation-based approaches for solving problems in a practical setting. A paving operation and a building construction project are presented and discussed as case studies in this paper in order to illustrate the extent to which practical construction problems are covered in the course.

### **1.1. Current Use of Simulation in Construction Education**

Lansley (1986) postulated that human beings learn best when they start from a scenario that they are familiar with and then progress to those which are new to them. Simulation-based approaches used in construction education adopt this same approach because they require that the students first get a good understanding of the systems they plan to simulate (analyze and design). This knowledge ensures that students make accurate abstractions of those systems when modeling using general purpose simulation templates.

The use of games and simulation modeling tool kits are the two commonly used applications of simulation-based approaches in construction education. Examples of games used in construction management education include CONSTRUCTO by Halpin (1976); a road game by Harris & Evans (1977); a dam construction game by Al-Jibouri & Mawdesley (2001); bidding games by Au et al (1969), AbouRizk (1993) and AbouRizk et al (2009). Others include Easy Plan by Hegazy (2006); an equipment replacement game by Nassar (2002), and STRATEGY by McCabe (2000). These games were developed at Universities and are applied in teaching students the concepts of construction management.

A number of simulation modeling tool kits exist on the market which can be used for simulation modeling. Examples of other packages in use in the simulation domain include Promodel, ARENA® (Rockwell Automation), Repast Suite (Argonne National Laboratories), SLAM (Pritsker et al. 1989) and AnyLogic® (XJ Technologies). Some of these software packages support multiple simulation paradigms such as discrete event simulation, system dynamics and agent-based simulation, while others only support one paradigm.

These tools provide an environment in which the students can be immersed so that they interact with

uncertainty and variability associated with real systems that they will be dealing with when they graduate.

## **2. SIMPHONY IN CONSTRUCTION MANAGEMENT EDUCATION**

Simphony is a simulation environment which supports a discrete event simulation paradigm. It is comprised of a simulation engine, templates, modeling features and an interface. The interface of the current version of Simphony, Simphony.NET 4.0, is shown in Figure 1. Simphony supports the development and use of different simulation templates. A simulation template is defined as a collection of abstract elements that are used in simulation modeling. Templates that comprise elements which are generic are referred to as general purpose templates, while those with elements customized for modeling a specific domain are referred to as special purpose templates.

Simphony.CYCLONE and Simphony.General are general purpose templates with generic, easy to use modeling elements that can be used to represent a wide spectrum of systems in the construction domain. The general purpose template has a total of 6 categories of modeling elements and a total of 27 generic modeling elements. Each element has a unique appearance, properties and simulation behavior. When building a model, elements are dragged from the templates' window and dropped onto the modeling surface and then linked by relationships (arrows) which provide a route for the flow of entities throughout the entire model. Relationships also enable the modeler to represent their logic as they build the model.

The special purpose templates supported by Simphony in previous versions and the current version include tunneling, steel fabrication, aggregate production, range estimating, earthmoving and PERT (Hajjar and AbouRizk 1999; AbouRizk and Mohamed 2000). These special purpose templates enable practitioners to make use of their knowledge and skills in solving real problems using simulation-based approaches. This is made possible by ensuring that all modeling elements developed in each of these templates have a close resemblance with each aspect of the domain they represent.

Simphony also supports the development of other user specific simulation software (special purpose templates). This provision gives Simphony a competitive edge over other simulation modeling software as a research tool and as a tool for use in regular class instruction. It should be noted that developing these special purpose templates in Simphony requires computer programming proficiency, time and effort, and a clear understanding of the domain for which the template is being developed.

Students in this course are taught how to develop their own templates and how to use general purpose templates and special purpose templates for real construction applications.

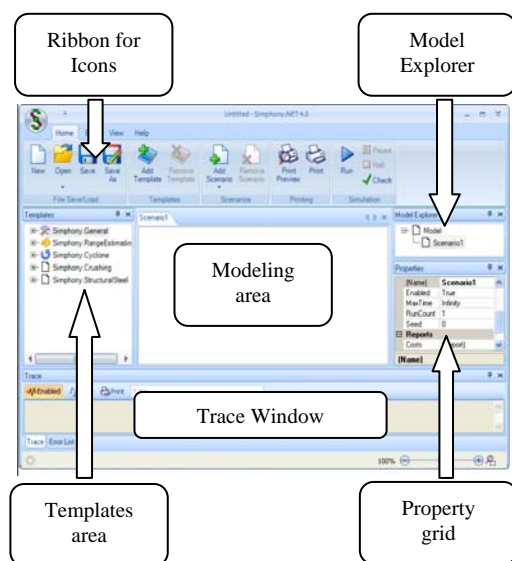


Figure 1: The User Interface of Symphony

The following sections discuss other aspects covered in the course using a simulation-based approach.

## 2.1. Creating Abstractions of Systems to be Simulated

Constructing an accurate abstraction of a system to be simulated is a skill which is developed over time as one continuously engages in simulation modeling. It is a process referred to as abstraction and is the first phase in modeling a process or a system.

Abstracting a system involves determining the level of detail to which the system will be emulated in the modeling environment, the servers (resources) in our system, the customers, the different state changes (events) and the interaction of the different parts of the system. Once all this has been established, it is drawn on paper in the form of a schema. This diagram is then transferred onto computer modeling elements.

This stage of the modeling process helps the students understand the system they are attempting to model. Visually representing this abstraction on the modeling surface using arrows and modeling elements and the modeling surface within Symphony reinforces that and enables them to assess the validity of their final model and possible outcome.

## 2.2. Processing Simulation Models by Hand

Students are taught how to process simulation models by hand so that they gain an appreciation of what takes place behind the scenes when a simulation model is run on computer. This knowledge helps the students to verify and validate the models they build. Students are taught how to populate events on an event list, how to transfer events that have occurred onto a chronological list, and how to advance the time of the whole system using MS Excel. They are also taught how to compute system production rates and vital statistics such as utilization. The same models are then run on the

computer to validate the results obtained from the hand calculations.

## 2.3. Statistical Aspects of Simulation

Simulation is used when analyzing complex systems which are affected by multiple factors resulting in random behavior. One effective way of modeling such systems in a simulation-based approach is to represent uncertain behavior of the system in the form of statistical distributions. These statistical distributions may be constructed from previously collected empirical data or from expert knowledge.

Simphony can be used for statistical modeling of such systems because it supports a number of commonly used statistical distributions and also provides a framework for performing Monte Carlo simulations. In order for modelers to perform successful stochastic simulation studies, they have to carry out comprehensive input modeling, a Monte Carlo simulation, and an output analysis.

Input modeling is the process of fitting statistical distributions to data, and testing how well the selected distributions fit to that dataset (i.e., goodness of fit tests). “Fitting” distributions is simply computing their statistics, such as their boundaries, shape and location parameters. Triangular distributions are commonly fitted to data captured from expert opinion because they don’t require complex techniques to be constructed. The distributions often fitted to empirical data include triangular, uniform, beta, exponential, uniform, lognormal, etc. Tests for goodness of fit include visual assessments of shapes of empirical and theoretical probability density functions (PDFs) and cumulative density functions (CDFs), Kolmogorov-Smirnov tests, and Chi-Square tests. Simphony does not support these types of analysis, although there are a number of commercial packages available such as BESTFIT, Crystal Ball® (Oracle) and @Risk (Palisade) which do. The choice of the distribution to use ultimately depends on whether it is supported by the intended simulation engine to be used and how well it fits to the dataset.

After input modeling has been completed, the selected distributions are entered into a pre-built simulation model and then the simulation is run. This type of simulation is referred to as Monte Carlo simulation. Monte Carlo simulation is defined as a numerical solution to a problem in which the system or phenomenon associated with the problem is modeled through a random sampling process (Bielajew 2001). Reliable results can be guaranteed for experiments with higher number of simulation runs (i.e., higher sample size).

Output analysis, on the other hand, involves investigating to determine whether the simulation results are accurate. This is done by carrying out tests for normality of the output (i.e., Shapiro-Wilk) and by computing confidence intervals for the output data. Figure 2 shows a flow chart of a typical process of carrying out an authentic stochastic simulation experiment.



In this course, students are taught how to do input modeling by hand calculation and then by using @Risk (Palisade). Monte Carlo simulation using Symphony is also covered, together with output analysis.

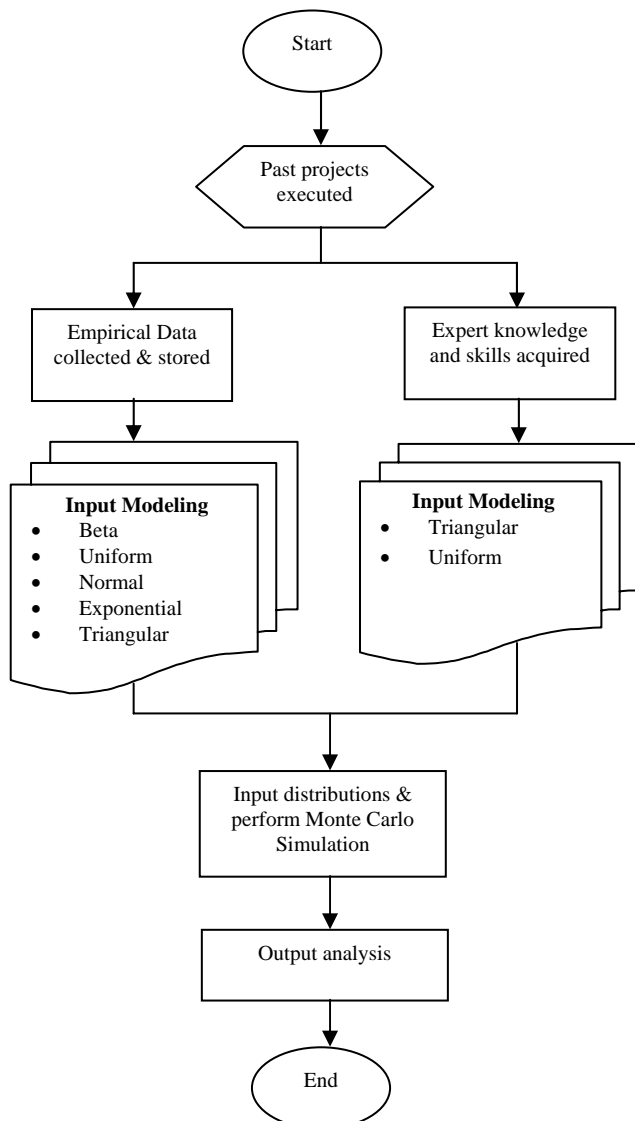


Figure 2: A Flow Chart of a Typical Stochastic Simulation Process

## 2.4. Validation and Verification of Simulation Models

After building a simulation model, it is important to review the model layout and inputs to ensure that it is an accurate abstraction of the domain represented; a process referred to as model validation. Thereafter, verification has to be done, which involves confirming that the logic embedded within the model is compatible with the way the modeler wanted the model to behave.

Simphony provides a number of features that assist modelers to verify the behavior of their models. For example, it has a counter element, charting elements, trace properties for each modeling element and a trace window. A counter is used to track the flow of entities in any part of the model and also provides insight into

time the time entities last flowed through a specific part of a model. The trace features allow a modeler to display simulation results or events in text format, as the model runs. Charting elements allow for the visualization of data generated from the simulation.

Iconic visualization is another feature that is very useful for verifying models in simulation. Although the current general template does not have features for visualizing events as they unfold during simulation, some special purpose templates such as an earthmoving template support iconic visualization of trucks flowing through a model. There are also a number of integrity checks which Simphony performs before a model is run. These checks serve as warnings to modelers of any problems in logic that may exist in their models. The process of validation and verification is a very important phase in simulation modeling to guarantee accurate results. However, this process takes a lot of time and it is challenging.

## 3. CASE STUDIES

A number of different case studies are presented when instructing the course so that students get an appreciation of what is required to successfully model construction processes. The emphasis is to enable students to develop a sense of some of the key issues in process interaction modeling. Examples of these issues include:

- Identification of what the entity or entities in the model will represent
- The parts of the system to model explicitly as resources
- The time units to use for modeling and
- The stopping criteria for the simulation.

The two case studies presented in this paper include a high-rise building construction project and an asphalt paving operation. Both are repetitive construction processes which are resource intensive and have a number of interacting cyclic processes.

### 3.1. Building Construction Project

The project presented has 26 identical rectangular floors to be constructed, each with 15 bays @ 34ft by 34ft (10.3 m by 10.3 m), as shown in Figure 3. First, forms are installed for a given floor. Once the forms for the entire floor are installed, reinforcing steel is placed for that floor. Then concrete placement commences, with each bay requiring 70CY (54 m<sup>3</sup>) of concrete, which is moved from a hopper to the placement area via buggies. Once concrete placement is done, forms can be removed after the concrete has cured. Once forms are removed they are moved by cranes to the next floor and the process repeats until the 26 floors are done.

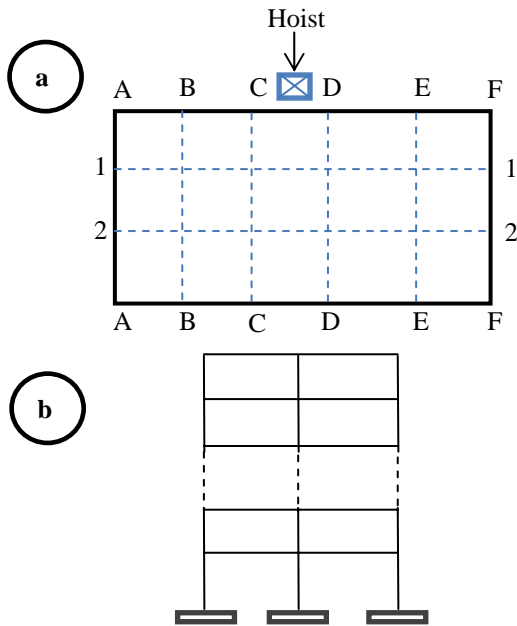


Figure 3: (a) Schematic Layout of Building Construction Project Floor Plan. (b) Right Side Elevation of Building Construction Project

An abstraction of the entire building construction process is summarized in the flow chart presented in Figure 4. When the base model was run, results showed that the average production rate of the project would be 0.55 completed floors per month and the total duration would be 565,600 minutes or approximately 47 months, assuming a 10 hr, 5 day, 4 week-month work shift configuration.

The benefit of using simulation in such an analysis is that the modeler can easily identify and experiment with numerous issues with the objective of arriving at an optimal system, hence saving time and money. For example, resource utilization results can be used to identify resources which are idle most of the time, allowing for waste to be eliminated. In this analysis, it was decided that the number of buggies were to be optimized since they had a very low utilization value. Sample results of the optimization experiment for getting the adequate number of buggies are summarized in Table 1. Results were plotted on a graph in MS Excel (shown in Figure 5) from which it was deduced that the optimal number to use for the project without increasing the project duration is 8, instead of the originally assigned number of 14.

Table 1: Summary of Results of Buggy Optimization

Parameter	Scenario 1	Scenario 2	Scenario 3
# of Buggies	12	6	3
Utilization (%)	2.5	5.0	10.0
Queue Length	0	0.1	0.26
Waiting Time (Min.)	0	0.30	1.30
Project Duration (Min.)	565600	565800	565900

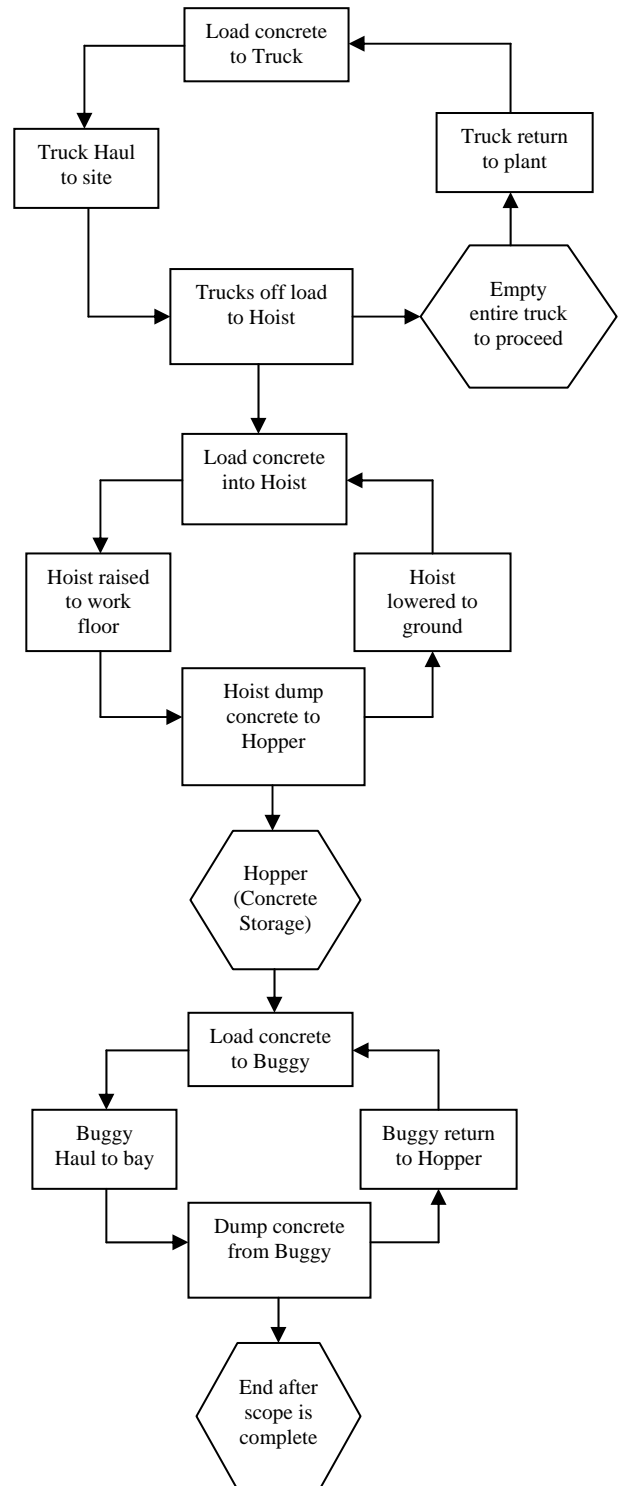


Figure 4: Flowchart of the Building Construction Processes

A sensitivity analysis was carried out for this building construction project to identify the bottle-neck activities. The results for this analysis are summarized in Table 2 for some activities in this project.

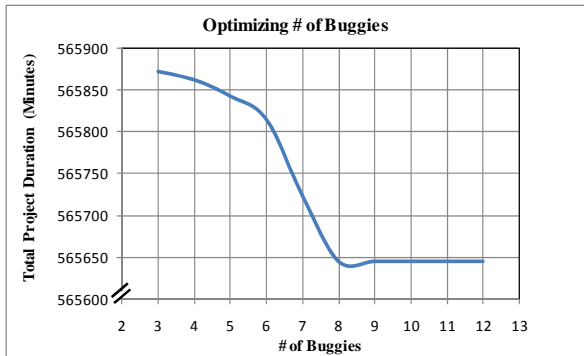


Figure 5: Number of buggies vs. total project duration in minutes

Table 2: Results of a Sensitivity Analysis on selected Activities in the Building Project

Activity Eliminated	Total Project Duration (Min.)	% Project time reduction
Erect & level forms	563200	0.4
Place rebars	562400	0.6
Concrete truck travel to plant	565700	0
Concrete truck returning to site	565600	0
Fill buggy, transport & dump concrete	565700	0
Curing slab	549400	2.9
Fly forms	562300	0.6
Hoist Concrete	280400	50.4
Lower Hoist	373900	33.9

From the results summarized in Table 2, it is evident that hoisting of concrete and lowering the hoist are the activities mainly controlling the project duration. This result was confirmed by a mean utilization value of 100% obtained from running a simulation on the base model. This implies that improvements would have to be made to the hoist cycle if the project duration is to be reduced. Increasing the capacity or number of hoists could be one option. Another option could be to buffer the truck cycle and hoist cycle with a sizeable hopper.

Curing of concrete is another activity that has an impact on the overall project duration, although it is not as large as the hoisting of concrete and lowering of the hoist. If its impact is to be reduced, the use of accelerators and other super setting additives in the concrete could be considered in order to reduce the curing time.

### 3.2. Asphalt Paving Operation

The paving operation presented assumes that the sub-grade and base course construction have been completed. The surfacing is to be done in asphalt concrete which is delivered by trucks and dumped onto a paver that spreads it. The spread asphalt is then compacted by a drum compactor, after which a

pneumatic compactor makes finishing passes, as shown in Figure 6. Some constraints are imposed on the sequence of the operation which include: a paver has to spread a complete parallel pass before it releases the pass to a drum compactor. A pass requires 15 paver skips to be spread, after which the paver repositions for another pass. The drum compactor also releases a completed section to the pneumatic compactor. Each section requires 5 paver skips. The parking lot being paved is completed after 4 parallel passes have been made by the pneumatic compactor. It can be assumed that the operation is not constrained by the delivery of asphalt by trucks.

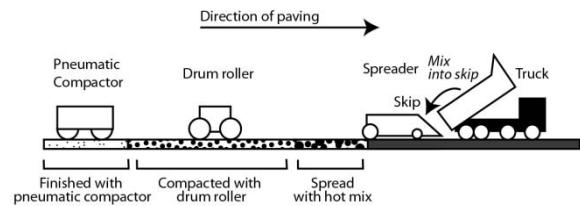


Figure 6: Paving Equipment Diagram (adapted from Halpin 1987)

The objective in this case study is to obtain the total project duration and plot a velocity diagram in Symphony. The generated velocity diagram can then be used to analyze the operation to ensure continuous resource utilization, which is good for the equipment condition and the overall project production rate. Symphony provides a favorable framework in which to experiment with the construction buffers, in order to achieve a continuous construction process without increasing the overall project duration.

A schematic layout is presented in Figure 7 which represents an abstraction of the operation. After the simulation is run, a total project duration of 1006.5 minutes is obtained. The velocity diagram shown in Figure 8 is also generated for the base scenario.

The construction buffer between the paver and the drum compactor is then reduced from 15 paver skips to 5 paver skips in order to synchronize all three cyclic processes and achieve continuous liner production lines for the drum and the pneumatic compactor. The same total project duration of 1006.5 minutes was obtained with this adjustment with production lines for all three pieces of equipment, almost parallel and linearly continuous throughout the project. This is illustrated in Figure 9.

A modeler can opt to explore the option of reducing the total project duration while maintaining continuous linear production lines by either increasing the number of resources for each construction cycle or by proposing to use equipment with higher production rates. This scenario-based analysis is very simple in this case, since a simulation model for the operation exists which can be experimented with.

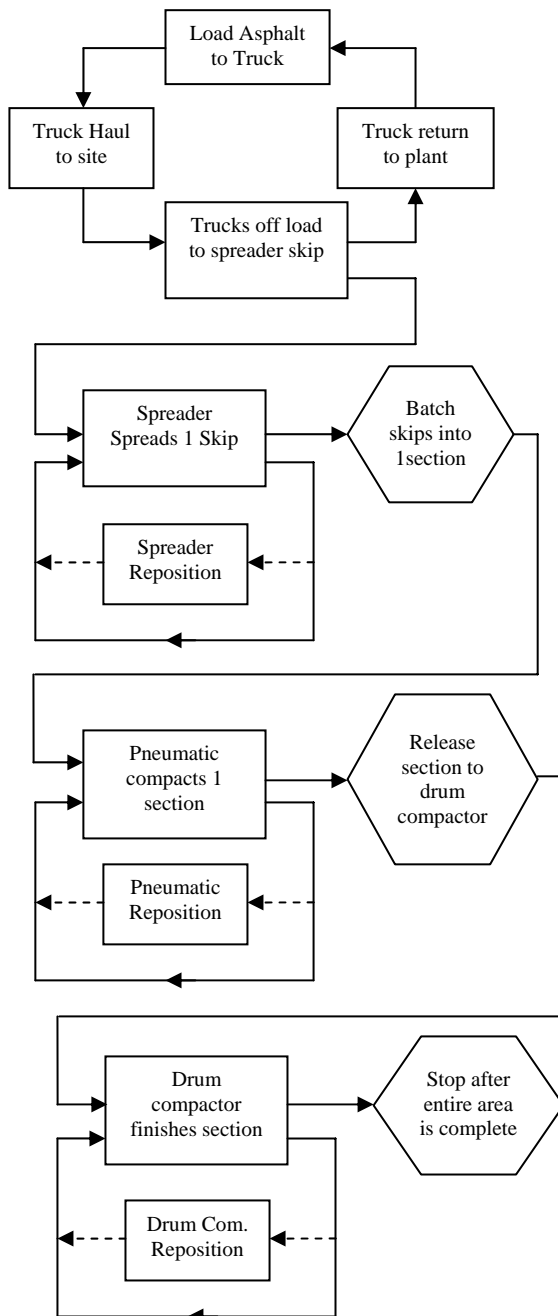


Figure 7: A Flowchart of the Paving Operation

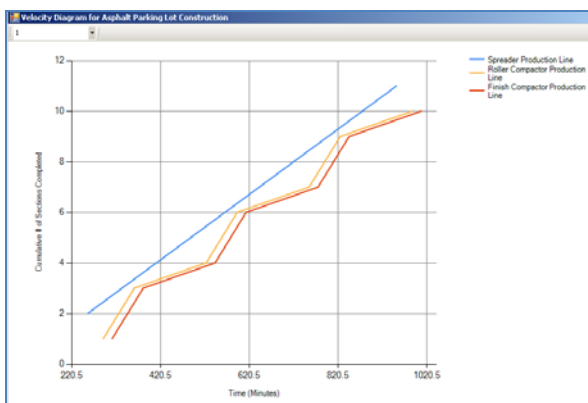


Figure 8: Velocity Chart for the Base-Case Scenario

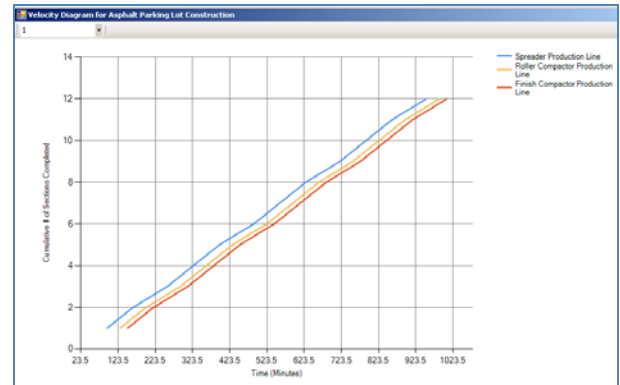


Figure 9: Velocity Chart for the Synchronized Operation

The issues dealt with in these two case studies are easy to investigate once base models for the operations are constructed in a simulation environment. It would be time consuming and tedious to carry out the same analysis without a simulation based approach hence its use in teaching this course.

Other issues that can be easily studied include the impact of the impact of equipment maintenance programs and uncertain events such as plant failure on the duration of such project.

#### 4. CONCLUSIONS

Simulation is a popular approach which provides an environment conducive to modeling and experimentation of problems which cannot be solved analytically due to their complex nature. Its use in academia is wide spread and continues to grow because it allows for instructors to bring uncertain and variable phenomena into typical classroom settings, giving students the chance to interact with them and develop the necessary knowledge and skills required to deal with them in a real life setting. Simulation also provides a relatively cheap test bed for students to experiment with complex systems before they go into the field to implement them, especially during in their research phase of their University programs.

Simphony is an easy-to-use simulation software package which allows the construction of models using its general purpose or special purpose templates, but also allows for the development of user specific simulation tools (i.e., other special purpose templates). It has been successfully used in courses at the University of Alberta during lectures for demonstrating case studies, and by students in solving their course assignments and exams. Simphony is also used by students who decide to pursue research in the area of simulation and has been shown to yield acceptable results.

When used appropriately, simulation-based approaches in construction management education can be very effective, especially when there is a good balance in the use of construction management games, simulation modeling tool kits and traditional methods of instruction.

## REFERENCES

- AbouRizk, S., and Mohamed, Y., 2000. Symphony-An Integrated Environment for Construction Simulation. *Proceedings of the 2000 Winter Simulation Conference*, pp. 1907-1914. December 10-13, Orlando (Florida, USA).
- AbouRizk, S. M., Hague, S., & Elmira, M., 2009. *Developing a Bidding Game Using High Level Architecture. Proceedings of the 2009 ASCE, International Workshop on Computing in Civil Engineering*, Austin, (Texas, USA), 513-522.
- AbouRizk, S. M., 1993. Stochastic Simulation of Construction Bidding and Project Management. *Microcomputers in Civil Engineering*, 8 (2), 343-353.
- Al-Jibouri, S & Mawdesley, M., 2001. Design and experience with a computer game for teaching construction project planning control. *Engineering Construction and Architectural Management*, 8 (5), 418-427.
- Au, T., Bostleman, R., & Parti, E., 1969. Construction Management Game – Deterministic Model. *Journal of Construction Division*, 95(CO1),25-38.
- Bielajew, A. F., 2001. Fundamentals of the Monte Carlo method for neutral and charged particle transport. Accessed at: <http://www-personal.umich.edu/~bielajew/MCBook/book.pdf> on 30<sup>th</sup> June 2011.
- Chang, D. Y., and Carr, R. I., 1987. "RESQUE: A resource oriented simulation system for multiple resource constrained processes." Proc., 1987 PMI Seminar/Symposium, Milwaukee, 4-19.
- Hajjar, D., and AbouRizk, S. M., 2002. Unified Modelling Methodology for Construction Simulation. *Journal of Construction Engineering and Management*, 128(2), 174- 185.
- Halpin, D. W., 1976. CONSTRUCTO – An Interactive Gaming Environment. *Journal of the Construction Division*, 102(CO1),145-156.
- Halpin, D. W., 1977. "CYCLONE - Method for modeling of job site processes." *Journal of the Construction Division, American Society of Civil Engineers*, 103(3), 489-499.
- Harris, F. C., & Evans, J. B., 1977. Road Construction – Simulation Game for Site Managers. *Journal of the Construction Division*, 103(CO3), 405-414.
- Hegazy, T., 2006. Computer Game for Simplified Project Management Training. *1st International Construction Specialty Conference (CSCE)*, Calgary, (Alberta, Canada).
- Lansley, P. R. (1986). Modeling Construction Organizations. *Journal of Construction Management and Economics*, 4(1), 19-36.
- Liu, L.Y., 1996. ACPSS — animated construction process simulation system. *Proceedings of the 3rd Congress on Computing in Civil Engineering, American Society of Civil Engineers*, pp. 397-403, New York, N.Y.
- Martinez, J. C., Ioannou, P. G., 1994. "General Purpose Simulation with Stroboscope," *Proceedings of the 1994 Winter Simulation Conference*, pp.1159-1166, December 11-14, Orlando, (Florida, USA).
- McCabe, B., Ching, K. S., & Rodriguez, S., 2000. STRATEGY: A Construction Simulation Environment. *Proceedings of Construction Congress*, pp.115-120, Orlando, (Florida, USA).
- Nassar, K., 2002. Simulation Gaming in Construction: ER, The Equipment Replacement Game. *Journal of Construction Education*, 7(1), 16-30.
- Odeh, A. M., 1992. Construction Integrated Planning and Simulation Model, *PhD Dissertation, Department of Civil & Environmental Engineering; University of Michigan, Ann Arbor, MI*.
- Paulson, G. C., Jr., 1978. Interactive graphics for simulating construction operations. *Journal of the Construction Division, American Society of Civil Engineers*, 104(1), 69-76.
- Pritsker, A. A. B., Sigal, C. E., and Hammesfahr, R. D. J., 1989. *SLAM II – Network Models for Decision Support*. 1st ed. New Jersey: Prentice-Hall, Inc.
- Sawhney, A., and AbouRizk, S.M., 1995. HSM — simulation-based planning method for construction projects. *Journal of Construction Engineering and Management*, 121(3), 297-303.
- Shi, J.J., 1999. Activity-based modeling approach for construction. *Proceedings of the 1999 Conference of the Canadian Society for Civil Engineering*, pp. 187-196. June 2-5, Regina, (Saskatchewan, Canada).

# LOGICAL OPERATOR USAGE IN STRUCTURAL MODELLING

Ieva Zeltmate <sup>(a)</sup>

<sup>(a)</sup> Riga Technical University, Faculty of Computer Science and Information Technology  
Department of System Theory and Design

[ieva.zeltmate@gmail.com](mailto:ieva.zeltmate@gmail.com)

## ABSTRACT

Structural modelling is an approach that is developed to support expert's knowledge acquisition and representation in the process of complex system analysis, design and modelling. The framework of several integrated structural models is used to capture and manage knowledge about specific system morphology, operations, and behaviour under certain operating conditions. To minimize expert's workload, once acquired knowledge must be automatically transformed between related structural models. Methods for structural model transformations are used to keep important system characteristics and connectedness between system components and functionality. In this paper special cases of structural model transformations are described. In models additional elements called logical operators are used. The goal of logical operator usage is to depict in structural models possible combinations and operation compatibilities that exist between real system components. The formal method that allows transform and use logic operators between models is described and examples of structural model transformations are given.

Keywords: structural modelling, transformations, functional, behaviour, parameters

## 1. INTRODUCTION

System structural modelling is a type of modelling that can be used in system design and analysis, to acquire expert's knowledge about system structure and organisation (Grundspenkis 1999; Kresken 1996; Oliva 2004). System structural modelling becomes important when it is necessary to understand and to change complex system and also predict its behaviour (Grundspenkis 1999). Usually, when system analysis is performed, a domain expert knows that system components and relationships among them are logically connected. However, when a system model is created, not all knowledge of the system structure may be presented. Sometimes it is because the model is not expressive enough to represent all necessary knowledge, but sometimes all knowledge about system is not known. Relationships between elements in the model show the connections between objects and logical operation sequences in the system in terms of normal functioning (Grundspenkis 1999; Oliva 2004).

Although a system structural model rarely answers how the relationships are connected or how elements collaborate or exclude one another. To address these questions it is important to represent coherence between element's relationships. Therefore additional elements, called logical operators, are used in Structural modelling approach (Grundspenkis 1997; Zeltmate and Grundspenkis 2008). Logical operators are extra characteristics (Capiluppi 2007) integrated in system structural models (Zeltmate and Grundspenkis 2008). Using logical operators allows to describe system's conformity to rules and order of activities. System representation with logical operators clearly defines existing paths, relationships, and flows in the models that describe real world systems.

Structural modelling (SM) is systematic and domain model-based approach that can be used in system structural modelling. The goal of SM is to support consecutive, structural system analysis, design and reasoning about systems using acquired knowledge. There are two different and integrated paradigms in structural modelling: a) morphological structure model (MSM) and b) functional structure model (FSM). MSM lacks characteristics which are necessary for diagnostic and predictive reasoning. Therefore functional representation and transformation algorithm to transform acquired knowledge from MSM into FSM were created. Functional structural model can be derived in the space of functions, behaviour, and parameters (Grundspenkis 1997; Grundspenkis 1999). Additional attributes and specific cases for model transformations are acquired using continuous SM approach. Representation of logic of the relationships in structural models offers a qualitative, more detailed knowledge about system and allows identifying relationship compatibilities.

In Structural modelling, the logic is defined in a morphological structural model and transformed to functional structural model. To minimize an expert's workload and time during knowledge representation process, it is essential to maintain and transform newly acquired logical operators automatically. Logical operator visualisation and interpretation in the MSM as well transformations from the MSM to the FMS in a space of functions were overviewed previously by (Zeltmate and Grundspenkis 2008). In this paper, I consider transformations from MSM to FMS in the



space of behaviour and parameters and corresponding examples. The transformations in the space of behaviour are analogous to the FMS transformations in the space of functions.

The paper overviews specific structural model transformation cases in the space of behaviour and parameters where the logical operands are used. This paper has five sections which are organized as follows. In section two concepts of structural models are presented to introduce with formal aspects of structural modelling approach and used notations. In section three graph transformation concepts and transformation examples are given. The section four includes an example of a cooling system of an internal combustion engine representation. Three structural models and corresponding transformation cases are shown. In conclusions a short summary of paper is given and tasks for future work described.

## 2. THE CONCEPTS OF STRUCTURAL MODELS

In Structural modelling approach structural and functional aspects of a system are unified in one framework. Using several types of interconnected models allows to observe and represent a system from different viewpoints (Grundspenkis 1999; Abu-Hanna and Jansweier 1994; Oliva 2004). Structural models are visualized as directed graphs (Tutte 2001) where the system components are depicted as nodes and relationships between components as directed edges. Hence nodes represent structural components of models. Structural component can be any collection of elements that is acquired in the process of conceptual system decomposition. MSM represents the set of objects and relations between objects at the system decomposition level selected by experts. In the morphological structural model, three structural components are used to encapsulate the domain knowledge: objects, flows, and contacts (see Figure 1).

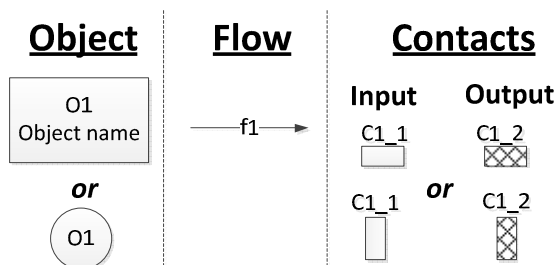


Figure 1: MSM components

Objects represent small and independent units of the domain knowledge. Although in the MSM objects are nodes, each object can be decomposed and inspected in detail. A flow is an interaction path between one object's output and another object's input (Grundspenkis 1997). Flows are used to represent functional and causal dependencies in system. Flows are identified by a small letter 'f' followed by a number. Each flow can perform one or more actions and execute one function. The knowledge about a flow includes the

function's name in the appropriate frame. This enables the connection between structural models and automates model transformations. An object's contacts include the information about the Input/Output state, the name of the flow, the function, and behaviour states. Contact names are acquired in the knowledge acquisition process. A contact's context or meaning does not depend on its vertical or horizontal position. If interpreted in an application domain, abstract objects correspond to components of a given system, and contacts represent their inputs and outputs (Grundspenkis 1997).

Structure, functions and behaviour are interconnected in a system and influence each other. That is why it is necessary to create and view corresponding structural models as interconnected. In SM the transformation is formal method that relates system morphological and functional aspects. The morphological structure model can be transformed into a functional structural model in function, behaviour, and parameter space (Grundspenkis 1997; Grundspenkis 1999; Zeltmate and Grundspenkis 2008). This means that transformation provides iterative represented knowledge transfer from MSM to FSM. In order to perform transformation several steps and rules must be considered (Grundspenkis 1997).

The behaviour state and parameters affect the action and time, during the function execution, and describe the system dynamics. From the structural point of view, behaviour state specifies constraints or boundaries, within which the function is implemented under certain circumstances. The behaviour state is a qualitative characteristic of a flow and describes the way how flow acts while it maintains steady stream from one object to another. Flow executes some function and implements some process in a system. The parameters are variables with certain values that specify the function behaviour. The topology of the functional structural model in the behaviour space (FSM BS) is derived from the MSM automatically (Grundspenkis 1997). The basic structural components for the behaviour representation are: behaviour state, behaviour, and flow (see Figure 2).

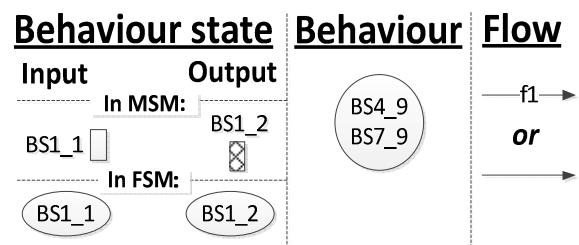


Figure 2: FSM BS components

Behaviour states (see Figure 3) can be represented in both structural models (MSM and FSM). They encapsulate the knowledge about the behaviour of a system that operates under "normal" conditions. When an object performs an action with a specific function, a determined behaviour is applied. The behaviour

expresses itself in the action, when the function is executed. It can be set at the output of an object in the model and assessed at the input of another object. Comparing the behaviour states at the output and input, it is possible to diagnose system functioning faults (Grundspenkis 1997; Capiluppi 2007). Accordingly, behaviour states for each system component are identified and described in the component's (object's) contacts. The representation of the behaviour states in the MSM is similar to the representation of contacts. Each contact has more than one possible behaviour states, but it can be only in one state at a given moment. That state can be recorded only in one structural model.

In FSM behaviour states are visually represented using ellipses. Behaviour states specify the flow at in inputs and outputs of a given object. If a system functions correctly, the behaviour at the beginning of flow (output) is the same as at the end of flow (input). A formal model transformation behaviour states in FSM BS into pairs. The transformed states define the behaviour of the function (see Figure 3). First, the input behaviour state is represented, followed by the input behaviour state. The behaviour includes behaviour state set that describes the flow representing the connection between two behaviour states.

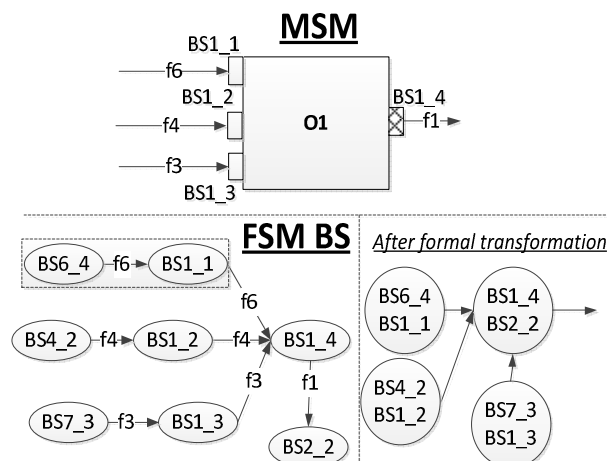


Figure 3: Behaviour representation in MSM and FSM

Each behaviour state has a parameter set. A parameter has a value or a set of values that characterize the state of the behaviour. A parameter can be linked to other parameters. When the system is described, external and internal relationships between the parameters are identified. Identifying parameters and relationships between them is the main task of building an explicit functional structure model in the parameters space (FSM PS). The FSM PS is used in the structural modelling approach to represent system dynamics and enables diagnostic reasoning (Grundspenkis 1997). Logical operators for external relationships are acquired automatically from the MSM, but for the internal relationships the logic must be acquired and specified by an expert. Three components are used to represent parameters in the FSM PS when the system functions correctly: parameter (P1\_1), parameter set (PS1\_1), and

flow (f1; see Figure 4). When the system functions with faults (Grundspenkis 1997, Capiluppi 2007), an additional element—defect—is used. Defect is a cause which can create fault in object behaviour and as a consequence a failure of system functioning.

In a space of parameters (Grundspenkis 1997) the decomposition of the parameter set is considered. First step in the FSM PS development is to acquire a set of parameters from the MSM (or FSM BS), which is done automatically. At the next step, each set is decomposed. An expert represents the knowledge about the parameters and relations between them in a frame hierarchy. In the FSM PS diagrams, the parameters are identified by a letter 'P', followed by symbols from the parameter set (for example, 6\_4) that correspond to the object number and the parameter set, followed by an additional number for each parameter.

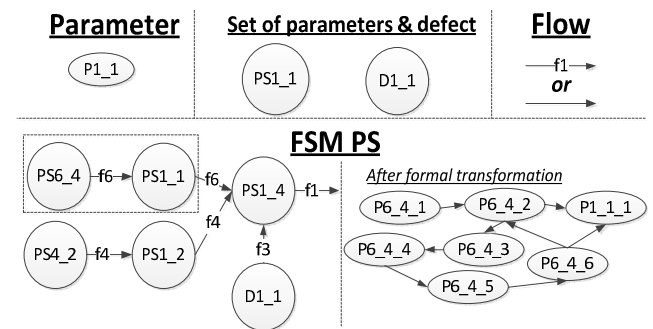


Figure 4: Parameters and corresponding elements

At least one member of the parameter set has a relationship with external parameters. An external parameter is a parameter, which can be found in another behaviour state for the same or another object. Accordingly, an internal parameter is a parameter, which is a part of the object's behaviour state. When the parameter space is decomposed, we first describe the parameters, which are connected with an external parameter in the output contact behaviour state (OBS). The parameters, which are connected to an external parameter in the input contact behaviour state (IBS), are described last.

If system functions faulty then acquired structural model differs from the structural models for a normally functioning system. The FSM PS must be extended using element defect in the model representation. If the defect which have caused faults does not seriously affect the system's functioning (that is, does not cause system failures), then only the FSM PS differs. More serious functioning problems (say, a lost system component or relationship) also appear in other structural models. The defects are usually caused by misbehaving objects located in higher or lower decomposition levels. These objects can be the system's components or external objects. A defect is identified similarly to the parameter set, because it can be caused by the object parameters that logically appear in the OBS. A defect is identified by a letter 'D', followed by the number of the object that is affected by this defect,

followed by the symbol ‘\_’ and the number of defect. Defects can be detected by measuring the parameter values. If the values changed and are outside the defined and allowed ranges then the system functions differently. Unexpected changes in the structure models can be observed in two situations: 1) when an expert has not represented all the knowledge about the system or the system has an alternative functioning case and 2) when the system operates with faults. If the sources of the fault are known, the expert can represent about the timing and origins of it in the frame hierarchy. This knowledge can be used in the case of the system malfunctioning to diagnose the cause of fault. Using the FSM PS an event tree can be created to find the defects which caused faults. Change of parameter value is called an event. A causal order of two events is interpreted as a causal relationship. Causal relationship chain allows to generate event tree structure (Grundspenkis 1997; Grundspenkis 1999). Event tree represents cause and consequence relationships. Since in system exists organization, change in one element impact other related elements. Organization is defined as relationship between two system elements A and B that can be a condition and can impact element C properties, characteristics, values or state (Ashby 2004). Because not all parameters are observed and measured in real systems (especially complex ones), event trees play an important role in diagnostics. Diagnosis allows to detect causes why system behaviour has changed. Logical operators are essential elements in event trees. Transformations are performed to acquire functional representation of system and transfer logic between structural models. Once acquired and represented logic can be transferred also to event trees. Next, I describe possible concepts for the graph transformations.

### 3. GRAPH TRANSFORMATION CONCEPTS

The logic in graphs between arcs is represented using operators (AND; OR). Logical operators in models are used for visual interpretation of structural and causation equations. The notation allows to map flows and to determine causes and expected results for certain actions and changes. The logic and flows combined representations can be interpreted in the “if-then” rule format. The logical operators determine what function, what behaviour, and what parameters are active for the incoming or outgoing flows. Specific symbols represent logical operators in structural models (Zeltmate and Grundspenkis 2008). If there are not complex relationships between flows that require complex logical expressions/rules with parentheses, then the notation is quite simple (see Figure 5.a.–5c.).

A square on the flow or on the relationship/link between flows specifies a logical AND operator and a triangle specify an OR. Symbols ‘,’ and ‘;’ separate flow identifiers and draw attention to the operators AND and OR. Shapes and symbols on or above a flow refer to other flows affected by the current flow (see Fig 5.a and 5.b). For each case that maps a structural component’s relationships, a rule can be designed. If

this notation is implemented in a software tool that supports structural modelling approach then automatically acquired rules can be used to reason about the system. A shape on a relation between flows refers to the incoming flows, and symbols above the shape between the flow names refer to the outgoing flows (Figure 5.c). Figure 5.a corresponds to the case when the flow f1 affects the flow f2. If the current object has only one input and one output flow then there is no need for logical operators. In Figure 5.b, operator OR (a triangle) shows how the outgoing flows f2 and f3 are affected by the flow f1. Similarly, a square and a comma (Zeltmate and Grundspenkis 2008) correspond to the operator AND. In Figure 5.b the link connecting the flows f1 and f2 and the square shows that two flows operator affect two other flows. A line between flows with a vertical crossing dash and parentheses correspond to brackets in logical expressions like it is shown on Figure 5.d.

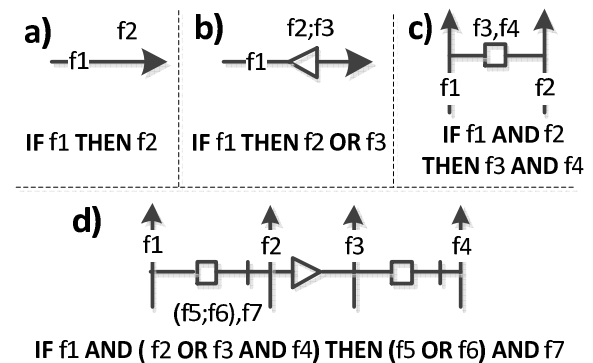


Figure 5: Logical operator representation

The described notation represents logical links between objects in structural models, their relationships, functions, and behaviour. In the case of normal functioning, four basic combinations of input and output flows exist (Zeltmate and Grundspenkis 2008): one to one; one to many; many to one; many to many. There are also situations when only output or only input flows for the object are detected, e.g., when there is a transient object in the system that accumulates or produces something (energy, matter, information). If an object does not consume resources and there are no incoming flows then the corresponding outgoing flows drain after some time. An example of such object is an electric battery. In a malfunctioning system, an object with only incoming or only outgoing flows indicates the presence of a fault, such as the disappearance of or change in some relationships (and perhaps objects). These cases and the corresponding transformations must be also described to diagnose system defects. Transformations from MSM to the functional structural models are similar in the space of behaviour and parameters and in the space of functions. Flows and logic are common components of all structural models. Thus, the knowledge about these components can be shared between models. Structural model transformations (Grundspenkis 1997) enable knowledge

reuse in different contexts. Further in this section several examples are given. The chosen representation form shows possible input and output flow combinations, using logical operators.

The first combination case is simple and no specific notations are used to understand the logical causation relationship (see Figure 6). However the logical operator transfer in FSM BS and FSM PS differs from operator transfer in FSM FS which have been presented previously (Zeltmate and Grundspenkis 2008). The logical expressions for the case (a) are as follows:

1. **IF BS3\_2 THEN BS1\_1; IF BS1\_2 THEN BS2\_1.** An output behaviour state at output (for example, BS1\_2) corresponds to exactly one input behaviour state (for example, BS2\_1). Since these behaviour states are connected with one flow, they are usually equal, and the parameters that specify the behaviour can be equal, too. If the parameters are not equal, their values are at least in a predefined range.
2. **IF (BS3\_2, BS1\_1) THEN (BS1\_2, BS2\_1).** If the flow f1 implements the behaviour of object O3 then the flow affects the behaviour of object O1.
3. **IF PS1\_2 THEN PS2\_1.** If one parameter set is used then it is connected to another parameter set to provide the required functionality. (Both sets belong to normally functioning objects) If two parameter sets belong to one flow then the names of the parameters in the sets are identical and the connection between the sets is strong: even though the two parameter sets belong to two objects, they can differ only in the parameters' values and in automatically assigned id like P1\_2\_1.

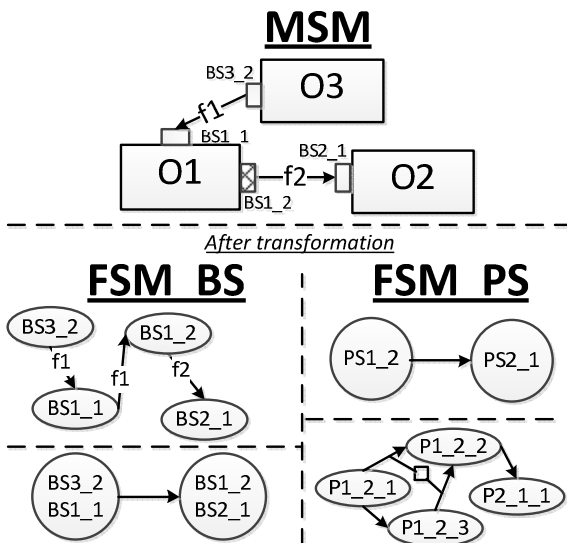


Figure 6: MSM and FSM BS, FSM PS, case (a)

If two parameter sets belong to two flows, the connection between them is weaker and the parameters in the sets can be different. I will not show the decomposition of FSM PS in other combination cases because logical operators for the parameter internal relationships are set by expert in the same manner as for the MSM.

The second combination covers three different cases (Zeltmate and Grundspenkis 2008). In the first two cases there is one input flow and several output flows connected with an AND (b) or an OR (c) of logical operator. In the third case (d) there is one input flow and several output flows connected with logical AND and OR operators (See Figure 7). The logical expressions for the case (d) are listed below:

1. **IF (BS1\_2, BS2\_1) THEN ((BS2\_2, BS4\_1) OR (BS2\_3, BS5\_1)) AND (BS2\_4, BS6\_1).** The behaviour realized by the object O1 affects the object O2 behaviour making O2 behave in two different ways.
2. **IF PS1\_2 THEN PS2\_1 AND IF PS2\_1 THEN (PS2\_2 OR PS2\_3) AND PS2\_4.** When the system functions normally, there is an expression which is true for the system: IF PS1\_2 THEN (PS4\_1 OR PS5\_1) AND PS6\_1. The parameter set determined at the beginning of the flow f1 affects the parameter sets for the flows f2, f3s and f4.

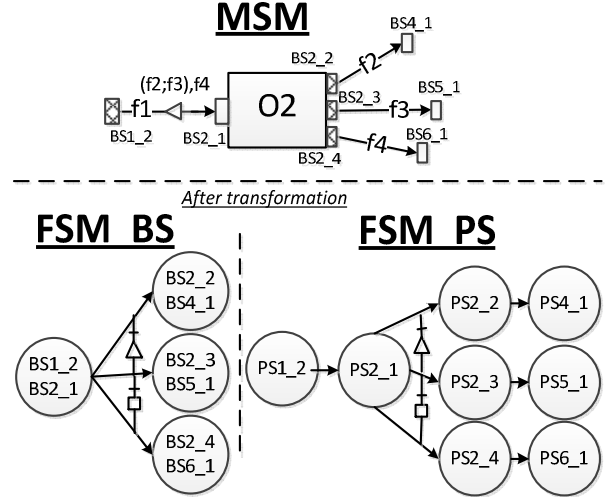


Figure 7: MSM and FSM BS, FSM PS, case (d)

The third combination can be divided into three different cases (Zeltmate and Grundspenkis 2008), too. These cases are common in situations when several input flows affect one output flow. In this paper the example of case (f) is represented (see Figure 8). The logical expressions for the case (f) are as follows:

1. **IF ((BS4\_2, BS2\_1) OR (BS5\_2, BS2\_2)) THEN (BS2\_3, BS3\_1).** Object O2 behaviour can be affected by the other two object (O4

and O5) behaviours. Because of the OR operator, only one of the objects (O4 or O5) affects O2.

2. **IF PS4\_2 THEN PS2\_1; IF PS5\_2 THEN PS2\_2; IF (PS2\_1 OR PS2\_2) THEN PS2\_3.** When the system functions normally, then the following expression holds: IF (PS4\_2 OR PS5\_2) THEN PS3\_1. The parameter sets at the beginning of the flows f2 or f3 can affect the parameter set for the flow f5.

In the case of a defect both logical operators (AND and OR) can be used between the flows. The OR operator means that the defect affects the functionality of the system only occasionally. The AND operator means that the defect always affects the functionality of the system.

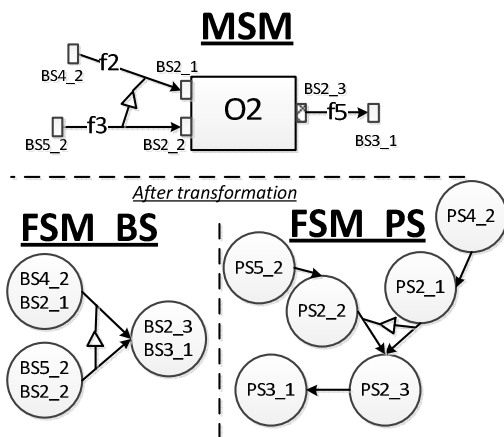
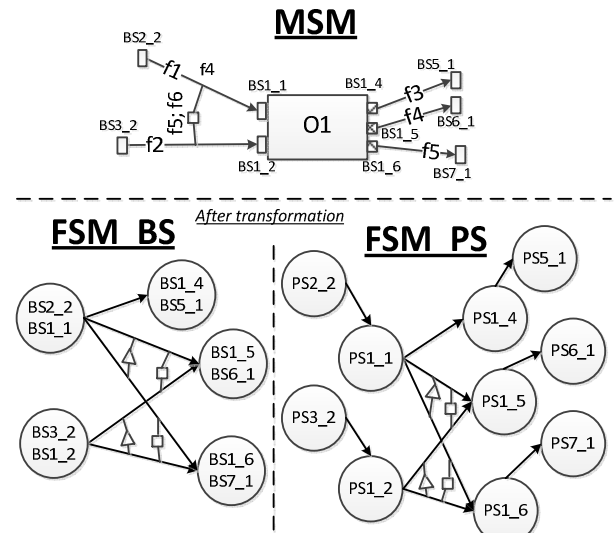


Figure 8: MSM and FSM BS, FSM PS, case (f)

The fourth combination of flows is the most general one. It covers the case of several input flows and several output flows, all possibly linked with various logical operators. This combination case has several sub-cases (Zeltmate and Grundspenkis 2008). I present one of them in Fig 9.

The following logical expressions hold for the case (m):

1. **IF (BS2\_2, BS1\_1) THEN (BS1\_4, BS5\_1) AND IF (BS2\_2, BS1\_1) AND (BS3\_2, BS1\_2) THEN (BS1\_5, BS6\_1) OR (BS1\_6, BS7\_1).** The behaviour implemented by the objects O2 and O3 affects the object O2 behaviour.
2. This complex case has several expressions:
  - (IF PS2\_2 THEN PS1\_1) AND (IF PS3\_2 THEN PS1\_2) AND (IF PS1\_1 AND PS1\_2 THEN PS1\_5 OR PS1\_6) AND (IF PS1\_5 THEN PS6\_1) AND (IF PS1\_6 THEN PS7\_1)
  - (IF PS2\_2 THEN PS1\_1) AND (IF PS1\_1 THEN PS1\_4) AND (IF PS1\_4 THEN PS5\_1)
3. Also two other expressions are true for the system in the case of normal functioning:
  - (IF PS1\_1 AND PS1\_2 THEN PS6\_1 OR PS7\_1)
  - (IF PS1\_1 THEN PS5\_1)



The MSM alone cannot be used to determine the influence of defects, because the defects can happen in the lower or higher level of the hierarchy. If a flow ends and there are no more incoming or outgoing flows from this object, then one of the following is true:

1. the object is connected to other objects beyond the system boundary and the environment is not shown in the MSM;
2. the system is not represented correctly; for example, an abstraction level shows objects or flows that belong to other decomposition levels;

3. there is a defect in the system caused by the expert's improper system understanding.

Determining all structural models and all rules for all cases in a system manually takes a lot of time and is costly. Using several structural models and a transformation methodology (Grundspenkis 2002) makes automated knowledge transfer possible (this includes the knowledge about the logical operators). Using logical operators in the FSM in the space of behaviour is similar to that in the space of functions (see Figure 11).

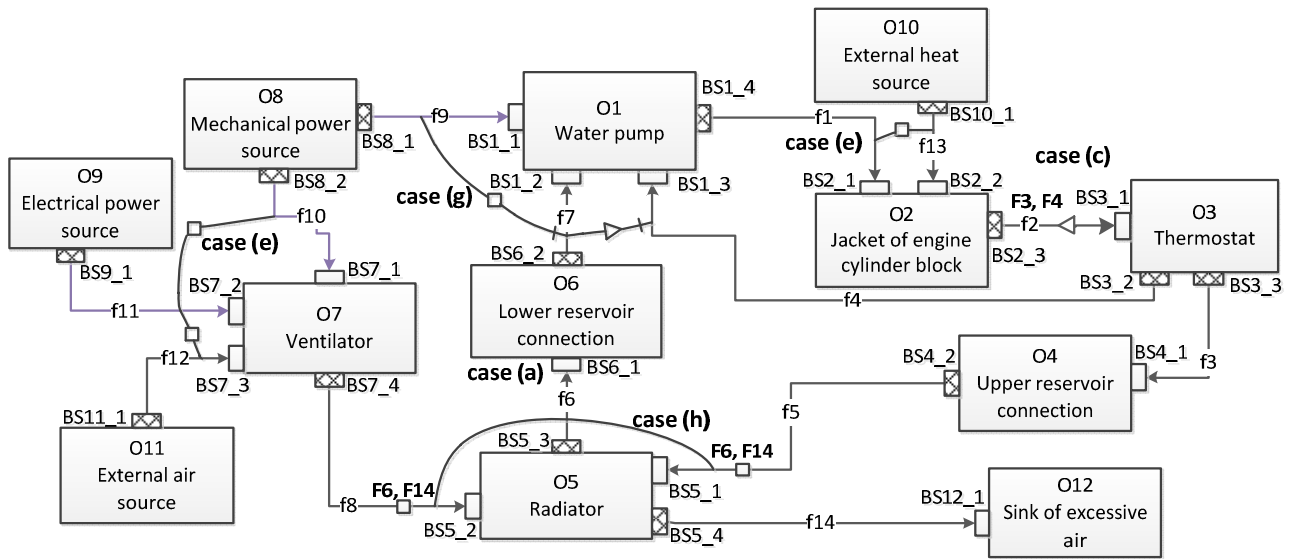


Figure 10: The morphological structure of an ICE cooling system

Each node in the diagram represents the behaviour state of the outgoing flow (above) and the behaviour state of the incoming flow (below). In the case of the normally functioning system these two states are strongly connected: they are either equal or at least similar. The behaviour state order in a node determines the direction of the flow. One can notice that there is no

explicit connection between nodes (BS8\_1, BS1\_1) and (BS8\_2, BS7\_1). However, there should be a connection, because all behaviour states of an object are somehow interconnected. It is implied that there is a logical AND operator between these nodes.

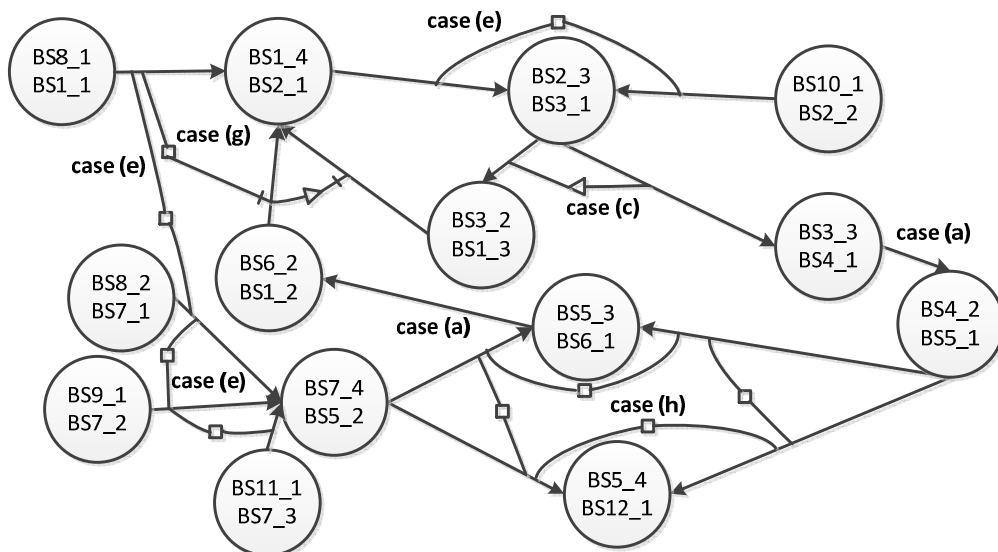


Figure 11: The functional structure model of an ICE cooling system in the space of behaviour



The next structural model that can be produced by the transformations is a functional structural model in the space of parameters. This structural model is similar to the FSM in the state of behaviour before grouping the behaviour states (see Figure 12).

Additional rules can be obtained in the parameter space. An example of an additional rule is: IF PS8\_1 AND (PS6\_2 OR PS3\_2). An FSM PS represents causation relationships between parameter values that are connected with external parameters. Two parameter sets for each flow fully reflect the changes in the system in the case of a defect or rule changes. In the former case, there will be at least two flows: one caused by the defect and the other caused by a system object—usually

connected with a logical AND operator. It is possible for some flows to change direction because of a defect. This change of direction can be detected through the change in parameter sets if logical operators are used: if the current values of the parameters differ from the original values, then one can create an event tree (Grundspenkis 1997) and determine the cause of the problem. If no logical operators are used, it is not possible to obtain correct rules that describe the system's functions, and build the causation chain. The functional structure model in the space of parameters makes it possible to speculate about the consequences of various parameter changes, such as made on purpose or in the case of faults.

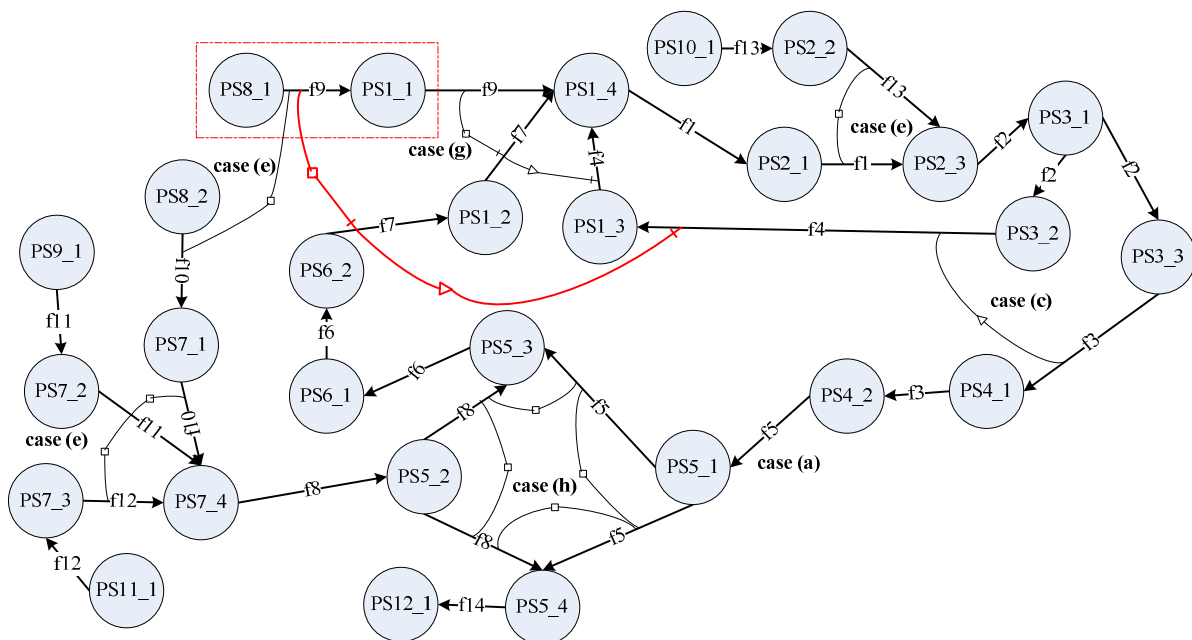


Figure 12: The functional structure model of an ICE cooling system in the space of parameters

## 5. CONCLUSION

Structural modelling is a flexible systematic approach to representing a normally functioning system, subject to changes, using a frame hierarchy. The knowledge about the system is stored in a knowledge base that is used to construct structural models. Combining several interconnected structural model makes it possible to explore different system aspects. Model transformations reduce time required for knowledge acquisition and representation by transferring logical operators between structural models. The system representation is built by experts and used to explore the system and to reason and make decisions about it. It is important to provide a convenient graphical notation for the knowledge that supports decision-making processes. Logical operators in structural models improve the understanding of system structural components and define relationships between them, which is necessary for understanding the principles of the system functioning.

In this paper, I discussed structural model transformations using logical operators in the space of behaviour and in the space of parameters. The

transformations with logical operators and automatic logical transformations improve the quality of analysis and decision making.

The arrangement of components according to the system rules is essential for reasoning. It should be noted that structural components and functions, including the system rules, are organized in a certain order subject to certain restrictions, making, e.g., some combinations of flows impossible. This partially eliminates possible misunderstanding that can arise in the decision process if only partial knowledge is available. Appropriate tools are needed to automate decision process, structural model construction, and transformations. The future work is connected with the transformation algorithm implementation in the tool which supports structural modelling approach.

## REFERENCES

- Abu-Hanna A. and Jansweier W. N. H., 1994. Modelling Domain Knowledge Using Explicit Conceptualization. IEEE-Expert 9(5), pp. 53-64.

- Ashby W.R., 2004. Principles of self – organizing system, in E:CO Special Double Issue Vol. 6 No. 1/2, reprint, (ed. Goldstein J.), pp.102-126
- Capiluppi M., 2007, Fault Tolerance in Large Scale Systems: hybrid and distributed approaches, Ph.D. Thesis, p. 122
- Grundspenkis J., 1997. Causal domain model driven knowledge acquisition for expert diagnosis system development, In: *Lecture Notes of the Nordic-Baltic Summer School on Applications of AI to Production Engineering*, Kaunas, Lithuania, (Wang K. and H. Pranevicius H., eds.), Kaunas University of Technology Press, pp. 251-268.
- Grundspenkis J., 1999. Reasoning Supported by Structural Modelling. In: *Intelligent Design, Intelligent Manufacturing and Intelligent Management*, (Wang K. and Pranevicius H., eds.), Kaunas University of Technology Press, Technologija, pp. 57-100.
- Grundspenkis J., 2002. Reasoning in structural model-based diagnosis. In proceedings of the 4th Int. Conf. on Quality, Reliability and Maintenance, QRM2002, Oxford, Professional Engineering Publishing, London, UK, pp. 295-298.
- Kresken T., 1996, System Modelling. Technology Briefing Report, Sintef, Renaissance Esprit Project 22010, p. 11.
- Oliva R., 2004. Model structure analysis through graph theory: partition heuristics and feedback structure decomposition. *System Dynamics Review*. Vol. 20, Issue: 4, pp. 313 – 336.
- Tutte W.T., 2001. *Graph Theory*, Cambridge University Press, New York, p 333.
- Zeltmate I., Grundspenkis J., 2008. Formal Method Of Functional Model Building Based On Graph Transformations. In proceedings of the International Workshop on Modeling & Applied Simulation, International Mediterranean and Latin American Modeling Multiconference, pp. 140-147

#### **AUTHORS BIOGRAPHY**

Ieva Zeltmate is researcher at the Department of System Theory and Design of Faculty of Computer Science and Information Technology in Riga Technical University. At the moment she is working on the final version of her thesis. Her scientific interests are associated with system analysis, design and development of knowledge representation systems, structural modelling, graphs, networks, complex and intelligent systems and knowledge representation within frames.

# A SURVEY ON PERFORMANCE MODELLING AND SIMULATION OF SAP ENTERPRISE RESOURCE PLANNING SYSTEMS

Manuel Mayer<sup>(a)</sup>, Stephan Gradl<sup>(b)</sup>, Veronika Schreiber<sup>(c)</sup>, Harald Kienegger<sup>(d)</sup>, Holger Wittges<sup>(e)</sup>, Helmut Krcmar<sup>(f)</sup>

<sup>(a) (b) (c) (d) (e) (f)</sup> Chair for Information Systems, Technische Universitaet Muenchen,  
Boltzmannstr. 3, 85748 Garching, Germany

<sup>(a) (b) (c) (d) (e) (f)</sup> {mayerm, gradl, schreibv, kienegger, wittges, krcmar}@in.tum.de

## ABSTRACT

IT industry faces the need of a robust, reliable and scalable architecture to support enterprise-scale IT systems. One of such complex systems are enterprise resource planning (ERP) systems. ERP systems support company-wide processes and are considered as a critical success factor for a reliable business operation. Despite the importance of performance prediction of IT systems shown in literature, there is not much work done in the context of ERP systems. This paper presents the results of an online survey conducted in December 2010. 36 IT decision makers of various industries and company sizes took part in this survey, which contained questions about whether and to which degree performance modelling and simulation is currently implemented in companies, which tools are used to measure the performance of ERP systems, and which requirements are expected.

Keywords: performance measurement, performance modelling, simulation, ERP systems

## 1. INTRODUCTION

With an increasing number of functionalities added to systems by manufacturers and system developers, modern computer systems are becoming more and more complex. One problem that arises is how to choose the right system and components for a certain problem. According to Risse (2006), the selection process is typically driven by different factors, such as functional requirements, performance demands and economic constraints. While architecture design decision-making involves addressing functional requirements, selecting necessary software (components) in the context of standard software, and considering tradeoffs due to the presence of economic constraints, questions arise about how the system performs, how the system scales if expected workload increases, which components might be potential bottlenecks, and what the system evaluation criteria would be. The objective of (technical) performance evaluation is to give answers on these questions and to present techniques to get the performance values (e.g. Jain 1991; Sauer and Chandy 1981).

The literature shows the importance of performance evaluation of IT (e.g. Robertazzi 2000). However, there are only a few publications regarding performance evaluation of ERP systems (Rolia et al. 2009). Therefore, the aim of this work is to analyze the status quo of ERP performance evaluation and prediction in companies' data centres. For that purpose, essential performance evaluation techniques have been examined. Based on this knowledge, an online survey has been developed. The focus concentrates on the following main questions:

- Which demands do companies have regarding performance evaluation of ERP systems?
- Is performance of ERP systems modelled and predicted in companies, and if so, which methods and tools are used?
- Are companies satisfied with the reliability of current performance evaluation techniques used?

The remainder of this work is as follows: after giving a short overview of performance evaluation techniques in section 2, related work concerning ERP systems is presented in section 3. Section 4 explains the structure of the online survey and gives a summary of the results. Section 5 outlines interesting issues for future work.

## 2. THEORETICAL BACKGROUND

According to Ferrari (1986), the field of performance evaluation dates back to the thesis of Alan Scherr (1965). Since then the discipline of performance evaluation has been addressed by many different text books, for example by Sauer and Chandy (1981) or Raj Jain (1991). Following Hu and Gorton (1997), the factors *functionality*, *reliability*, *speed*, and *economic efficiency* have to be considered in conjunction with performance. In the majority of cases, the first two factors, functionality and reliability, are addressed by system designers. As a result, the main focus of performance evaluation community is speed and economic efficiency. While speed is often described by response time and throughput, economic efficiency

reflects the need to design and implement a system at the lowest cost.

To evaluate the performance of a system, first it is necessary to define some criteria or metrics. The right selection of the metric depends on the scenario, for example, in a user oriented analysis response time is of more interest than throughput. On the other hand, performance evaluation of networks mostly concentrates on the throughput achieved.

Once the metrics, their relationships and effects on performance parameters are known, it is equally important to select a proper workload in a particular environment. According to Jain (1991), workload types can be divided into single instruction, instruction mixes, kernels, synthetic programs and application benchmarks. Single instruction, instruction mixes and kernels are primarily used for hardware-related performance evaluations. Synthetic programs (also called synthetic benchmarks), such as the Zachman test for SAP ERP systems (Boegelsack, Wittges, and Krcmar 2010; Kühnemund 2007), are designed to simulate real workloads. Their only objective is to consume system resources, but often they are too simple to accurately reflect real system issues, such as unrepresentative disk or memory references.

Finally, after choosing the metrics and workload, the right evaluation technique is fundamental to get significant results. Most authors in literature, for example Heidelberger and Lavenberg (1984) or Jain (1991), distinguish between three different techniques for performance evaluation: measurement, analytic modelling and simulation. Ferrari et al. (1983) merge analytic modelling and simulation as one technique, since both require the construction of a system model. The selection of the right technique depends on the design stage of the system.

Compared to other performance techniques, measurement provides the highest accuracy at highest cost, but can only be used after the system has been built. There are different purposes for system measurement: to get information to characterize and model workloads, to validate system models, and to get insights into system behaviour to improve system performance. Typically, a monitoring tool is used for data collection. Snodgrass (1988) defined monitoring as "the extraction of dynamic information concerning a computational process, as that process executes". The main challenge in monitor design is to minimize the observation impact to the performance behaviour of the system, since a monitor requires a certain amount of resources every time it is activated. (Ferrari, Serazzi, and Zeigner 1983)

Following Trivedi, Haverkort, Rindos and Mainkar (1994), analytic modelling can be used at very early design stages but provides only a limited accuracy. Jonkers (1994) divides performance modelling formalisms into two classes: deterministic and probabilistic. In deterministic models, all quantities such as timing parameters are fixed, while probabilistic models allow some degree of uncertainty. The latter

enable the solution of models that would otherwise be analytically intractable, due to the assumption of certain time distributions. Well-known representatives of probabilistic models are Markov Chains (Trivedi 1982; Trivedi et al. 1994), Petri Nets (Peterson 1981) and Queuing Networks (Bolch et al. 2006; Kleinrock 1976a; Kleinrock 1976b; Lazowska et al. 1984; Tijms 2003)

Higher accuracy but associated with higher costs can be achieved with simulation. In summary, simulation imitates the operation of a real-world process or system over time (Banks et al. 2004). A simulation can take place at any point in the life-cycle of the product. A major advantage of simulation over analytic modelling is that it can be used to create very detailed, accurate models. The other side of the coin is the fact that detailed models are often time consuming and difficult to design. A wide variety of simulation types exist and can be categorized in stochastic or deterministic and static or dynamic simulation (Jain 1991).

### 3. RELATED WORK

According to Rolia et al. (2009), there is not much work done in the context of performance evaluation and prediction of ERP systems. Rolia et al. (2009) investigate response time behaviour of a SAP ERP system using Layered Queuing Networks (LQN) models. The workload is composed by a fixed mix of sales and distribution requests. The requests are repeated cyclically 20 times, which corresponds to experiment duration of about 40 minutes. The results presented show that the LQN model offered mean response time predictions within 15% of measured values for a wide range of load levels. Gradl et al. (2010) pursue a similar approach for ERP business process modelling. A case study of an existing production planning process shows how LQN models can be exploited as a performance analysis tool.

Seelig et al. (2008) discuss performance evaluation techniques by comparing the results of one analytic and one simulative approach. The environment used in this work is a SAP Web Service application. The authors conclude that the simulation approach is suitable even in a very early stage of software development, enabling the software architect to identify potential hotspots prior to actually implementing the software components. On the other hand, quantitative models not only validate the simulation results, but also enable the architect to evaluate variants of the modelled component without much additional effort.

An evolutionary model generation for ERP performance simulation is presented by Tertilt et al. (2010). In performance models, some of the components in complex ERP systems are handled as black boxes (e.g. due to intellectual property). To increase the prediction accuracy of these components, an evolutionary algorithm is used. First results of the prototypical implementation showed the feasibility of the depicted approach as long as the data used for modelling was equally distributed. However, if the

measured performance data is unbalanced, the error increases significantly.

Besides the analytic and simulative approaches mentioned, different efforts have been made to measure performance of ERP systems. Jehle (2010) analyzes the performance behaviour of a portal system by SAP. Performance measurements are effected on identically configured test systems, both on native and virtualized environments. The results show that under certain circumstances a virtualized environment even increases the performance of the portal system being studied, namely, if the test workload is moderate and causes excessive Input/Output (I/O) loops. Similarly, Boegelsack, Wittges, and Krcmar (2010) investigate the scalability and performance of a virtualized SAP system on the basis of a quantitative approach and gives recommendations how to configure a SAP system for heavy workload. It is also shown that the average performance of a SAP system increases if a container-based virtualization solution is used, and decreases with a Xen-based virtualization solution.

#### 4. SURVEY PREPARATION AND RESULTS

To understand whether and to which degree performance evaluation is currently implemented in companies, an online survey has been conducted. The main questions mentioned above can be broken down as follows:

- Which demands do companies have concerning performance evaluation of ERP systems?
- Do companies evaluate the performance of ERP systems? If so,
  - Which methods and metrics are used?
  - Which parts or layers of their systems are evaluated?
  - Which tools are used?
- Are companies satisfied with currently used methods?
- How big is the interest in this topic?
- What are the benefits expected?

Based on these questions, the basic structure shown in Figure 1 forms the backbone of the questionnaire.

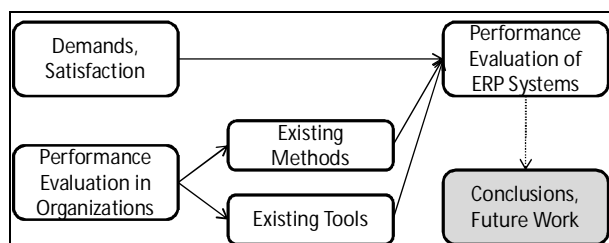


Figure 1: Research Model

##### 4.1. Structure

Besides introduction and conclusion, the final version of the survey contains four blocks of questions

concerning performance evaluation of ERP systems. The first one deals with questions about the general performance evaluation of ERP systems in organizations, the second one with questions about the concrete system, and the last two ones contain questions about demographic data of organizations and persons responding to the survey.

#### 4.2. Data Collection and Demographics

The online survey was initiated on two big internet platforms in December 2010 and continued for a 6-weeks period until January 2011. The chosen platforms were the German-speaking SAP User Group (DSAG) and the SAP Developer Network (SDN).

These two platforms were chosen for the survey due to the composition of their members. Both are specialized platforms to SAP ERP systems. We received 36 fully completed and usable questionnaires from 116 participants, which equals to a return rate of 31.03%. This sample is not as big as we had expected, but still sufficient for a reliable evaluation. Due to the limit theorem, the normal distribution is sufficiently exact for a sample of  $n > 30$  (e.g. Pal and Sarkar 2005).

The industry most represented was IT services (25%), followed by public administration (19.44%), manufacturing (11.11%), banks, assurances and financial services (11.11%), automotive (8.33%), mechanical engineering and construction (8.33%), and others (16.68%). Companies with less than 2000 employees represented almost 60% of the participating organizations. Most answers were received from IT consultants (33.33%), followed by IT managers (27.78%) and IT project leaders (27.78%). 97% of the participants of the survey are long year veterans regarding ERP systems. It can be expected that these results are based on a high level of experience.

#### 4.3. Results

The survey shows that more than 97% of the interviewed persons are very open-minded to performance evaluation of ERP systems. About two-thirds of the participants believe that performance evaluation is very important. All participants measuring performance use common metrics, namely, response times (100%), capacity utilization (100%), and throughput (37.5%).

##### 4.3.1. Demands, Expectations

The participants were asked about their demands on performance evaluation of ERP systems. The result is shown in Figure 2.

To get a better understanding of which factors are most important for the respondents, the response options A (not important at all) to E (highly important) were given weights from 0 to 4. Result F (no answer given) is also given the weight 0. Subsequently, the percentages have been multiplied by the weight and normalized.

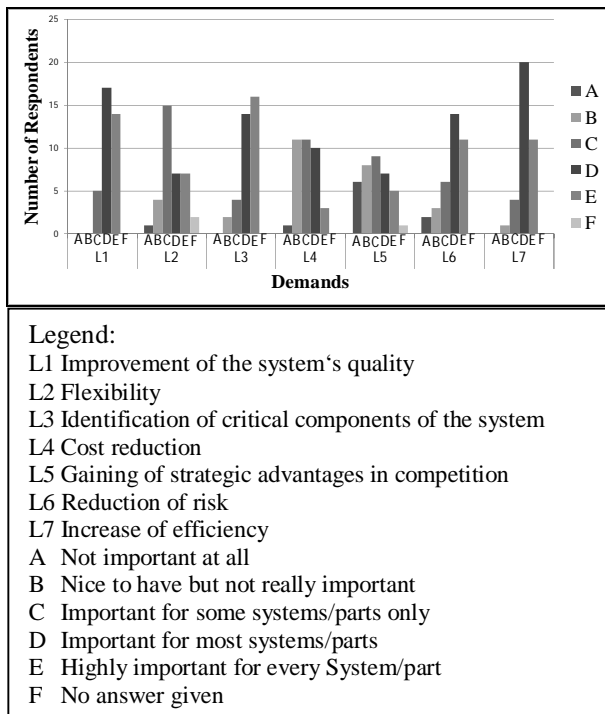


Figure 2: Demands on Performance Evaluation of ERP Systems

The result showed that L1 (to improve the system quality) and L3 (identification of critical system components) are most important for the respondents, closely followed by the requirement L7, to increase efficiency. Two other demands, which were not included in the poll, but were added by the respondents, are operational safety and system stability.

#### 4.3.2. Performance Evaluation of ERP Systems in Organisations

For the further course of the questionnaire it was important to check if companies already evaluate the performance of their ERP systems (see Figure 3).

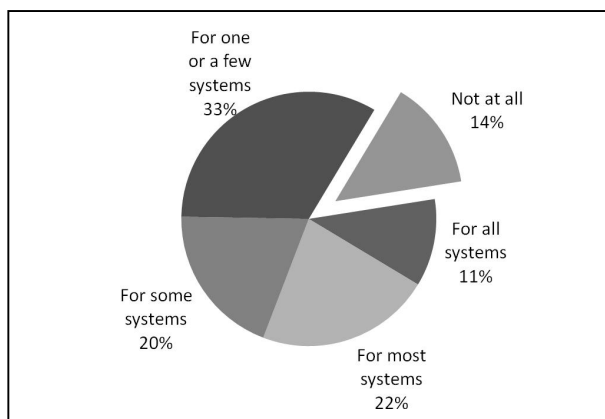


Figure 3: Performance Evaluation of ERP Systems in Organisations

About 86% of participants' companies have already evaluated performance of ERP systems at least once. On the other hand, those who use performance

evaluation on a regular basis for most or all of their systems are clearly underrepresented with only about 33% of the participants.

In this context, it would be interesting to see, whether there are dependencies between these results and the company size or IT budgeting. Unfortunately, the sample size does not allow giving statistically relevant statements. The participants that are not evaluating the performance abandon it due to not enough resources (60%) or competencies (40%).

#### 4.3.3. Established Tools

Only half of the participants of the online survey are using tools and methodological approaches to measure performance. All considered companies use inherent tools of the ERP system. The survey did not ask for the reason, but it can be assumed that these ERP integrated tools are used not only to avoid potential compatibility problems and version release dependencies when using third party tools, but also to avoid additional licensing costs.

Another question about tools is shown in Figure 4. It has been asked, how flexible these tools are in respect of application field, functionality and modifications.

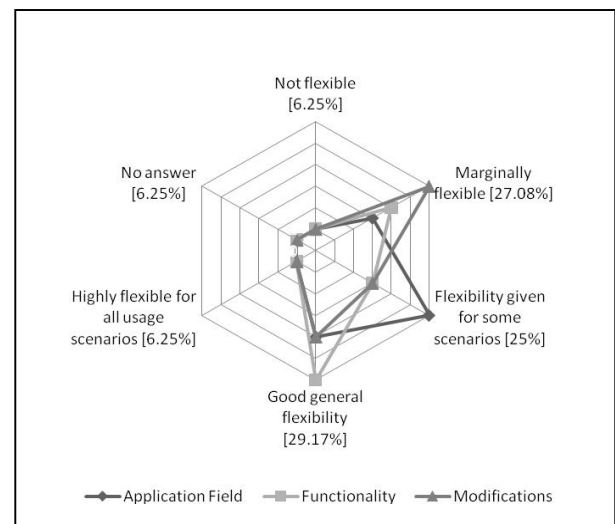


Figure 4: Flexibility of Performance Tools

More than 58% of the respondents are not or only partially satisfied with the flexibility of the product. As we had seen before, all of them are using ERP inherent performance tools, although the given answers indicate that they are not satisfied with the flexibility of their tools. Normally, one would assume that the customer would consider switching to a different tool in such cases. But it seems that the assumption taken above is backed by this result. In addition, the assumption is also supported by the fact that only 12.50% had switched the performance tool in the past.

In conclusion, the respondents use the offered ERP integrated tools and products but are not really satisfied. Interestingly, the major dissatisfaction comes with the suitability for performance measurements of future workload characteristics (e.g. with increased loads).



#### 4.3.4. Established Methods

The next question concentrated on methods used for performance evaluation purposes. Here, multiple answers were possible, because the organizations might use different methods in different phases in the life cycle of an ERP system, or they could use different methods to compare these results and thus get a probably more reliable overview of system's performance.

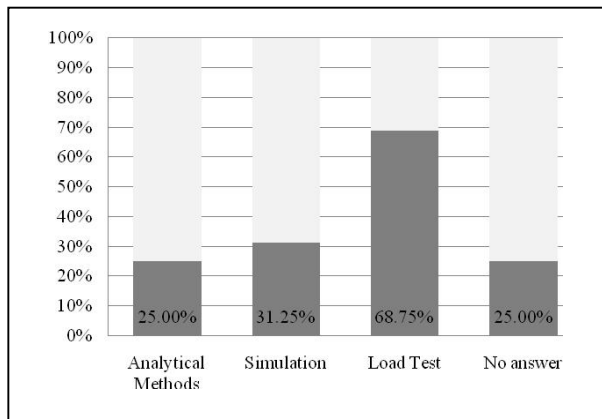


Figure 5: Methods Used for Performance Evaluation

Figure 5 shows that almost 70% of the respondents are using workload tests, 25% use analytical models and 31.25% use simulations. One of the reasons might be the concentration on performance behaviour of the current system configuration, without the intention to further investigate alternative scenarios with different workloads. In addition, we asked the participants about the drawbacks of their currently used methods. Most of them said that on the one hand these methods are not entirely flexible and on the other hand there is no end-to-end view across all layers.

In retrospect, interesting aspects for future data acquisitions are the reasons why there are no further tools other than ERP integrated instruments used or developed, and which functional improvements and additions are required to cover all aspects to entirely evaluate performance of the whole system and software stack.

#### 4.3.5. User Satisfaction

Another important question concerned with the satisfaction of reliability, namely, the robustness of tools versus failures and the reliability of performance results.

As shown in Figure 6, no respondent is absolutely satisfied with reliability. This confirms the already gained knowledge about the weaknesses of currently used tools. Surprisingly, one-third of the respondents, who all measure performance regularly and most likely did not switch the tools in the past, question the reliability. The lack of confidence correlates to the results about functional weaknesses of currently used tools.

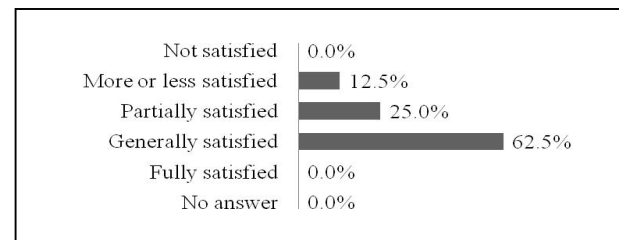


Figure 6: User Satisfaction Concerning Performance Evaluation Reliability

## 5. CONCLUSIONS AND FURTHER RESEARCH

The survey shows that almost all participants (97%) are open-minded to performance evaluation of ERP systems and 86% of them currently measure performance at least for some components or systems. All of them measure the performance in the phase of operation and maintenance. Almost 70% use workload tests to measure performance, 25% are using analytical models, and 33% are using simulations. Performance predictions for increasing workloads are rarely made. This is reasoned by functional weaknesses of the tools used.

Interestingly, only ERP inherent and no third party or own developed tools are used. These integrated tools only partially satisfy the demands of the participants. Although only the minority of participants is satisfied with the reliability of performance evaluation results, almost no one has changed the method or tool used, due to the lack of know-how, resources, and high implementation costs. Comparing these results with well-established performance evaluation techniques in literature, it is assumed that there are a lot of potentials to improve methods and tools in practice.

The advantages seen in performance evaluation of ERP systems include mainly improvements in efficiency (85.29%), risk reduction (76.47%), increased flexibility (50%) and cost reduction (44.12%). Consequently, further research should be done to combine existing performance evaluation techniques in literature with requirements in practice. To identify the demands in detail, additional data acquisitions are necessary. For this purpose and with the assistance of the expert communities, semi-structured interviews are being considered. Based on these findings, aligned methods for performance evaluation of ERP systems can be developed.

## ACKNOWLEDGMENTS

The online survey has been conducted via [www.2ask.de](http://www.2ask.de). For a copy of the survey questionnaire, results, and analysis please contact the authors.

## REFERENCES

- Dsag* - Deutsche Sap Anwendergruppe. Available from: <http://www.dsag.de> [February 12 2011].
- Sdn* - Sap Developer Network. Available from: <http://www.sdn.sap.com> [February 12 2011].

- Banks, J., Carson, J., Nelson, B. L., Nicol, D., 2004. *Discrete-Event System Simulation*. Englewood Cliffs, NJ, USA:Prentice-Hall.
- Boegelsack, A., Wittges, H., Krcmar, H., Year. Scalability and Performance of a Virtualized Sap System. *Proceedings of the 16th American Conference on Information Systems*, Paper 13. August 12-15, Lima, Peru.
- Bolch, G., Greiner, S., de Meer, H., Trivedi, K. S., 2006. *Queueing Networks and Markov Chains: Modelling and Performance Evaluation with Computer Science Applications*. Hoboken, NJ, USA:John Wiley and Sons.
- Ferrari, D., 1986. Considerations on the Insularity of Performance Evaluation. *IEEE Transactions on Software Engineering* (12:678-683).
- Ferrari, D., Serazzi, G., Zeigner, A., 1983. *Measurement and Tuning of Computer Systems*. Englewood Cliffs, NJ, USA:Prentice-Hall.
- Gradl, S., Mayer, M., Wittges, H., Krcmar, H., Year. Modelling Erp Business Processes Using Layered Queueing Networks. *Proceedings of the 12th International Conference on Enterprise Information Systems*, 255-260, Funchal, Portugal.
- Heidelberger, P., Lavenberg, S. S., 1984. Computer Performance Evaluation Methodology. *IEEE Transactions on Computers* (33:1195-1220).
- Hu, L., Gorton, I. "Performance Evaluation for Parallel Systems: A Survey," 9707, University of New South Wales, Sydney, Australia.
- Jain, R., 1991. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modelling*. New York, NY, USA:Wiley/Interscience.
- Jehle, H., 2010. *Performance Measurement of an Sap Enterprise Portal System in a Virtualized Environment*. University of Technology Munich.
- Jonkers, H., Year. Queueing Models of Parallel Applications: The Glamis Methodology. *Proceedings of the 7th International Conference on Computer Performance Evaluation*, 123-138, Secaucus, NJ, USA.
- Kleinrock, L., 1976a. *Queueing Systems, Volume 1: Theory*. Hoboken, NJ, USA:John Wiley & Sons.
- Kleinrock, L., 1976b. *Queueing Systems, Volume 2: Computer Applications*. Hoboken, NJ, USA:John Wiley & Sons.
- Kühnemund, H. "Documentation for Slcs V.2.3.," SAP AG, Linux Lab, Walldorf, Germany.
- Lazowska, E. D., Zahorjan, J., Graham, G. S., Sevcik, K. C., 1984. *Quantitative System Performance: Computer System Analysis Using Queueing Network Models*. Englewood Cliffs, NJ, USA:Prentice-Hall.
- Pal, N., Sarkar, S., 2005. *Statistics: Concepts and Applications*. New Dehli, India:Prentice-Hall.
- Peterson, J. L., 1981. *Petri Net Theory and the Modelling of Systems*. Englewood Cliffs, NJ, USA:Prentice-Hall.
- Risse, T., 2006. *Design and Configuration of Distributed Job Processing Systems*. Thesis (PhD). Technische Universität Darmstadt.
- Robertazzi, T. G., 2000. *Computer Networks and Systems: Queueing Theory and Performance Evaluation*. New York, New York, USA:Springer.
- Rolia, J., Casale, G., Krishnamurthy, D., Dawson, S., Kraft, S., Year. Predictive Modelling of Sap Erp Applications: Challenges and Solutions. *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, 1-9, Pisa, Italy.
- Sauer, C. H., Chandy, K. M., 1981. *Computer Systems Performance Modelling*. Englewood Cliffs, NJ, USA:Prentice-Hall.
- Scherr, A. L., 1965. *An Analysis of Time-Shared Computer Systems*. Massachusetts Institute of Technology.
- Seelig, M., Kluth, S., Porzucek, T., Copaciu, F., Naumann, N., Kühn, S., Year. Comparison of Simulation and Performance Modelling - a Case Study. *Proceedings of the 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems*, 49-56, Washington, DC, USA.
- Snodgrass, R., 1988. A Relational Approach to Monitoring Complex Systems. *ACM Transactions on Computer Systems* (6:157-195).
- Tertilt, D., Leimeister, S., Gradl, S., Mayer, M., Krcmar, H. "Towards an Evolutionary Model Generation for Erp Performance Simulation," in: *IADIS International Conference on Intelligent Systems and Agents (ISA)*, Freiburg, Germany, 2010.
- Tijms, H. C., 2003. *A First Course in Stochastic Models*. Hoboken, NJ, USA:John Wiley & Sons.
- Trivedi, K. S., 1982. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. Englewood Cliffs, NJ, USA:Prentice-Hall.
- Trivedi, K. S., Haverkort, B. R., Rindos, A., Mainkar, V., Year. Techniques and Tools for Reliability and Performance Evaluation: Problems and Perspectives. *Proceedings of the 7th International Conference on Computer Performance Evaluation*, 1-24, Secaucus, NJ, USA.

# UNDERSTANDING THE PERFORMANCE BEHAVIOR OF A SAP ERP SYSTEM FOR THE USE OF QUEUING MODELS

Stephan Gradl<sup>(a)</sup>, Manuel Mayer<sup>(b)</sup>, Alexandru Danciu<sup>(c)</sup>, Holger Wittges<sup>(d)</sup>, Helmut Krcmar<sup>(e)</sup>

<sup>(a) (b) (c) (d) (e)</sup> Chair for Information Systems, Technische Universitaet Muenchen, Boltzmannstrasse 3,  
85748 Garching, Germany

<sup>(a) (b) (c) (d) (e)</sup> [\[Stephan.Gradl, Manuel.Mayer, Danciu, Holger.Wittges, Krcmar\]@in.tum.de](mailto:[Stephan.Gradl, Manuel.Mayer, Danciu, Holger.Wittges, Krcmar]@in.tum.de)

## ABSTRACT

ERP systems support the management of a company's resources. As a large number of business-relevant processes are supported by ERP systems, the performance and availability of those systems is crucial for the success of a company (Apache Software Foundation 2010). We analyze the response time of 49350 requests. Furthermore, we interpret the system's internal behavior by fetching and analyzing the statistical data. As results we can show that queuing models can be used for evaluating the performance of SAP ERP systems as the response time behavior follows the assumptions of queuing theory, resulting in nearly constant resource consumption per user interaction task, independent of the number of parallel requests. By analyzing the reasons for these results, important insights into the performance behavior of SAP ERP systems for performance analysis and prediction are achieved.

Keywords: ERP, SAP, Performance, Load Test, Measurement, Analysis

## 1. INTRODUCTION

ERP systems support the management of a company's resources. As a large number of business-relevant processes are supported by ERP systems, the performance and availability of those systems is crucial for the success of a company (Krcmar 2009). In particular, we focus on the performance analysis of the SAP Enterprise Resource Planning (SAP ERP, formerly SAP R/3) application (Schneider 2008). SAP ERP is an integrated backend application with tens of thousands of installations worldwide designed for tracking and managing business processes in midsize and large enterprises. From a technical perspective, this application is built on top of a software integration platform that provides primitives to control the concurrency offered by application server and database server, the layered use of servers, asynchronous messaging, and priority scheduling for certain types of processing. According to Jain (1991), there are several classical approaches for capacity planning and performance evaluation of computer systems like measurement (benchmarking and stress testing),

simulation and analytical modeling. To evaluate the performance using simulation techniques, the system has to be modeled. Performance modeling of an ERP system requires deep knowledge about the structure and its performance behavior. To achieve accurate and significant results using queuing models the system has to follow certain performance criteria (Chen et al. 2008):

- The CPU time per user interaction task has to be independent from overall system utilization.
- The CPU utilization has to increase linearly with the number of concurrent load steps.
- The response time has a constant section that is followed by a linearly increasing section, ending in an exponentially increasing behavior.

In the following we describe a case study we performed on an SAP ERP system to analyze the performance behavior of this system when set under heavy parallel. We measure the response time behavior of the system as a black box, and then go a step further and analyze the internal behavior of the ERP system by fetching and interpreting the system's statistical records. Section 2 of this paper provides the research context of this work, while Section 3 provides an overview of the required definitions. In Section 4 we describe how we measure the response time of the system, give a brief overview of the architecture of the system under test and the used benchmark, and point out the method we used to create load. Section 6 then follows with the measurement results, as well as with their interpretation, and the analysis of the statistical records of the ERP system. Finally, in Section 7 we conclude our results and give an overview about the next steps and future work.

## 2. RELATED WORK

The key literature about performance measurement and analysis of (enterprise) software systems are the books of Jain (1991) and Lilja (2000). These authors describe elaborately the whole process of performance measurement, pointing out what performance is, how it is measured, and which factors affect the performance of a software system. We are basing our work on the

definitions made in these books, and adopt them to the fields of ERP. The importance of performance analysis is pointed out by Menascé (2002). An overview of the factors that determine the performance of an application is given by Bailey (2005) and Hollingsworth (2005). For the performance of an ERP system, we refer to (Schneider-Neureither 2004). In this book, the author explains in detail the effect of the SAP architecture and configuration on its performance, focusing on the solution of concrete operational problems. Although the book is written as an administrator manual, it provides a good overview of the factors affecting the SAP system's performance. An overview of existing SAP benchmarks is given in (Prior 2003).

A scientific approach for the measurement of ERP performance behavior – in this case focusing on the effects of virtualization – presented by Jehle (2009) and Bögelsack (2010). While Jehle is focusing on the response time behavior using a load test, Bögelsack (2008) is analyzing the system's internal matters, especially the CPU time, for interpreting its effect on the system performance.

Jin (2007) shows a method for performance prediction of legacy information systems. As the internal architecture of the investigated productive information system is not known, the authors used a method that is based on a black box approach for predicting the technical performance of this legacy information system with historical values. This approach combines benchmarking, production system monitoring, and performance modeling (BMM) by analyzing and correlating the performance values derived from the benchmarks and monitoring. Based on the measurements, a model is created and used for the performance prediction.

In (Rolia et al. 2009), an LQN model for the performance prediction of an SAP ERP system is introduced. In this approach the statistical records provided by the SAP system are used for performance analysis and prediction. In addition, the authors also used CPU values gathered from an SAP tool called saposcol. The workload used is based on a sales and distribution scenario, very similar to the workload that is applied in the SAP SD benchmark. Buffers, both from the applications server and the database, having a significant impact on the overall performance, are not taken into account.

### 3. DEFINITIONS

For measurement the performance of an application, one first has to define what is understood as performance in the given context, and which metric(s) are considered for the representation of an application's performance. Our understanding of performance is best shown by the following definition, taken from (Schneider-Neureither 2004).

Generally spoken, the performance of a data processing system is its ability to match the requirements in response time and throughput.

As already given by this definition, the most popular metrics for an application's performance are response (or execution) time and throughput. Nevertheless there are other, like the number of accesses to a special resource or energy need. In our context of ESOA, the response time as defined by Nudd (2000) and shown in figure 1 the time from the moment the request is sent (T1) until the time, when the response is completely received (T3) – is the more relevant metric, as the service calls are considered to be short running by time critical (in contrast to a batch job, where in general the throughput is more relevant than the single response time).

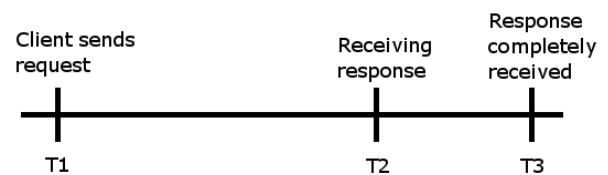


Figure 1: The Structure of the response time (according to Menascé (2002))

Even though it is obvious that the response time and the throughput are connected, in our work we focused on the response time, since the response time is the actual time a user waits while performing a task. On a high dispersion of response times, the throughput could be still eligible, while some high response times are unacceptable (e.g. due to Service Level Agreements).

### 4. MEASUREMENT

Following (Lilja 2000), the most common benchmark strategy is the fixed-computation approach in which the total time required to execute the benchmark is used as the performance metric. The complementary approach is to fix the amount of time, where the total amount of computation completed in this time period is used as performance metric. The most flexible benchmark strategy is to derive a third dimension from some combination of the execution time and the amount of computation completed within this time. In this way (using this third dimension as performance metric), execution time and computation can be kept variable. The Hierarchical Integration Benchmark (HINT), for instance, uses quality improvements per second as performance metric, defined as a function of the problem being solved by the benchmark program. Table 1 summarizes the strategies that can be used in a benchmark program to exercise the system under test.

For our case study we fixed the amount of computation while measuring the time, needed by the sap ERP System to execute it.

As introduced later on, the benchmark consists of the creation of a material master record in the SAP ERP system. For this task, a WebService in the SAP system has been identified, which has been used for creating load on the system. This service is called in parallel by an own implementation of a Java load generator. For measuring the system behavior, we combine the black

box approach described by Kruse (2009), and the glass box approach used by Malik (2010). In the following section, we illustrate the architecture of the system under test, the benchmark, and the load generator.

Table 1: Benchmark Strategies (based on (Lilja 2000))

Time	Computation	Performance Metric
Variable	Fixed	Execution Time
Fixed	Variable	Consumption completed
Variable	Variable	Third dimension

## 5. SYSTEM ARCHITECTURE

To provide an understanding of the ERP system architecture shown in figure 2, we derive the system components from the ERP process step-by-step by analyzing the recorded trace and the abstraction of the trace entries. These components are described in detail in (Schneider 2008). The process step of calling a program involves many components of the SAP system (see figure 2).

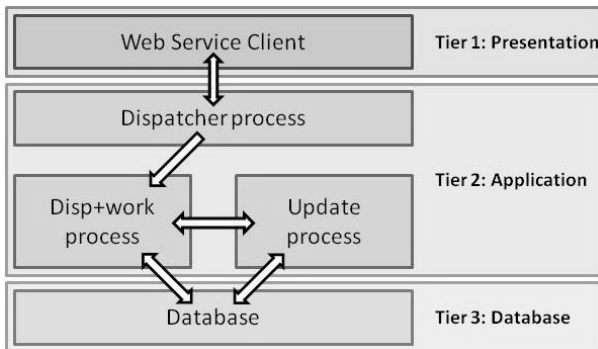


Figure 2: Simplified SAP ERP system architecture

Searching for the program includes access to internal buffers as well as access to the database tables. This access is made by the so called disp+work processes of the SAP system. Such processes are responsible for executing programs, processing user or WebService requests, and accessing the database. Before a request is associated to one of the disp+work processes of the SAP system, a dispatcher process is accessed. The dispatcher process manages all other processes in the SAP system, and his primary task is to assign a user request to a free disp+work process. In our model, we assume the database as a black box.

After the SAP system got the information which program has to be executed, it loads a compiled version of the program from the database and executes it. Sometimes such compiled programs are held in the internal buffers of the SAP system to avoid database accesses.

After the request is processed, the data should be saved to the database. This is done by the disp+work process(es) together with a process called update process. This process receives data and stores it in corresponding database tables.

Simultaneously, a lock on a central table is established, which may be described as a little

repository of all available material master records (MMR) within the system. This lock is not set by the disp+work process itself; it triggers a so-called enqueue process. The only task of the enqueue process is to set locks on any tables in the SAP system, and to manage such locks. After the lock was set successfully, the disp+work process can store the data into the central MMR repository.

### 5.1. Benchmark

For the load run processed in this case study, the Production Planning Integration Case Study (Weidner 2006) has been used. In detail Web Services creating different kinds of material master records, bill of materials, routing, etc. has been chosen. As shown in table 2, these Web Services have an average complexity with read, insert and update statements to the database. Each execution of the case study does not depend on another one and therefore can be executed anytime using the SAP Web Service Interface.

As already introduced, the programs in a SAP ERP System are running on an infrastructure containing dispatching, locking, buffering and database access mechanisms (Schneider 2008). In order to understand the performance of a SAP ERP system, it is necessary to use a workload that uses these components. This leads to different kinds of database queries that characterize the Web Service calls from a technical side.

Table 2: Database accesses for material creation

	DBRows	Read to Buffer	ReqTime
Direct Read	6	222	5134
Sequ. Read	2754	176	6352196
Insert	122	0	12633
Update	1	20	686

Table 2 shows different kinds of database accesses. As the labels Insert and Update are self-explanatory, "Direct reads" are always in the form of "select single" and fetch exactly one row from the database, while queries in the form of "select \* from..." are named Sequential reads. The column DBRows shows the number of rows that are fetched directly from the database without being served by the buffers, while the column read2Buffer shows the number of requests that could be served from caches. The column ReqTime contains the time requests took. This entry does not include requests time served by the caches, since these times can be neglected, according to the technical documentation of the Transaction STAD.

In addition, the workload emulates an existing business process, affecting the already mentioned key components (buffers, locks, database accesses, etc) on the technical layer.

For the load test, an ERP installation with an application server and a database server, both hosted on a physical server with 16 GB of Ram and 4 Cores running at 2.8 GHz, has been used. The application

server and the database were provided in virtual containers using SUN Solaris Zones. The application server was configured with 30 work processes and the database (MaxDB, version 7.7) with up to 150 parallel connections. For the case study, we used an SAP system with customizing and data of the SAP International Demo and Education System (IDES).

The underlying database for this installation, containing one IDES client, has a size of 220 GB and uses Unicode. The database data files are provided on two internal SAS Discs configured in a performance raid with raid level 0.

## 5.2. Load Generator

Load generation is done by the own implemented Java application Load Generator. The Load Generator uses threads for parallelization, and the Axis2 framework (Apache Software Foundation 2010) for calling the web service.

A load run is conducted by a stepwise increasing number of parallel service calls. To minimize the amount of overhead, we initialize the payload once, cloning the object tree for each call, and changing the material number. The load procedure calls the service  $i$  times in parallel for every step  $i$  in the load test (where  $i = 1$  to  $n$ ,  $n$  = the number of maximum parallelism set for the load test), then waits for all results to be stored, and finally increases  $i$  by a given step size (step size 1 in our case). Doing this, it is assured that every sequence, which is a load unit of a certain number ( $i$ ) of parallel requests, is not affected by the request before. For this purpose, a configurable wait time after each sequence has been implemented, too. This short time frame between the sequences is used to fetch the later discussed statistical values from the SAP system, in order to keep these values available and the amount of data, which has to be transferred, as small as possible.

This results in a response time distribution matrix containing the response times for all ( $i$ ) sequences, and a total number of calls of  $m\_calls = (n * (n + 1)) / 2$ .

Experiments showed that the initialization of caches on first requests results in non proportional and unpredictable long runtimes. To avoid these “cold start” effects, we perform a settling phase of three times forty requests before starting the measurement. In this way we assure that all caches are initialized, as can be seen by the cache hit statistics provided by the SAP system.

## 6. RESULTS AND ANALYSIS

In this section we provide two views of the system under test. At first, we interpret the response time behavior, seeing the system as a black box (cf Ludwig (2007)). Afterwards, we switch to the glass box view, taking a deeper look at the system’s internal behavior, analyzing the statistical data, and providing an illustration of where the presented response time behavior originates from.

### 6.1. Measurement

Figure 3 shows the response time results for five load tests running from one to 140 parallel requests. The diagram contains 49350 response time results, resulting in an easily recognizable behavioral pattern. Using figure 4, we will explain this pattern in the following.

Taking a close look at the response time diagram, one can see that the pattern can be split up in three parts. The first part (marked as 1) is where the number of parallel requests is smaller than the number of available work processes on the ERP system. All requests can be handled by the system in parallel. An increasing number of parallel requests slows down the response times for all requests, as can be seen by the difference between 1.1 (one request) and 1.2 (30 parallel requests). Nevertheless, this slowdown is affecting all requests evenly.

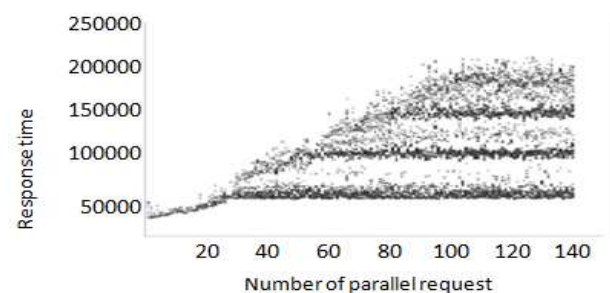


Figure 3: Response time diagram for five load tests

When the number of parallel processes exceeds the number of available work processes, the message queuing used by the ERP system for load balancing becomes visible. The first block of  $n$  requests is processed in parallel (where  $n$  is the number of available work processes) in a constant time (2), independent of the overall number of requests. This is comprehensible, as the surplus of the requests stays in the queue and thus is not consuming any relevant resources.

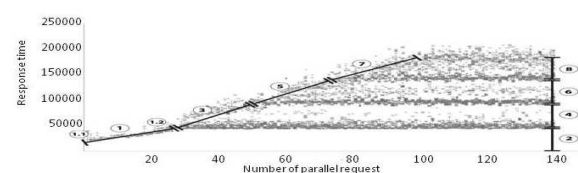


Figure 4: Response time diagram analysis

The surplus of requests is processed after work processes finish the first request. A repeating pattern is recognizable, as the surplus of requests (6) is processed as the requests in section 1. This continues (3 and 7, 4 and 8), up to approximately 100 parallel requests. From this point on, the maximum response time stays constant – due to a timeout of all longer running requests.

As a result we can say that the system’s capacity is at about 100 parallel requests. Passing this limit will result in requests not being successfully answered. But



whatever excessive load generated, at least these 100 requests will be successfully processed.

The response time of a single request though is quite unpredictable.

## 6.2. Glass Box View

To analyze the response time in a glass box view, the relevant components addressed in section 1, and the time slots they use, have to be introduced. The response time in the SAP ERP ABAP stack is quite complex and consists of the following components:

- wait time
- roll in time
- load/generation time
- database time
- enqueue time
- roll out time
- rolled out time
- time in workprocess

As illustrated in figure 1, each request that arrives at the system has to be assigned to a work process in order to get processed. This assignment is done by the dispatcher process. The time for this step is called wait or queue time. The overall time the program is processed by the work process is called time in work process.

As shown in figure 5, this time consists of several components. The first processing step in the work process is to load the process context in its memory. This is called roll in time. If a program calls a remote service, the time, while it has to wait for the response, is also assigned to the roll in time. Then the program has to be loaded into the process memory or generated (compiled) from source code, if it is called the first time. While the request is processed, the work process fetches data from the database. This is aggregated into the metric database time.

The SAP ERP kernel has its own mechanism to control concurrent access to database objects, the so called enqueue process. If a resource is busy, then a work process has to wait until it can be obtained. This slot is called enqueue time. As soon as the work process is finished, the information has to be unloaded from the process memory into the systems shared memory, which is measured by the roll out time. If a request consists of more than 1 work process calls, then the time between the end of the first call and the beginning of the next call is registered as rolled out time.

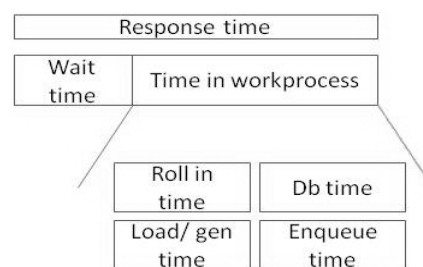


Figure 5: Response time components

These dependencies are shown in figure 3 (response time). In this illustration, the Time in work process consists of the components roll in time, load/generation time, database time, and enqueue time. The roll out time is part of the rolled out time and is not part of the response time, as the response is sent to the client before, in order to reduce the system response time. Different other metrics are not collected, but calculated by the kernel.

The CPU time is a subset of the response time and cannot be assigned to a special component, but is returned by the operating system timer. In the case, of UNIX, the timer works with 100 Hz and consequently the CPU time is always a multiple of 10ms. To analyze the system behavior, the introduced metrics have to be measured. This is done by the SAP ERP kernel completely independent of any ABAP application. The kernel logs a set of performance metrics, like the components of the response time, the program and user name, response size and other important information for performance analysis of the system.

After a request is processed, the work process collects the available information, calculates additional metrics, and stores it in the shared memory. This memory can be accessed by all work processes, and so the performance metrics of the full. As soon as the buffer is full, it is written to binary files on the file system. Every hour a new file named stat is created, while the old one is renamed to stat\_<number>. In this way, the statistical records can be accessed as long as the maximum number of stat files is not reached. As soon as this maximum number is exceeded, the oldest file will be deleted. The number of these stat files (and with it the amount of statistical records that are held) is controlled by parameters of the SAP system. For this case study, their values have been increased to hold all the data of a load run. As the logging of statistical records is a standard functionality of the kernel that is always turned on, this method of monitoring can be referred to as non intrusive as mentioned in JAIN, as it does not add an additional load on the system in comparison to a productive usage of the ERP system.

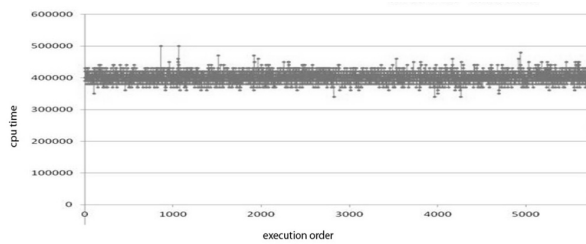


Figure 6: CPU time analysis

The statistical values of each program are displayed using the Transaction STAD in the SAP Gui. To evaluate the big number of these performance values in the presented case study, they have to be exported to a file or a database. For this, several BAPIs exist to deliver the statistical records using RFC function calls. The most complete values are delivered by the BAPI SAPWL\_STATREC\_DIRECT\_READ. As this BAPI is not exported as a web service by the SAP system, it has been copied and modified to meet the requirements for the export as a web service. Thereby, the feature of obtaining the statistical values as soon as the load run is over has been integrated in the presented load generator and analysis tool. To have these performance values easily available, a database has been created to store these values of each load run.

Taking a closer look on the measured values, as discussed in section 5, a clear pattern can be observed. Using the glass box approach, this pattern will be explained by analyzing the process times of each component of the SAP ERP systems kernel. As the system has been warmed up, before the load test has been executed, load/generation time can be neglected, as the ABAP code is already in the corresponding buffer. The same effect can be observed in productive systems, as the size of the buffers is adapted to the expected workload. In addition, the amount of data that has to be rolled in at the beginning of the processing phase is too small to have influence on the systems performance.

The determining factor for the overall response time are the CPU time, total database time, commit time, which is a part of the total db time, and the queuing time. The measured values of these metrics will be introduced in the following to illustrate the performance behavior of the web service used in this case study.

Figure 6 shows the CPU time in nanoseconds according to the timestamp the request has been processed. As the load run has been executed with a growing number of parallel requests, this graph shows that the CPU time, as expected, does not depend on the number of parallel requests. The overall response time is not negatively affected by the CPU time under a growing number of parallel requests.

Figure 7 illustrates the database commit time in nanoseconds; also according to the timestamp the request has been processed. This graph shows a few irregularities, which are created by the log writer mechanism of the underlying database. Besides these

irregular values, it does not show any dependencies to the number of parallel requests.

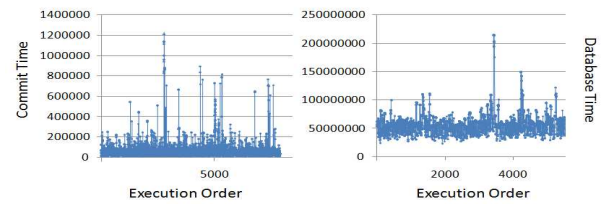


Figure 7: Commit time and database time

The total database time is shown in the right part of figure 7. There is no indication that the performance of the database component does significantly change with the number of parallel requests. As the ERP system has a certain number of work processes, it has been assumed that the database response time does not change significantly with a rising number of parallel requests to the SAP ERP system, as there are not more requests arriving at the database than work processes in the system exist. But even in the starting phase, where less parallel requests had to be served by the system and the database, the performance of the database is in a certain frame.

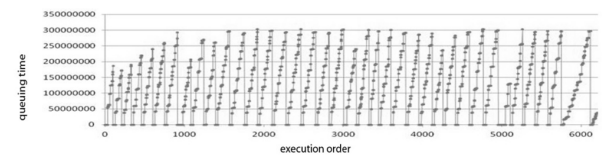


Figure 8: Queuing time

An explanation of the measured values shown in figure 4 is found in figure 8 showing the queuing time in nanoseconds; according to the timestamp the request has been processed.

The graph shows a significant dependency to the number of parallel requests. As in the starting phase the queuing time is small, it grows with more requests sent to the system. With 30 work processes configured, the first 30 requests had no or a negligible queuing time, while the remaining requests had to wait until one of the first requests has been finished. Consequently, the queuing time rises with the number of parallel requests until the sequence is finished.

## 7. LIMITATIONS

This case study does not regard the usage of different technical users on the SAP system. However, in a well configured productive SAP system, the shared memory that is intended to hold the user contexts is big enough to hold all data in shared memory without the need to put some data into a slower memory area. Therefore, the simplification of reducing the amount of different users does not reduce the significance and validity of the case study's results.

In addition, the impact of a suboptimal system configuration to system response times is not part of this work. Using a non optimal resource distribution,

several system components may respond differently compared to the response times and components' behavior in general measured in this case study. For instance, in some special cases the work process is assigned to a user as long as he is logged into the system. This work aims at understanding the general concept of the response time of an ERP system that is well parameterized, and not at the analysis of performance problems.

Equally, the different behaviors of ERP systems that are used for development purposes, so called "DEV" Systems, are not regarded in this case study.

A challenging task when analyzing complex software systems is the decision how many components should be integrated in the analysis. Even in this defined example, the paper demonstrates that a lot of data is gathered from several components in the SAP system and that even the database can be described in a more detailed way. In (Gradl et al. 2010) an eight level architecture was presented to limit the effort of building the architecture of the SAP system. By analyzing the SAP system and its traces, it was discovered that the lowest level is the response time level of the database. As the SAP system does not provide more detailed information about the database, the database has been treated as black box. Response time values of the database are derived directly from the SAP system.

## 8. CONCLUSION

On the first view, the response time behavior of the analyzed service seems to be complex. But on a second view it reveals a quite predictable behavior. Using a queuing mechanism, the SAP system processes several requests in parallel, and at the same time it regulates the resource consumption by limiting the number of parallel processed requests to the number of work processes available. This leads to a stable environment not being affected by the amount of parallel requests, and a guaranteed response time for some of the requests.

Taking a look inside the system, it can be seen that the resource consumption behind the messaging layer is quite constant. Considering the queuing, this is not surprising. Only the commit time shows some volatility, caused by the fluctuation of parallel processes accessing the database. This volatility of the commit time also causes the jitter in the response time diagram.

In summary we draw the conclusion, that the dispatching time represents the difference between response times, while the processing time is constant independent of how many parallel requests are processed. While this sounds obvious at first, it is an important validation for performance analysis and prediction.

As next steps, we use the information acquired in this work and performance data gathered to parameterize a layered queuing model based on (Gradl et al. 2010) to predict the performance of a SAP ERP system via simulation (Woodside 2002).

## REFERENCES

- Apache Software Foundation "Apache Axis2/Java," 2010.
- Bailey, D.H., and Snaveley, A. "Performance Modeling: Understanding the Present and Predicting the Future," in: *Euro-Par 2005 Parallel Processing*, Springer, Berlin / Heidelberg, 2005, pp. 185-195.
- Bögelsack, A., Jehle, H., Wittges, H., Schmidl, J., and Krcmar, H. "An Approach to Simulate Enterprise Resource Planning Systems," in: *6th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems, MSVVEIS-2008, In conjunction with ICEIS 2008*, U. Ultes-Nitsche, D. Moldt and J.C. Augusto (eds.), INSTICC PRESS, Barcelona, Spain, 2008, pp. 160-169.
- Bögelsack, A., Krcmar, H., and Wittges, H. "Performance Overhead of Paravirtualization on an Exemplary ERP System," in: *12th International Conference on Enterprise Information Systems*, Funchal, Madeira, Portugal, 2010.
- Chen, S.G., and Lin, Y.K. "Performance analysis for Enterprise Resource Planning systems," *Industrial Engineering and Engineering Management*, 2008. IEEM 2008. IEEE International Conference on, 2008, pp. 63-67.
- Gradl, S., Mayer, M., Wittges, H., and Krcmar, H. "Modeling ERP Business Processes Using Layered Queueing Networks," in: *12th International Conference on Enterprise Information Systems*, Funchal, Portugal, 2010.
- Hollingsworth, J.K., Snaveley, A., Sbaraglia, S., and Ekanadham, K. "EMPS: An Environment for Memory Performance Studies," in: *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05) - Workshop 10 - Volume 11*, IEEE Computer Society, 2005, p. 223.222.
- Jain, R. *The art of computer systems performance analysis* Wiley, New York, 1991.
- Jehle, H. "Performance-Messung eines Portalsystems in virtualisierter Umgebung am Fallbeispiel SAP," in: *CVLBA Workshop 2009. 3. Workshop des Centers for Very Large Business Applications (CVLBA)*, Arndt, H.-K.; Krcmar, H., Magdeburg, Deutschland, 2009.
- Jin, Y., Tang, A., Han, J., and Liu, Y. "Performance Evaluation and Prediction for Legacy Information Systems," in: *ICSE '07*, IEEE, Minneapolis, 2007.
- Krcmar, H. *Informationsmanagement* Springer-Verlag New York, Inc., 2009.
- Kruse, H.G. *Leistungsbewertung bei Computersystemen*, (1. ed.) Springer-Verlag, Berlin, Heidelberg, 2009.

- Lilja, D.J. *Measuring Computer Performance - A practitioner's guide* Cambridge University Press, 2000.
- Ludewig, J., and Lichter, H. *Software Engineering - Grundlagen, Menschen, Prozesse, Techniken*, (1. ed.) dpunkt.verlag GmbH, Heidelberg, 2007, p. 618.
- Malik, H. "A methodology to support load test analysis," in: *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 2*, ACM, Cape Town, South Africa, 2010, pp. 421-424.
- Menascé, D.A. "Software, performance, or engineering?," in: *Proceedings of the 3rd international workshop on Software and performance*, ACM, Rome, Italy, 2002, pp. 239-242.
- Menasce, D.A., and Almeida, V.A.F. *Capacity Planning for Web Services: metrics, models, and methods* Prentice Hall, Upper Saddle River, NJ, Upper Saddle River, NJ, 2002.
- Nudd, G.R., Kerbyson, D.J., Papaefstathiou, E., Perry, S.C., Harper, J.S., and Wilcox, D.V. "Pace - A Toolset for the Performance Prediction of Parallel and Distributed Systems," *International Journal of High Performance Computing Applications* (14:3), August 1, 2000 2000, pp 228-251.
- Prior, D. "Who Sets the Pace in the SAP Performance 'Olympics'?", *Gartner*, 24.2.2003 2003, p 6.
- Rolia, J., Casale, G., Krishnamurthy, D., Dawson, S., and Kraft, S. "Predictive modelling of SAP ERP applications: challenges and solutions," in: *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Pisa, Italy, 2009, pp. 1-9.
- Schneider-Neureither, A. *Optimierung von SAP-Systemlandschaften*, (1. ed.) Galileo Press, Bonn, 2004.
- Schneider, T. *SAP-Performanceoptimierung*, (5 ed.) Galileo Press, Bonn, 2008, pp. 247-249.
- Weidner, S. "Integrations-Fallstudie PP (SAP ECC 5.0)," 2006.
- Woodside, M. "Tutorial Introduction to Layered Modeling of Software Performance," 2002.

# OBJECT-ORIENTED MODEL OF FIXED-BED DRYING OF COFFEE BERRIES

Emílio de Souza Santos<sup>(a)</sup>, Paulo Cesar Corrêa<sup>(b)</sup>, Brian Lynn Steward<sup>(c)</sup>, Daniel Marçal de Queiroz<sup>(d)</sup>

<sup>(a)</sup> Regional Development Analyst, Agricultural and Environmental Engineer, CODEVASF, Brasília/DF – Brazil.

<sup>(b), (d)</sup> Associate Professor, Department of Agricultural Engineering, Viçosa Federal University, Viçosa/MG – Brazil.

<sup>(c)</sup> Associate Professor, Dep. of Agricultural and Biosystem Engineering, Iowa State University, Ames/IA – USA.

<sup>(a)</sup> [emilio\\_ss@hotmail.com](mailto:emilio_ss@hotmail.com), <sup>(b)</sup> [copace@ufv.br](mailto:copace@ufv.br), <sup>(c)</sup> [bsteward@iastate.edu](mailto:bsteward@iastate.edu), <sup>(d)</sup> [queiroz@ufv.br](mailto:queiroz@ufv.br)

## ABSTRACT

The aims of this work were to validate a new model of a fixed-bed drying process with experimental data, compare this model with a traditional one (MSU Model), and analyze the sensitivity of each model parameter. A low level model of a thin-layer was developed, based on mass and heat balance between the drying air and the product. It also considered the product physical properties variation during drying. Thus a high level model was created by connecting four thin-layer models, in order to represent a thick layer. The proposed model was validated by the analysis of model performance index between the experimental data and simulated results. Finally, the efficiency indexes of the proposed model and the MSU model were compared. The models and the simulation were done using OpenModelica® 1.6.0, based on the Modelica language. The proposed model had shown good performance indexes compared with the MSU model.

Keywords: Modelica, physical properties, MSU Model, model performance.

## 1. INTRODUCTION

After harvest, coffee is very perishable due to the high moisture and sugar content. Thus, coffee passes through a drying operation, to enable safe storage and minimize quality degradation prior to subsequent processing. Drying is defined as an operation of moisture removal through to simultaneous heat and mass transfer (Henderson et al. 1997). It is also one of the preservation methods of agricultural products which are most often used and is the most energy-intensive process in industry (Dincer 1998). Moreover, drying is a difficult food processing operation mainly because of undesirable changes in the quality of the dried product. Such changes can occur because of substantial changes in the product's chemical and physical properties during the drying process.

In a drying system, two fundamental tasks must be performed. First, energy must be provided to the product to be dried so that moisture contained in the product is vaporized. Second, the vaporized moisture must be removed by the drying system (Henderson et

al. 1997). Thus, air used to dry the product, after passing through a thin-layer of product, experiences a decrease in temperature and an increase in its humidity ratio. Consequently, thick-layer drying is a complex dynamic system, considering that the drying kinetics will change due to the changes in air and product properties through time and space.

In order to reduce costs and robustly control drying systems, simulation of the process is important, as a mean to understand the system dynamics and then optimize the system. Many models have been developed to simulate fixed-bed drying. A commonly used model is the MSU model (Michigan State University). This model consists of a system of four differential equations resulting from mass and energy balances of a control volume (Dalpasquale et al. 2008). The MSU model is easy to apply to different agricultural products because it assumes that some product properties is equal to water properties, e.g. specific heat. However, this model tends to overestimate the drying rate (Brooker et al. 1992), due to the model assumptions.

Currently, with the rapid improvement of computer and software technology and the widespread availability of computational resources, high fidelity models can be implemented with simulation that more closely represent reality. Some recent examples of new drying models include the work of Izadifar and Mowla (2003), who considered the theory of simultaneous mass and heat convection and internal mass diffusion in their model; Guiné et al. (2007) who based their model on the liquid diffusion theory, considering the product shrinkage and physical properties variation during drying; and Lecorvaisier et al. (2010), who considered the dynamic phenomena of air turbulence.

In addition, the dynamics of the drying system are coupled with those of various components that are necessary for the operation of the entire system including motors, fans, conveyors, heaters and control devices. Each component performs a specific operation that effect the entire system and the system affects the operation of each component. The simulation of this type of system is usually performed by simulating each component individually. This approach is taken due to the difficulty of developing an overall system model that represents many physical domains including the

electrical, mechanical, thermodynamic, hydraulic, chemical, and control domains.

Recently though, modeling technologies have been advanced to ease the development of complex, multi-domain, physical models. For example, *Modelica* is an object-oriented equation-based programming language which enables multi-domain modeling, meaning that model components corresponding to physical objects from several different domains that can be described and connected (Fritzson, 2003). Therefore, the *Modelica* language is promising for drying system modeling, since this language can enable modeling of all the physical domains and mechanisms in this type of system.

The aims of this work were to: (1) model the drying process of a fixed-bed coffee dryer considering the variation of the product physical properties and drying air flow; (2) validate the model by comparing the simulate results with experimental data; (3) compare this model with the MSU Model; and (4) analyze parameter sensitivity in order to understand and simplify the proposed model.

## 2. METHODOLOGY

The present work was conducted in the Laboratory of Evaluation of Physical Properties and Quality of Agricultural Products in the Brazilian Grain Storage Training Center – CENTREINAR, Federal University of Viçosa, Viçosa, MG, Brazil.

### 2.1. Modeling

The proposed model to describe the fixed-bed coffee dryer was built *Modelica* language. Real physical objects are represented on *Modelica* language as *object* block. Each object is an instance of a specific *class*, also called *model*. Furthermore, *models* contain both the state of the object, represented by *variables*, and the behavior of the object, represented by *equations*.

Object-oriented language enables a process known as *inheritance*. *Base-class* states (variables) and behaviors (equations) can be extended to a *sub-class*. Thus, the model hierarchy is organized by levels, where models are classified as a higher level than the inherited model. Figure 1 shows the schema of the proposed fixed-bed-dryer model and the model hierarchy.

The coffee fixed-bed was considered to be a thick layer composed of a finite number of thin layers. Each thin layer is a control volume that has a fixed transversal area  $A$  and a variable thickness  $L$ . The drying kinetics in each control volume were described using thin layer drying theory and were modeled as a *Modelica* model class.

*Modelica* model classes are connected with connector class which represents the conservation relations between model classes. In this case, the connector variables were air properties. The connector's flow variable is the velocity ( $V$ ) which is proportional to mass flow rate. The connector's potential (non-flow) variables are the temperature ( $T$ ), pressure ( $p$ ) and relative humidity ( $rh$ ) of the air. The sub-index  $a$

denotes an input variable and the sub-index  $b$  denotes an output variable. Figure 2 shows a scheme of the thin layer model (*DryLayer*) with the inlet and outlet connector.

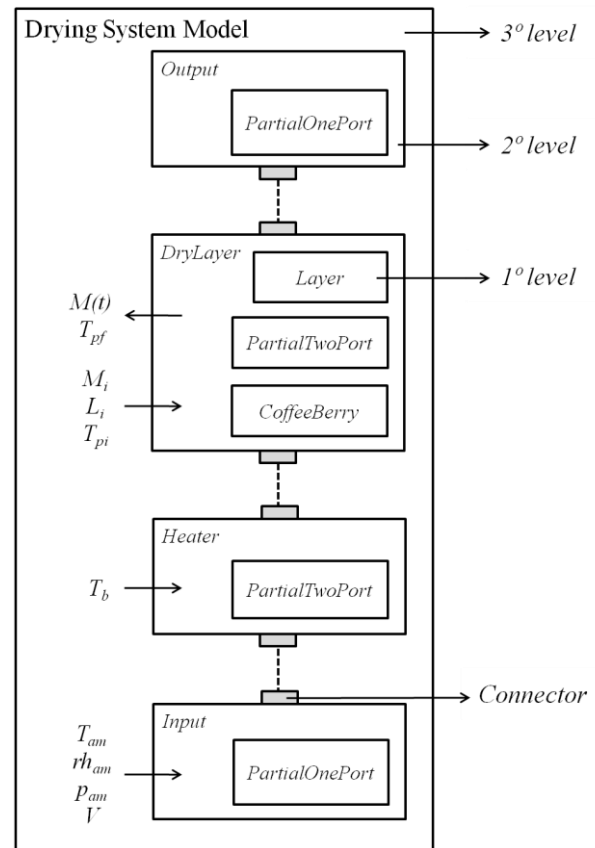


Figure 1: Hierarchy schema of the thick layer drying

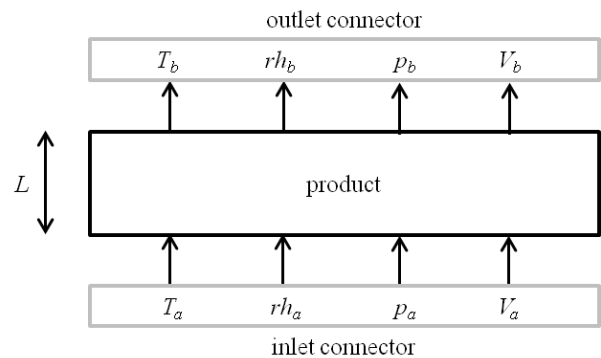


Figure 2: Control volume and air connectors of the *DryLayer* model.

The first-level-models were created containing the air psychrometrics equations and properties. The second-level-model to represent the air inlet and outlet are named *Input* and *Output*, respectively. These models need only a single connector, inlet or outlet. Thus, a first-level-model, named *PartialOnePort*, was created containing the psychrometrics equations and a connector. The others second-level-models need a pair of connectors (inlet-outlet). Thus, a first level model was created, named *PartialTwoPort*.



The saturation vapor pressure was described as function of temperature by an empirical equation proposed by Keenan and Keyes (Henderson et al. 1997).

$$\ln\left(\frac{p_{vsat}}{R'}\right) = \frac{A + BT + CT^2 + DT^3 + ET^4}{FT - GT^2} \quad (1)$$

where:  $p_{vsat}$  is the air saturation vapor pressure, Pa;  $T$  is the air temperature, K;  $R'$ ,  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $F$  and  $G$  are the adjusted empirical coefficients.

The air vapor pressure, humidity ratio, enthalpy and density were calculated using basic psychometrics equations. These equations were formulated based on Dalton's law, mass and energy conservation principles.

$$H = \frac{p_v}{1.605(p - p_v)} \quad (2)$$

$$\rho_{ar} = \frac{(1 + H)(p - p_v)MM_{H_2O}}{RT} \quad (3)$$

$$p_v = rh p_{vsat} \quad (4)$$

$$h = c_{ar}(T - T_0) + H[h_{g,0} + c_v(T - T_0)] \quad (5)$$

where:  $H$  is the air humidity ratio, kg of vapor per kg of dry air;  $p$  is the air absolute pressure, Pa;  $p_v$  is the air vapor pressure, Pa;  $\rho_{ar}$  is the humid air density, kg m<sup>-3</sup>;  $MM_{H_2O}$  is the water molar mass, 18.02 kg kmol<sup>-1</sup>;  $R$  is the universal gas constant, 8314 J mol<sup>-1</sup> K<sup>-1</sup>;  $rh$  is the air relative humidity, dimensionless;  $h$  is the humid air enthalpy, kJ kg<sup>-1</sup>;  $h_{g,0}$  is the water latent heat at the reference temperature, 2502.5352 kJ kg<sup>-1</sup>;  $T_0$  is the reference temperature, 273.15 K;  $c_{ar}$  is the dry air specific heat, 1.0069 kJ kg<sup>-1</sup> K<sup>-1</sup>; and  $c_v$  is the vapor specific heat, 1.8757 kJ kg<sup>-1</sup> K<sup>-1</sup>.

The air and vapor mass flows are calculated in the first level models. These variables are used in the mass and energy balance equations.

$$m'' = \rho_{ar} V \quad (6)$$

$$m_v'' = \left(\frac{H}{H + 1}\right) m'' \quad (7)$$

where:  $m''$  is the air mass flow, kg m<sup>-2</sup> s<sup>-1</sup>;  $m_v''$  is the vapor mass flow, kg m<sup>-2</sup> s<sup>-1</sup>;  $V$  is the air velocity, m s<sup>-1</sup>.

The first-level-model named *CoffeeBerry* was created containing the mathematical equation that describes the product physical properties as functions of its moisture content. This model is stored in a Package named *Product*, which can be used to storage many types of product models.

The coffee volume, density, porosity and specific heat were described using the equations 8 to 11, which were developed by Junior (2001). Assuming the coffee berry to be a perfect sphere, the product average diameter can be calculated using equation 12.

$$v = 10^{-9} (621.46 + 152.78 M + 12.417 M^2) \quad (8)$$

$$\rho_p = 420.8490 + 198.8201 M - 53.8475 M^2 \quad (9)$$

$$\varepsilon = 10^{-3} (432.324 + 114.307 M - 32.317 M^2) \quad (10)$$

$$c_p = 0.9447 + 3.6197 M - 1.9920 M^2 \quad (11)$$

$$v = \frac{\pi d_p^3}{6} \quad (12)$$

where:  $M$  is the product moisture content, kg of water per kg dry mass (d.b.);  $v$  is the product volume, m<sup>3</sup>;  $\rho_p$  is the product density, kg m<sup>-3</sup>;  $\varepsilon$  is the product porosity, dimensionless;  $d_p$  is the product average diameter, m; and  $c_p$  is the product specific heat, kJ kg<sup>-1</sup> K<sup>-1</sup>.

The drying constant was calculated using the equation 13, proposed by Young and Dickens (Henderson et al. 1997). This equation relates a general drying constant to a referential one, obtained experimentally. The empirical equation of Junior (2001) was used as the referential drying constant. The saturation vapor pressure ratio in equation 13 was assumed to be equal to 1 due to the equation 14 already considering the variation of relative humidity. The referential velocity is equal to 0.2166 m s<sup>-1</sup>, the value used in Junior's experiment.

$$k = \frac{1}{3600} k_r \left(\frac{p_{sat}}{p_{satr}}\right)^{0.46} \left(\frac{V}{V_r}\right)^{0.7} \quad (13)$$

$$k_r = -0.1196 + 1.4180 \cdot 10^{-3} T_c + 6.9938 \cdot 10^{-5} T_c^2 + 0.6545 rh - 0.5369 rh^2 - 7.5170 \cdot 10^{-3} T_c rh \quad (14)$$

where:  $k$  is the drying constant, s<sup>-1</sup>;  $k_r$  is the referential drying constant, h<sup>-1</sup>;  $p_{satr}$  is the referential saturation vapor pressure, Pa;  $V_r$  is the referential air velocity, m s<sup>-1</sup>;  $T_c$  air temperature in Celsius degree, °C.

The product water latent heat was described using equation 15, adjusted by Junior (2001). The Modified Henderson equation (16), adjusted by Correa et al. (2010), was used to describe the hygroscopic equilibrium between the product and the air.

$$h_{fg} = (2502.49 - 2.43 T_c) (1 + 7.7866 \cdot 10^6 \exp(-19.6621 M^{0.0499})) \quad (15)$$

$$1 - rh_e = \exp(-0.0001(46.8549 + T_c) M_e^{1.8299}) \quad (16)$$

where:  $h_{fg}$  is the water latent heat inside the product,  $\text{kJ kg}^{-1}$ ;  $M_e$  is the product equilibrium moisture content, d.b.; and  $rh_e$  is the equilibrium relative humidity, dimensionless.

A first-level-model named *Layer* was created to define initial values of the variables: the initial layer thickness ( $L_i$ ), initial moisture content ( $M_i$ ) and initial product temperature ( $T_{pi}$ ). The initial layer thickness of the control volume has to be as small as possible, in order for the thin layer drying theory to be applicable.

The *Input* model extended the *PartialOnePort* model and has the structure to receive the boundary variables of the air. The *Output* model only needs to extend the *PartialOnePort* model.

The *Heater* model was created to represent the air heater. This model extended the *PartialTwoPort* model and has the structure to receive the variable  $T_b$ , which represent the temperature that the air is heated (drying temperature). As result, all outlet air variables are calculated by the *Heater* model.

The *DryLayer* model was created to represent the control volume of product. This model extended the *PartialTwoPort*, *CoffeeBerry* and *Layer* models and has the mass and energy balance of the drying process.

The exponential equation, proposed by Sherwood (Henderson et al. 1997), was used to describe the thin layer drying of the product. This equation assumes that the drying rate is proportional to the difference between the moisture content and the equilibrium moisture content, proportionality given by the product drying constant.

$$\frac{dM}{dt} = -k(M - M_e) \quad (17)$$

where:  $dM/dt$  is the drying rate,  $\text{s}^{-1}$ .

The heat balance equation considering on *DryLayer* model neglects the heat transfer due to conduction between the particles and radiation between the particles and the dryer's walls. Moreover, it was assumed that all heat transferred by the air is used to the product water evaporation and to heat the product. Thus the changes in air properties can be related to the variation on product moisture content, using the mass (18) and heat (19) balance equations. The convective heat transfer coefficient was calculated using the Barker's empiric equation (21), presented by Brooker et al. (1992).

$$m''_{vb} - m''_{va} = -\frac{dM}{dt} \frac{\rho_p L}{1 + M} \quad (18)$$

$$m''_b h_b - m''_a h_a = \frac{dM}{dt} \frac{\rho_p L}{1 + M} h_{fg} - \rho_p L c_p \frac{dT_p}{dt} \quad (19)$$

$$m''_b h_b - m''_a h_a = -h'(T_a - T_p) \quad (20)$$

$$h' = 0,9918 c_{ar} m''_a \left( \frac{3600 d_p m''_a}{0,06175 + 16,5 \cdot 10^{-5} T_c} \right)^{-0,34} \quad (21)$$

where:  $dT_p/dt$  is the product temperature rate,  $\text{K s}^{-1}$ ;  $T_p$  is the product temperature,  $\text{K}$ ; and  $h'$  is the heat transfer coefficient  $\text{W m}^{-2} \text{K}^{-1}$ .

The pressure variation through the product layer was calculated using the equation 22, which assumes that the total pressure variation is equal to the vapor pressure variation plus the friction loss. The friction loss was calculated using the Darcy equation (23) and the friction factor was calculated using the Ergun equation (24) (Henderson et al. 1997).

$$p_b - p_a = (p_{vb} - p_{va}) - p_f \quad (22)$$

$$p_f = f \rho_{ar,a} \frac{L}{d_p} \frac{V_a^2}{2} \quad (23)$$

$$f = \left( \frac{1 - \varepsilon}{\varepsilon^3} \right) \left( 3,5 + \frac{300(1 - \varepsilon)}{Re} \right) \quad (24)$$

$$Re = \frac{\rho_{ar,a} V_a d_p}{\mu_{ar,a}} \quad (25)$$

where:  $p_f$  is the friction loss,  $\text{Pa}$ ;  $f$  is the friction factor, dimensionless;  $Re$  is the modified Reynolds number, dimensionless;  $\mu_{ar}$  is the air dynamic viscosity,  $\text{Pa s}$ .

The layer thickness variation was calculated using the equation 26. This equation considers the dry mass conservation and that the volume shrinkage only happens on the thickness direction, due to the fixed transversal area of the dryer. The sub-index  $i$  denotes initial.

$$\frac{\rho_v L}{1 + M} = \frac{\rho_{v,i} L_i}{1 + M_i} \quad (26)$$

Many *DryLayer* model can be connect to form a third-level-model that represents a thick fixed layer. The model was implemented in *Modelica* language and simulated using the *OpenModelica*® 1.6.0 package.

## 2.2. Experimental Data

The experiment was conducted in a prototype dryer consisted by three chambers with equal dimension of  $0.57 \times 0.35 \times 0.64 \text{ m}$ . It was used coffee berries (*Coffea arabica* L.) variety *Mundo Novo*, manually harvested and pre-dried with ambient air.

The coffee was dried with dryer air (initial inlet air) with temperature at three levels: 40, 50, 60 °C. Three replicates were done for each temperature, resulting in nine experimental tests. Each test had a specific condition of ambient air and different initial moisture content of the product, presented on Table 1.

The coffee samples were withdrawn at predetermined periods at the heights of 0.10, 0.25, 0.40 and 0.55 m, in order to determine the product moisture content in different layers. For the simulation, it was assumed that each sample was withdrawn at the center of the layer, so the first and third layers have 0.20 m of thickness and the second and forth layers have 0.10 m of thickness.

Table 1: Drying and ambient conditions for each experimental tests.

<i>Test</i>	$T_s$ °C	$T_{am}$ °C	$rh_{am}$ %	$p_{am}$ kPa	$V$ m/min	$M_i$ d.b. %
40.1	40	23.5	55.9	93.843	8.60	17.91
40.2	40	25.4	52.7	93.683	8.81	20.41
40.3	40	20.2	67.7	94.216	8.27	20.25
50.1	50	22.9	48.0	94.376	9.20	38.56
50.2	50	23.2	62.2	93.750	8.82	18.56
50.3	50	19.8	69.6	94.136	8.72	18.10
60.1	60	22.9	57.7	93.790	7.10	17.45
60.3	60	16.7	45.7	94.083	8.27	32.32
60.3	60	24.3	43.9	93.817	8.81	19.70

where:  $T_s$  is the drying air temperature;  $T_{am}$  is the ambient temperature;  $rh_{am}$  is the ambient air relative humidity;  $p_{am}$  is the ambient air absolute pressure;  $V$  is the inlet air velocity;  $M_i$  is the product average initial moisture content.

### 2.3. Model Validation

The model statistical performance was evaluated by analysis of the relative standard deviation (*RSD*) and the performance index (27). This index evaluates the model performance by multiplying one dimensionless number that corresponds to the model precision, given by correlation coefficient ( $r$ ), and another that corresponds to the model accuracy, given by agreement index (28) (Willmoot et al. 1985).

$$i = r d \quad (27)$$

$$d = 1 - \frac{\sum_{j=1}^n (Y_j - \hat{Y}_j)^2}{\sum_{j=1}^n (|Y_j - \bar{Y}| + |\hat{Y}_j - \bar{Y}|)^2} \quad (28)$$

where:  $i$  is the performance index;  $r$  is the correlation coefficient;  $d$  is the agreement index;  $Y$  is the observed data;  $\hat{Y}$  is the model-predicted data;  $\bar{Y}$  is the average value of observed data;  $n$  is the number of observed data.

### 2.4. Model comparison

The proposed model performance was compared with the MSU model, in order to investigate the validity of the new model. The performance comparison between the models was done by analyzing the performance index of each model.

The MSU model is completely described by Brooker et al. (1992), and was implemented in *Modelica* language and simulated with the same conditions of the proposed model using the *OpenModelica*® 1.6.0 package.

If compared with the proposed model, the MSU model have the following particularities: constant physical properties; assumption of water latent heat being equal to the free water; neglecting of pressure variation due to the increasing of vapor pressure through the drying; and neglecting of the layer volume shrinkage.

### 2.5. Sensitivity Analysis

Model sensitivity to parameter variation was evaluated using the differential sensitivity analysis method. In this method, all model parameters were classified by means of the sensitivity coefficient (29). The sensitivity coefficient represents the ratio between the output variable variation (response) and the analyzed parameter variation (perturbation), while all other parameters remain constant (Hamby 1994). The model result while all parameters are held constant is defined as the “base case”. The sensitivity coefficient was calculated to response of  $M$ ,  $T_p$ ,  $rh$  and  $T$  of all layers, and a global value  $S_i$  was calculated as an average value of all calculated sensitivity coefficients. The analysis was done considering a variation of  $\pm 20\%$  of each model's parameter.

$$S_{X,\beta} = \frac{|X - X_i| \beta}{|\beta - \beta_i| X} \quad (29)$$

where:  $S_{X,\beta}$  is the sensitivity coefficient response of  $X$  to the  $\beta$  variation;  $X$  is the output variable value to the base case;  $X_i$  is the output variable value to the  $\beta$  variation;  $\beta$  is the value of the analyzed parameter to the base case;  $\beta_i$  is the varied value of the analyzed parameter.

## 3. RESULTS

### 3.1. Simulation Results

Figure 3 shows the experimental observed values and predicted simulate results of moisture content for test 40.3, 50.3 and 60.3.

It can be observed in Figure 3 that the drying capacity increases when the temperature increases. This

behavior occurs due to the increase of the product internal water diffusion and the increase of vapor pressure gradient between the water layer surface and the drying air.

Vossen (1979) indicated that a moisture content lower than 14 % (d.b.) is essential to ensuring safe storage of coffee. Furthermore, a low moisture content of coffee berries is important in the husking process. For the experimental conditions of the 40.3° C test, the 12 hour drying time was not enough to ensure a moisture content of 14 % (d.b.) in all layers. It is recommended for the industrial coffee drying process that the process end when the last product layer reaches the 14 % (d.b.) moisture content.

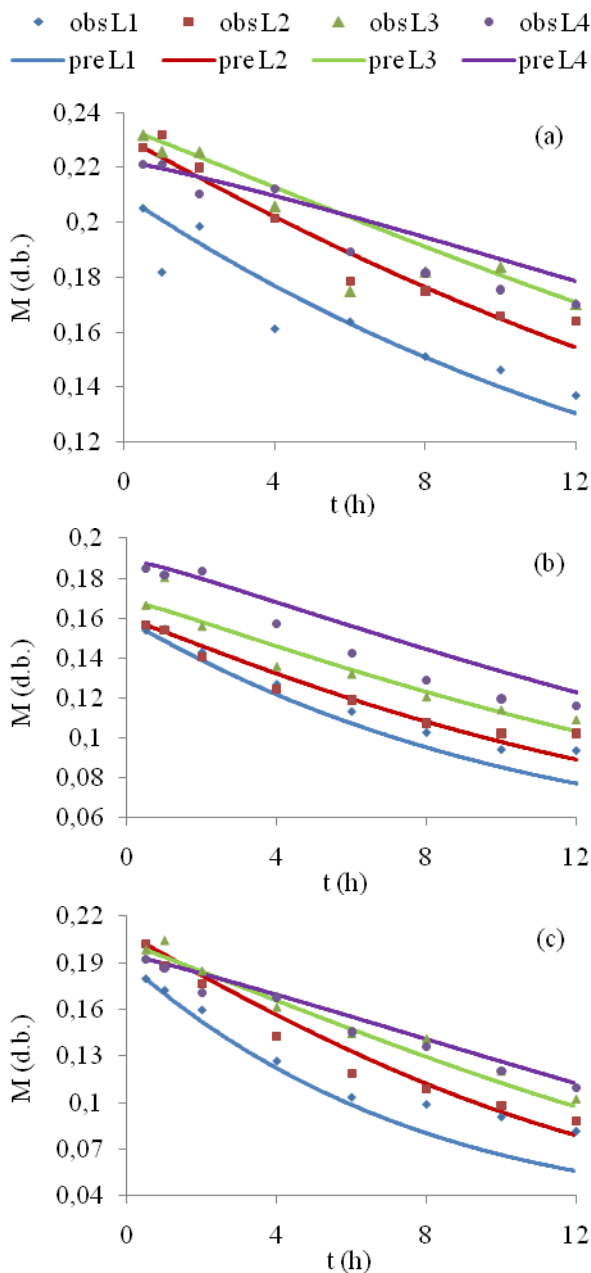


Figure 3: Observed (obs) and predicted (pre) values of moisture content of test 40.3 (a), 50.3 (b) and 60.3 (c).

Another important parameter to evaluate for process quality is the product temperature ( $T_p$ ). Clarke and Macrae (1987) stated that if the coffee berry reaches a temperature of 40°C, the product suffers physical and chemical changes that will decrease the quality of the coffee beverage. Thus, it is recommended that the maximum drying temperature for static dryers be 40°C (Sfredo et al. 2005). Figure 4 shows the simulate results of product temperature for the test 40.3, 50.3 and 60.3.

It can be observed in Figure 4 that  $T_p$  reached a value higher than 40 °C for layer 1, 2 and 4 in the 60.3° C test. Based only on the simulation results, the use of a temperature higher than 60 °C for drying will generate a low quality product. In the 50.3° C test, only  $T_p$  of layer 2 reaches a value higher than 40 °C. Probably the use a temperature of 50 °C can be used to drying coffee depending on the drying conditions and system control.

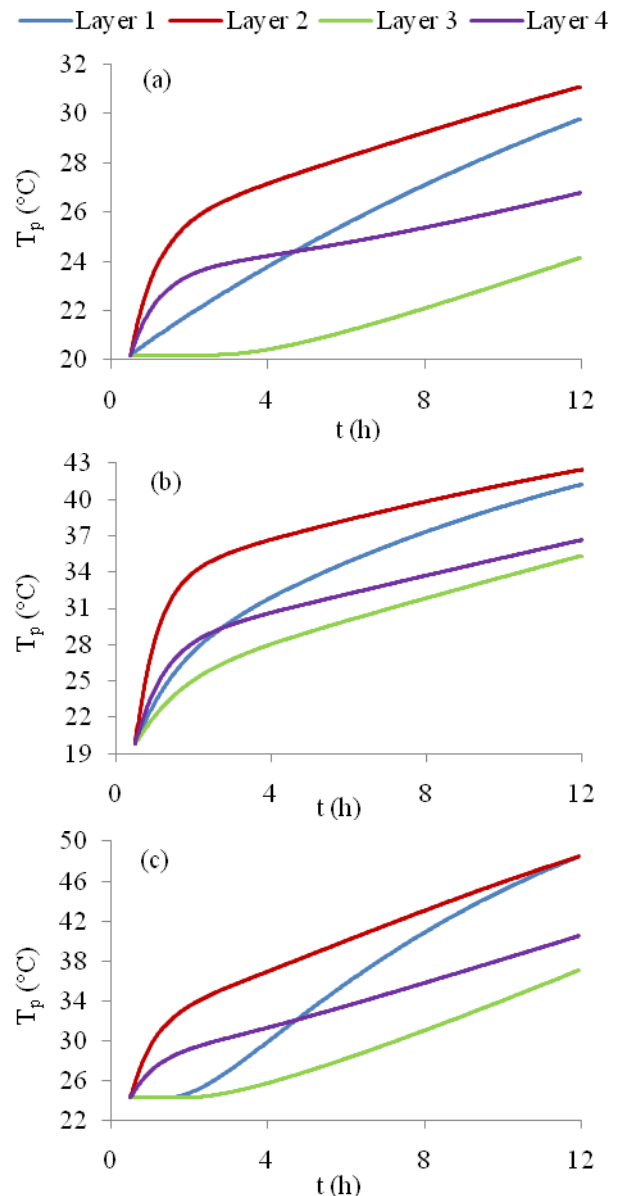


Figure 4: Predicted values of  $T_p$  for each analyzed layer of test 40.3 (a), 50.3 (b) and 60.3 (c).

The temperature of small particles is hard to measure incisively due to the size of the measurement equipment. Usually the product surface temperature or the transfer fluid temperature is used as a quality process parameter. In this contest the simulation results of product temperature can be used as a process control.

### 3.2. Validation

Table 2 shows the agreement index, correlation coefficient, performance index and the relative standard deviation of all simulated tests.

Table 2: Statistical validation parameters of each simulated test.

Test	<i>d</i>	<i>r</i>	<i>i</i>	<i>RSD</i> (%)
40.1	94.55	92.07	87.05	3.41
40.2	93.62	94.75	88.71	4.53
40.3	96.48	95.12	91.77	3.36
50.1	97.01	95.35	92.49	5.79
50.2	97.44	95.93	93.48	4.16
50.3	97.38	96.33	93.81	4.18
60.1	97.62	96.63	94.33	5.24
60.2	97.53	97.92	95.51	6.66
60.3	98.37	97.77	96.18	5.21

All simulated tests of the proposed model presented values of relative standard deviation lower than 10 %, indicating satisfactory representation of the studied phenomena (Chen and Morey, 1989; Madamba et al., 1996; Mohapatra and Rao, 2005).

Moreover, all simulated tests presented values of performance index were higher than 85 %, indicating accuracy and precision of the proposed model to describe the system (Camargo and Sentelhas 1997).

Based on the analyzed statistical parameters the proposed model can safely describe the fixed-bed system to drying coffee berries, for the range of analyzed conditions.

### 3.3. Model comparison

Figure 5 shows the experimental observed values and MSU predicted results of moisture content for tests 40.3, 50.3 and 60.3.

The results showed in Figure 5 corroborate with the affirmation of Brooker et al. (1992) that the MSU model tends to overestimate the drying results, due to the considerations inherent to the model.

Table 3 shows the agreement index, correlation coefficient, performance index and the relative standard deviation of all simulated tests using MSU model.

Almost all condition simulated using the MSU model presented values of relative standard deviation higher than 10 % and/or a performance index lower than 85 %, indicating that this model not satisfactory describes the studied phenomena.

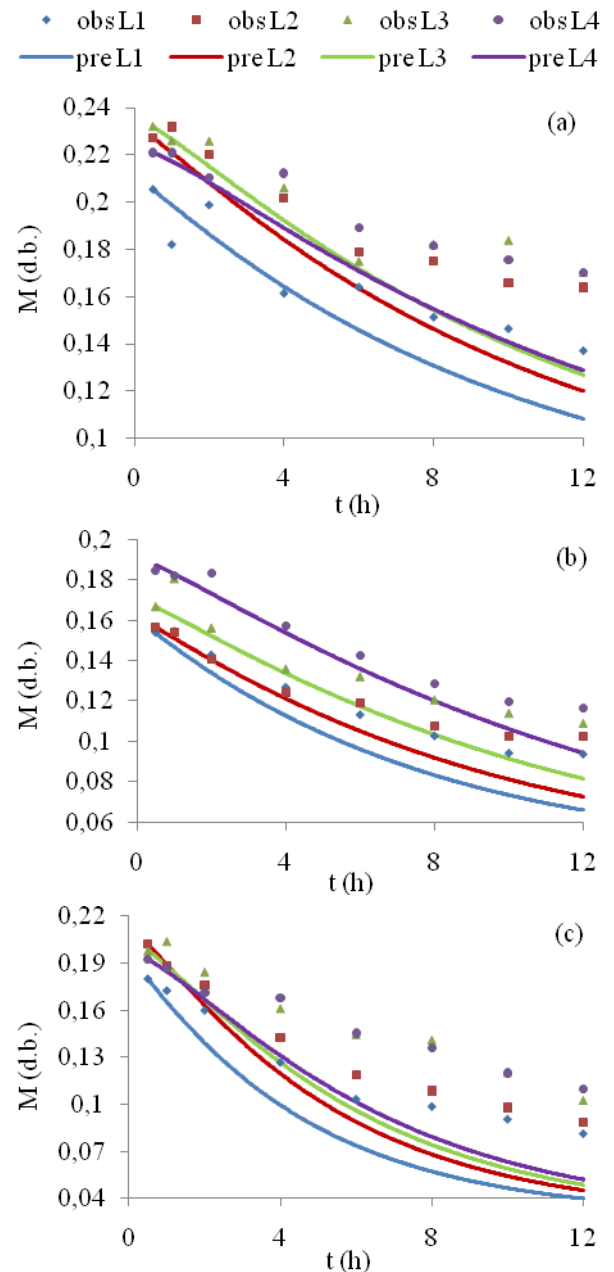


Figure 5: Observed and MSU predicted values of moisture content of test 40.3 (a), 50.3 (b) and 60.3 (c).

Table 3: Statistical validation parameters of each MSU simulated test.

Test	<i>d</i>	<i>r</i>	<i>i</i>	<i>RSD</i> (%)
40.1	86.72	92.96	80.62	7.96
40.2	92.8	94.23	87.45	5.99
40.3	84.85	93.67	79.48	8.41
50.1	79.04	94.66	74.82	20.91
50.2	87.75	95.14	83.48	12.86
50.3	92.78	98.33	91.23	7.95
60.1	84.04	96.85	81.39	21.58
60.2	85.55	94.92	81.21	19.78
60.3	81.35	88.14	71.7	26.83

It can be observed by comparing Table 2 and Table 3 that all values  $i$  and  $RSD$  of the proposed model are better than the MSU model, indicating that the proposed model has a better performance.

### 3.4. Sensitivity Analysis

Table 4 shows all sensitivity coefficients calculated for each analyzed parameter, organized in increasing order of the  $S_i$  value.

Table 4: Sensitivity coefficients of each analyzed parameter of the proposed model.

$S_i$	$S_M$ (%)	$S_{rh}$ (%)	$S_{T_p}$ (%)	$S_T$ (%)	$S_t$ (%)
$T_s$	119.11	170.24	75.93	85.03	115.16
$M_i$	91.26	46.80	27.65	50.43	46.30
$T_{am}$	29.91	95.58	41.98	34.82	45.30
$V$	53.37	44.12	22.94	31.69	34.71
$rh_{am}$	31.04	82.17	7.17	13.76	31.19
$\rho_{ar}$	30.67	46.81	22.61	28.56	29.87
$L_i$	27.58	42.77	27.17	30.97	28.85
$p_{am}$	23.34	33.78	20.66	24.01	24.32
$c_{ar}$	19.90	29.98	24.15	25.74	22.85
$k$	34.30	14.83	6.56	9.51	15.38
$M_e$	32.19	11.38	5.69	8.04	13.51
$h_{lv}$	14.15	18.98	6.50	9.83	12.73
$h_{fg}$	5.29	9.98	17.26	15.21	9.56
$k_h$	6.69	7.22	11.98	12.80	7.58
$d_p$	0.74	0.80	1.37	1.46	0.85
$c_p$	0.35	0.68	1.15	1.02	0.64
$c_v$	0.23	0.32	0.11	0.17	0.21
$\varepsilon$	0.05	0.08	0.01	0.03	0.04
$\rho_p$	0.01	0.02	0.01	0.01	0.01

It is observed in Table 3 that all input and initial parameters presented high sensitivity coefficients, higher than 24 %. The drying temperature ( $T_s$ ) presented the greatest sensitivity coefficient (higher than 100 %), probably due to the high interference on the mass and heat transfer parameters and air properties.

The mass and heat transfer parameters and air properties presented substantial sensitivity coefficients, higher than 7.5 %. The air properties are only dependent of the thermodynamic conditions. They are well known and described on literature and can't be changed in the model structure. The transfer parameters are dependent of the fluid and particle properties, since the fluid is the air, the product transfer kinetics, represented by specific empiric equations, are the unique reason on changes of the model results.

All product physical properties presented low value sensitivity coefficients, less than 1 %. The variation of these parameters can be neglected, so an

average or initial value can be used in order to simplify the model.

The unique product property that presented a considerable sensitivity was the water latent heat inside the product ( $h_{fg}$ ). Probably, the consideration of this parameter on the model is one reason for a better performance than MSU model.

Furthermore, other reasons that can explain the improved performance are the consideration of pressure variation and layer volume shrinkage. As can be seen in Table 4, the model is very sensitive to the variation of  $L_i$  and  $p_{am}$ , presented a  $S_M$  value of 27.58 % and 23.34 %, respectively.

On the other hand, physical properties variation does not contribute to performance improvement of the proposed model, since the model is not very sensitive to these parameters.

## 4. CONCLUSIONS

From this research, it can be concluded that the proposed model can accurately describe the fixed-bed drying system to dry coffee berries. Furthermore, this model had better performance than the MSU model. Probably this result was due to the consideration of water latent heat inside the product, pressure variation and layer volume shrinkage. However, future studies must be completed to investigate the use of the proposed model to describe other drying systems for other products.

## ACKNOWLEDGMENTS

The authors thank to CNPq and CAPES for the financial support essential to complete this work.

## REFERENCES

- Afonso Júnior, P. C., 2001. *Aspectos físicos, fisiológicos e de qualidade do café em função da secagem e do armazenamento*. Thesis (PhD), Viçosa Federal University.
- Brooker, D. B., Bakker-Arkema, F. W., Hall, C. W. 1992. *Drying and storage of grains and oilseeds*. Westport: The AVI Publishing Company.
- Camargo, A. P., Sentelhas, P. C. 1997. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no estado de São Paulo, Brasil. *Revista Brasileira de Agrometeorologia* 5(1):89-97.
- Chen, C.; Morey, R. V., 1989. Comparison of four ERH/EMC equations. *Transactions of ASAE*, 32:983-990.
- Clarke, R. J., Macrae, R. 1987. *Coffee Technology*, vol 2. Elsevier: London. 321p.
- Corrêa, P. C., Goneli, A. L. D., Afonso Júnior, P. C., Oliveira, G. H. H., Valente, D. S. M. 2010. Moisture sorption isotherms and isosteric heat of sorption of coffee in different processing levels. *International Journal of Food Science & Technology* 45:2016-2022.
- Dalpasquale, V. A., Sperandio, D., Silva, L. H. M., Kolling, E. 2008. Fixed-bed drying simulation of



agricultural products using a new backward finite difference scheme. *Applied Mathematics and Computation* 200:590–595.

- Dincer, I. 1998. Moisture loss from wood products during drying - Part I: moisture diffusivities and moisture transfer coefficients. *Energy Sources* 20:531–539.
- Fritzson, P. 2003. *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*. New York: Wiley-IEEE Press.
- Guiné, R. P. F., Rodrigues, A. E., Figueiredo, M. M. 2007. Modeling and simulation of pear drying. *Applied Mathematics and Computation* 192:69–77.
- Hamby, D. M. 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment* 32: 135-154.
- Henderson, S. M., Perry, R. L., Young, J. H. 1997. *Principles of Process Engineering*. Fourth edition. St. Joseph: ASAE Press.
- Izadifar, M., Mowla, D., 2003. Simulation of a cross-flow continuous fluidized bed dryer for paddy rice. *Journal of Food Engineering* 58:325–329.
- Lecorvaisier, E., Darche, S., Silva, Z. E., Silva, C. K. F. 2010. Theoretical model of a drying system including turbulence aspects. *Journal of Food Engineering* 96:365–373.
- Madamba, P. S.; Driscoll, R. H.; Buckle, K. A., 1996. Thin-layer drying characteristics of garlic slices. *Journal of Food Engineering*, 29:75-97.
- Mohapatra, D.; Rao, P.S., 2005. A thin layer drying model of parboiled wheat. *Journal of Food Engineering* 66:513-518.
- Sfredo, M. A., Finzer, J. R. D., Limaverde, J. R. 2005. Heat and mass transfer in coffee berries drying. *Journal of Food Engineering* 70:15-25.
- Vossen, H. A. M., 1979. Methods of preserving the viability of coffee seed in storage. *Seed Science and Technology*, 7(1):65-74.
- Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., Rowe, C. M. 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research* 90(5):8995-9005.

## AUTHORS BIOGRAPHY

**Emílio de Souza Santos**, Agricultural and Environmental Engineer, graduating in Agricultural Engineering (Universidade Federal de Viçosa). Regional Development Analyst at CODEVASF (Companhia de Desenvolvimento do Vale do São Francisco e Parnaíba). *Work Areas*: Operation, Maintenance and Safety of Hydraulic Infrastructures; Water and Environmental Management.

**Paulo Cesar Corrêa**, Agronomy Engineer, Ph.D. in Agricultural Engineering (Universidad Politécnica de Madrid). Associate Professor of Universidade Federal de Viçosa. Instructor and Researcher at CENTREINAR (Brazilian Grain Storage Training Center). Editor of Brazilian Journal of Stored Agricultural Products. *Research Areas*: Physical Properties of Agricultural Products; Storage and Process of Agricultural Products.

**Brian Lynn Steward**, Electrical Engineer, Ph.D. in Agricultural Engineering (University of Illinois). Associate Professor of Iowa State University. Member of ASEE (American Society for Engineering Education) and ASABE (American Society of Agricultural and Biological Engineers). *Research Areas*: Virtual Prototyping; Dynamics Systems Modeling and Simulation (off-road vehicles and agricultural systems); Machines and Agricultural Automation.

**Daniel Marcal de Queiroz**, Agricultural Engineer, Ph.D. in Agricultural and Biosystems Engineering (Purdue University), Associate Professor of Universidade Federal de Viçosa, Director of CENTREINAR (Brazilian Grain Storage Training Center); President of the Arthur Bernardes Foundation in; President of the Brazilian Society of Agricultural Engineering. *Research Areas*: Machine Design; Precision Agriculture; Grain Drying.

# SCHEDULING PATIENTS BASED ON PROVIDER'S AVAILABILITY

José Sepúlveda, PhD, MPH<sup>(a)</sup>; Waldemar Karwowski, PhD<sup>(b)</sup>; Francisco Ramis<sup>(c)</sup>, PhD; Pablo Concha<sup>(d)</sup>

<sup>(a),(b)</sup> University of Central Florida, Orlando, Florida 32816-2993, USA

<sup>(c),(d)</sup> Universidad del Bío-Bío, Concepción, Chile

<sup>(a)</sup> [jose.sepulveda@ucf.edu](mailto:jose.sepulveda@ucf.edu), <sup>(b)</sup> [waldemar.karwowski@ucf.edu](mailto:waldemar.karwowski@ucf.edu), <sup>(c)</sup> [framis@ubiobio.cl](mailto:framis@ubiobio.cl), <sup>(d)</sup> [pccerilkin@yahoo.com](mailto:pccerilkin@yahoo.com)

## ABSTRACT

We present the approach taken to schedule patient arrivals in the simulation of specialty clinics and their associated Ambulatory Surgery unit at the Orlando VA Medical Center (OVAMC), Orlando, Florida, USA. In these clinics, patients arrive by appointment. A primary care provider (PCP) electronically requests an appointment (consult) with a specialist and a reviewer decides whether and when the consult takes place. Requests for new patient appointments arrive electronically 24-7. The decision depends on the provider's availability. A clerk schedules the patients. Once a scheduling slot is found, a message is sent to the patient who then shows up at the scheduled time. If a specialist is not available within a reasonable time horizon (the goal is within two weeks), the request may be transferred to a community provider on a fee basis. We discuss issues associated to slot sizing and overbooking policy. The measures of performance include percent of requests scheduled within a week and percent of patients seen by the provider within two weeks.

Keywords: appointment scheduling, outpatient scheduling, object-oriented simulation, decision support system

## 1. INTRODUCTION

We presented elsewhere (Bozorgi and Sepúlveda 2011) the simulation of the specialty clinics and their associated ambulatory surgery unit at OVAMC. That simulation has been successfully used to streamline operations, including monitoring patient flow, setting priorities, supervising the assignment of resources, scheduling people, sequencing procedures, surgery scheduling and how the electronic waiting lists are kept.

A preliminary analysis of the initial simulation runs indicates that, once they arrive at the OVAMC for a scheduled appointment, patients can expect to be treated within a reasonable time. The problem seems to be in getting to the OVAMC in the first place; e.g., there are scheduling-related problems. For example, the EWL (Electronic Waiting list) printed on June 1 shows a few active consults (mostly "routine" cases) pending for over a year. The clinic wants at least 98% of all new requests approved and scheduled (or rejected) within a week of the time the request is received. The clinic

wants to schedule at least 90% of all new patients within two weeks of the first request. The percent of consults scheduled within a week of request and the percent of cases completed (patient seen by specialist) within two weeks of the consult request clearly do not meet expectations.

For the daily operations simulation, the main measures of performance (MOPs) are patient time [hours] in the system (from arrival to the clinic until the patient leaves the clinic) and resource utilization. For the revised problem (performance over an extended period of time), the main MOP is the time elapsed [days] from consult request by the PCP until the patient is seen by the specialist. We test different scheduling ideas that will improve compliance with VA goals as well as reduce the average time to treatment and the maximum length of time (beyond the initially provider-specified date) a patient is in the EWL.

## 2. PATIENT SCHEDULING

This work focuses on the approach taken to schedule patient arrivals in the simulation of the outpatient specialty clinics and their associated ambulatory surgery clinic. The simulation focuses on the scheduling process: A primary care provider (PCP) electronically requests an appointment (consult) with a specialist; requests arrive electronically; they are reviewed by a specialist who decides whether to take the case, who will take it, and its urgency. The request is then handed to a clerk for actual contacting the patient and scheduling. Once a scheduling slot is found, a delayed message is sent to the patient who then appears in the simulation at the scheduled time and date. Thus, in our situation, patients do not arrive at random.

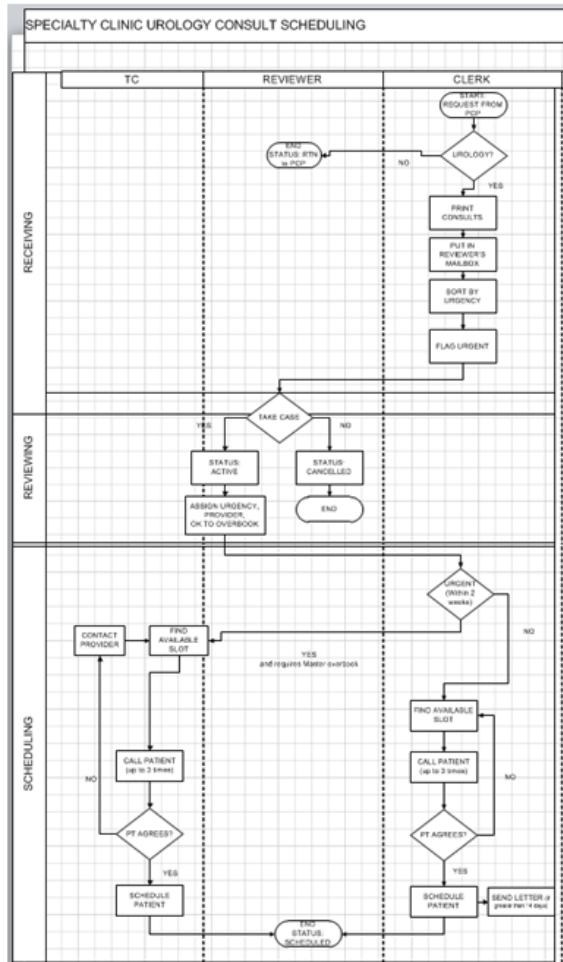
The scheduling decision depends on the provider's availability and willingness to overbook patients, if necessary. If the specialist is not available within a reasonable time horizon (the goal is within a month), the request may be transferred to a community provider on a fee basis.

### 2.1. Consults

The clinic wants at least 98% of all new requests approved and scheduled (or canceled) within a week of the time the request is received. The clinic wants to schedule at least 90% of all new patients within two weeks of the request.

Requests for new patient (“consult”) appointments arrive electronically 24-7. If no reviewer is available, the requests are automatically placed in an electronic waiting list (EWL).

Process maps (see Figure 1) were created through comprehensive and iterative interviews and unit/team staff feedback sessions. These results were validated through follow-up workshops and interactive meetings. Bottlenecks and inefficiencies were identified and presented to the unit staff.



**Figure 1 – Urology Consult Scheduling Process Map**

Each specialty clinic (ENT, Eye Clinic, General Surgery, GYN, Orthopedics, Podiatry, Urology) has one or more (depending on request volume) specialists that review the requests to determine if a new request will be accepted, who will provide the service, the urgency of the request and, for cases that need to be seen urgently, whether overbooking the provider is allowed. Given the same urgency, new requests are processed FIFO. A large majority of requests are accepted. Rejected requests are counted and send back to the originating PCP with an explanation. New patients are assigned to clinic providers on a round-robin basis.

Table 1 presents the daily workload of scheduling clerks by clinic. The second column (“ApptMade/day”) reflects the total number of appointments processed

daily (includes consults and follow-up appointments). The third column (“consults/day”) specifies the number of daily consults. The last column presents the ratio of follow-up appointments per consult. In addition to scheduling patients, clerks check patients in and out of the clinics. This is reflected in the “Check-in/day” column.

**Table 1 Scheduling clerk workload, Specialty Clinics**

Desk	ApptMade/day	Consults/day	Check-in/day	clerks	Follow-up/ Consult
AmbSurg	56.2	12.1	32.8	1	3.6
ENT	39.9	3.4	36.6	1	10.8
GenSurg, ENT	53.4	7.0	36.0	2	6.6
GYN	44.0	6.2	30.9	1	6.1
OPHTH	53.7	24.0	34.4	3	1.2
ORTHO POD	44.4	15.1	31.2	2	1.9
Urology	27.5	7.4	22.9	1	2.7
Total per day	319.2	75.1	225.0	11	3.2

After the reviewer decides what needs to be done, the request is returned to the clerk for scheduling. Urgent requests (patient needs to be seen within two weeks) are scheduled by the Team Coordinator (TC). If necessary, these requests may be overbooked. Non urgent request are scheduled by the clinic’s clerk and they are not overbooked.

## 2.2. Consult Scheduling

Depending on the clinic, scheduling slots may be 15, 20 or 30-minutes. Consults are requests for patients who have not been seen at the clinic within the last 24 months (or ever). Consults require longer time (two slots) with the provider so the provider can get familiar with the case. The assigned time should be sufficient to cover direct contact with patient and the necessary time to record the encounter’s details. If a provider has not finished recording and another patient is waiting for a scheduled appointment, the provider must finish the recording at a later time the same day.

The clerk (or TC) must first find two contiguous slots available for the assigned provider during the window established by the reviewer. If no slots are available, clerk (or TC) must inform the provider who may change the window or allow over-booking the time slots.

The next step is contacting the patient to find out his/her preferred time within the window established by the reviewer. Three phone calls are mandated (if needed, they take place at different times of the day, in different days). Once the patient agrees on a time, the consult request is scheduled. If the patient proposes an alternate time, the clerk may need to consult the provider. If the patient cannot be located, the consult is scheduled and a letter with the appointment time is mailed to the patient.

The consult scheduling process ends when the consult is scheduled. In the simulation, a message is sent to the patient to show up at the appointed time.

### 2.3. Ambulatory Surgery Consults

If a patient is determined to need surgery, a request for a new surgery consult is generated by the surgeon who provides an estimated time for the surgery depending on the procedure and the patient's health condition. Surgery requests are processed through a specialized scheduler (Surgery Coordinator) who finds a time where the surgeon and an operating room are available within one of the weekly time windows assigned to the surgeon's specialty, books the OR, schedules a pre-op visit for the patient, and makes sure that everything is ready and available the date of surgery. The Surgery Coordinator handles all requests for surgery from all specialty clinics. Once a surgery slot has been selected, the Surgery Coordinator asks the clinic's clerk to contact the patient to schedule the pre-op visit and the surgery.

### 2.4. Follow-up visits

After a consult, the provider may decide that a return visit if warranted, usually after the patient gets some exams done. All patient return visits ("follow-up"), as well as lab and imaging exams, are scheduled by the clinic's clerk. Follow-up visits are scheduled for one-slot.

Follow-up visits that need to take place within three months are scheduled at patient check-out time. Those with a larger time horizon are placed on a different electronic list, called the "Recall list." The patient is asked at check-out for a preferred time and the actual schedule is mailed via regular mail about three weeks before the appointment.

## 3. THE SCHEDULING PROCESS

For the surgical specialty clinics in the study, the ratio of consults to follow-up visits is 1 to 3.2 (on the average each new patient returns 3.2 times, i.e., about once every three months, for additional care related to the same consult). For this VA center as a whole, this ratio is closer to 1:6, which may reflect the incidence of chronic illnesses.

From a scheduling perspective, the main difference between consults and follow-ups is that urgent consults are handled by the scheduling supervisors ("team coordinators") while non-urgent consults and follow-up visits are scheduled by the unit's clerk(s).

### 3.1. Clerks

In addition to the clerical processing of new consults, clerks are responsible for scheduling consults and follow-up visits, as well as for checking patients in and out as they come for their appointments. Checking-in entails going over patient's data (address, phone, etc.) and scanning insurance cards. Checking-out involves getting from the patient the desired date and time for the

follow-up appointment, scheduling the appointment (if any) and scheduling labs and imaging exams (if any).

## 4. THE SIMULATION

Finally, we discuss the object-oriented simulation model developed to analyze the described as-is scheduling process situation and its potential alternatives.

We used Flexsim Healthcare as our simulation platform. Flexsim HC is an object-oriented language that provides a library of ready-made software "objects" that have a real-life physical counterpart.

A software object has structure (features) described by attributes (properties) and behaviors (operations). For example, a nurse "knows" what to do when told to transport a patient to a given location, including where to return the wheel chair after use. This helps to create models that "make sense."

Flexsim HC's library (see Figure 2) includes objects for patient arrival, queuing and processing as well as staff and equipment groups. There are also non-human objects ("items") which we used to model electronic requests for appointments.

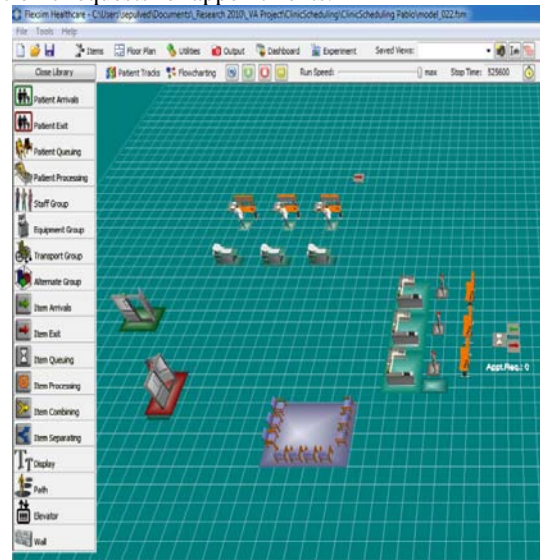


Figure 2 - Flexsim HC library of objects

### 4.1. Simulation details

When the model opens, two global tables are displayed: one is a Summary table; the other is the Appointments table (there is one for each physician).

The Summary table (Figure 3) has 5 rows: Number of requests processed; Average time until request processing; Average time from request until visit; Average patients attended per month; and Total patients attended.



Global Table - PFM_summary	
Name	PFM_summary
Rows	9
Number of requests processed	157.00
Average time until request processing	0.00
Time from request until visit (average)	827.11
Average Patients Attended per Month	0.00
Total Patients Attended	133.00

Figure 3 - Summary Table

The Appointment table (Figure 4) has, for each provider, 365 columns (one per day for a full year) and 20 rows. Each row represents an appointment block. There are up to 20 appointment blocks per physician each day.

Global Table - Appointments\_Physician\_1

Name	Appointments_Physician_1	Rows	20	Columns	365	
	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6
Row 1	470.00	1910.00	3350.00	4790.00	6230.00	999999.00
Row 2	530.00	1970.00	3410.00	4850.00	6290.00	999999.00
Row 3	590.00	2030.00	3470.00	4910.00	6350.00	999999.00
Row 4	650.00	2090.00	3530.00	4970.00	6410.00	999999.00
Row 5	710.00	2150.00	3590.00	5030.00	6470.00	999999.00
Row 6	770.00	2210.00	3650.00	5090.00	6530.00	999999.00
Row 7	830.00	2270.00	3710.00	5150.00	6590.00	999999.00
Row 8	890.00	2330.00	3770.00	5210.00	6650.00	999999.00
Row 9	950.00	2390.00	3830.00	5270.00	6710.00	999999.00

Figure 4 - Appointments Table

Consult requests (“items”) are generated in ItemArrivals1. Items join the ItemQueueing1 queue. The number of items in queue statistics is displayed.

On exiting the queue, items go to Appoint Processing1 where the appointment is processed by a reviewer. The appointment processing time is the time it takes the reviewer to determine if the request is valid.

The request is then queued to a clerk who spends some time finding an open slot for the desired provider. After hours and Saturdays and Sundays are off (block value = 99999) (see Figure 4). If the appointment block is empty (value=0), a patient can be scheduled in that block (two continuous slots are needed for consults; one slot for follow-up visits).

After finding a candidate slot, the clerk must call the patient. When a patient is scheduled, the corresponding block in the Dr.’s appointments table displays the appointment time (in minutes since time 0, e.g., midnight of day 1. For example 470 means 0750 of day 1; 1910 means 0750 of day 2; etc. (See Figure 4)

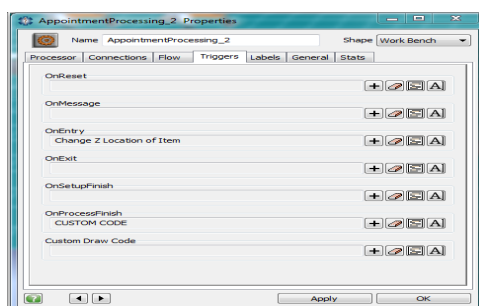


Figure 5 - Message triggering

After the appointment processing (“OnProcessFinish”) (Figure 5), a delayed message is triggered (“OnProcessFinish” CUSTOM CODE). Parameters transferred are (Figure 6): the node where message is sent (Arrival1; this is where the patient will show up in the simulation), the length of the delay (delay\_msg), the current object (person requesting the consult: current), the Dr # (rr\_variable), and if new or established patient (itemtype), and the current time. [“senddelayedmessage(Arrival1,delay\_msg,current,rr\_variable,getitemtype(item),time());”]

```

66  if (gettablenum(tabladoc,xx-1,dia) <= 920+(dia-1)*1440)
67  {
68      double delay_msg = (gettablenum(tabladoc,xx-1,dia)+appt_time)-time();
69      settablenum(tabladoc,xx,dia,delay_msg+time());
70      senddelayedmessage(Arrival1,delay_msg,current,rr_variable,getitemtype(item),time());
71      xx = 999;
72      dia = 999;
73  }
74  else
75  {

```

Figure 6 - Message sending

The delayed message will arrive at “Arrival1” (the entrance node to the simulation) at the scheduled time and will then create the patient.

## 4.2. Scenarios

Scenarios deal with varying number of resources (clerks, providers), varying slot length or number of slots assigned to a visit, and different hours of operations for the facility, individual specialty clinics or selected personnel.

## 4.3. Results

The simulation model accomplishes the following:

1. Keeps track of the number of unprocessed (electronic) requests and the average time elapsed until a schedule slot is assigned to a patient-provider pair.
2. Interacts with an appointments relational database of all scheduled appointments and surgeries, as well as scheduled vacations and other non-available times, on a rolling one-year horizon.
3. Once a visit is scheduled, the simulation model makes the patient show up for the appointment at the scheduled time.
4. Determines the average elapsed time from the initial request until the first visit takes place.
5. Determines the average number of repeat visits.
6. Determines the average elapsed time from initial request until the case is completed.
7. Estimates the average time after the clinic’s closing (“overtime”) the provider must remain in the office finishing the daily dictations.
8. Determines the average throughput (cases completed per month).
9. Determines average backlogs per month

10. Percent requests waiting more than one week for scheduling
11. Percent new requests waiting more than one month for first encounter with a provider.

Note that, except for walk-in cases, patients do NOT arrive at random (electronic requests do). Scheduled patients arrive at their scheduled appointment time (unless they are no-shows, a property that depends on the type of appointment and the specialty involved). Returning patients are scheduled must see their originally assigned provider.

#### ACKNOWLEDGMENTS

The authors acknowledge the assistance and contributions of all personnel at the OVAMC, Orlando, Florida, USA.

#### REFERENCES

- Bozorhi, A., and Sepúlveda, J., Ranking Measures of Performance for Ambulatory Surgery Units, IIE 61<sup>st</sup> Annual Conference and Expo, Reno, NV, May 21-25, 2011
- Chakraborty, S., Muthuraman, K., & Lawley, M. (2010). Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE - Transactions*, 42(5), 354.
- Garcia, H.-M. C., Yun, D. Y., Ge, Y., & Khan, J. I. (1996). Computer-aided resource planning and scheduling for radiological services. *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 2711, pp. 156-166). Newport Beach, CA, USA: Society of Photo-Optical Instrumentation Engineers.
- Guo, M., Wagner, M., & West, C. (2004). Outpatient clinic scheduling: a simulation approach (pp. 1981-1987). Washington, D.C.: Winter Simulation Conference. Retrieved from <http://portal.acm.org.ezproxy.lib.ucf.edu/citation.cfm?id=1161734.1162104&coll=portal&dl=ACM&CFID=79816932&CFTOKEN=49171663>
- Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9), 800-819.
- Mullen, P. M. (2003). Prioritising waiting lists: how and why? *European Journal of Operational Research*, 150(1), 32-45.
- Muthuraman, K., & Lawley, M. (2008). A stochastic overbooking model for outpatient clinical scheduling with no-shows. *IIE Transactions (Institute of Industrial Engineers)*, 40(9), 820-837.
- Qu, X., Rardin, R. L., Williams, Julie Ann S., & Willis, D. R. (2007). Matching daily healthcare provider capacity to demand in advanced access scheduling systems. *European Journal of Operational Research*, 183(2), 812-826.
- Sepulveda, J., Ramis, F., Concha, P., and Neriz, L., The benefits of a library of ready-made simulation objects: A comparison of process-oriented versus object-oriented healthcare simulation, *IIE Annual Conference & Expo*, Cancún, Mexico, June 5-9, 2010
- Vasilakis, C., Sobolev, B. G., Kuramoto, L., & Levy, A. R. (2007). A simulation study of scheduling clinic appointments in surgical care: Individual surgeon versus pooled lists. *Journal of the Operational Research Society* (Vol. 58, pp. 202-211). Houndmills, Basingstoke, Hants., RG21 6XS, United Kingdom: Palgrave Macmillan Ltd. Retrieved from <http://dx.doi.org/10.1057/palgrave.jors.2602235>
- Zhang, B., Murali, P., Dessouky, M. M., & Belson, D. (2009). A mixed integer programming approach for allocating operating room capacity. *Journal of the Operational Research Society*, 60(5), 663-673.

#### AUTHORS BIOGRAPHY

José Sepúlveda, PhD, MPH, is an Associate Professor in the Department of Industrial Engineering and Management Systems in the College of Engineering and Computer Science, University of Central Florida, Orlando, Florida, USA. His major areas of research interest are object oriented programming and expert systems applications in simulation; data mining; and industrial engineering applications in health care environments. He has written two books and numerous professional papers. His e-mail address is [jose.sepulved@ucf.edu](mailto:jose.sepulved@ucf.edu).

Waldemar Karwowski, PhD, is Professor and Chair at the Department of Industrial Engineering and Management Systems in the College of Engineering and Computer Science, University of Central Florida, Orlando, Florida, USA. He is the Director of the Institute for Advanced Systems Engineering at the University of Central Florida. His email is [waldemar.karwowski@ucf.edu](mailto:waldemar.karwowski@ucf.edu).

Francisco Ramis, PhD, is a Professor at the Departamento de Ingeniería Industrial, Universidad del Bío-Bío, Concepción, Chile. He is also the Director of the Center for Health Care Studies, Universidad del Bío-Bío. His email is [framis@ubiobio.cl](mailto:framis@ubiobio.cl).

Pablo Concha is a Senior Analyst at the Center for Health Care Studies, Universidad del Bío-Bío, Concepción, Chile. His email is [pcerilkin@yahoo.com](mailto:pcerilkin@yahoo.com).



# AN OVERALL DHM-BASED ERGONOMIC AND OPERATIONAL ASSESSMENT OF A MANUFACTURING TASK: A CASE STUDY

Nadia Rego Monteil<sup>(a)</sup>, David del Rio Vilas<sup>(b)</sup>, Diego Crespo Pereira<sup>(c)</sup> Rosa Rios Prado<sup>(d)</sup>

<sup>(a), (b), (c), (d)</sup> Integrated Group for Engineering Research (GII). University of A Coruna, Spain

<sup>(a)</sup>[nadia.rego@udc.es](mailto:nadia.rego@udc.es), <sup>(b)</sup>[daviddelrio@udc.es](mailto:daviddelrio@udc.es), <sup>(c)</sup>[dcrespo@udc.es](mailto:dcrespo@udc.es), <sup>(d)</sup>[rrios@udc.es](mailto:rrios@udc.es)

## ABSTRACT

An effective manufacturing workstation design should jointly consider both operational and ergonomic characteristics. Accordingly, the desirability of candidate solutions should rely on the definition and combined assessment of multiple Key Performance Indicators (KPIs). Digital Human Modelling and Simulation (DHMS) provides with a powerful tool capable to precisely recreate and assess human tasks in terms of single and separately treated performance measures. However, the lack of consensus in the adoption of ergonomic parameters and the need of integrating methods have been reported as a fundamental shortcoming when conducting an overall systematic data results analysis. Under an engineering approach, we propose the employment of a control chart integrating biomechanical, postural and operational performance indicators in a continuous framework as well as a Desirability Function (DF) allowing the comparison of several alternative designs. A DHMS-based case study of a manual task in the natural roofing slate manufacturing process is presented.

**Keywords:** Workplace Design, DHMS, Desirability Function

## 1. INTRODUCTION

During the design of new manufacturing workstations or even in the case of re-design of existing workstations several issues need to be simultaneously considered. As a consequence, workstation effective ergonomic design and productivity enhancement usually require the definition and use of multiple Key Performance Indicators (KPIs).

The development of the Digital Human Modelling and Simulation (DHMS) and its further integration in computer aided manufacturing environments represents the most relevant achievement for the efficient and complete consideration of human factors in production systems. However, DHMS alone is not enough in manufacturing ergonomics, because it does not provide appropriate methods for an overall ergonomic risk assessment (Fritzsche, 2010).

However, some authors have coped with the need of definition and integration of multiple KPIs (Wright and Haslam, 1999; Russell et al., 2007) and also proposing multiple objective functions (Ben Gal and Bukchin, 2002). The above reported references reveal a

lack of consensus in the adoption of ergonomic standards or methods. All of this turns out in a confounding panorama for fresh engineers striving to incorporate the ergonomic component into their designs.

It is in the aim of this paper to describe an overall analysis methodology for manufacturing workstations design and re-design. It consists of a task characterization control chart gathering ergonomic and operational KPIs within a continuous framework. This will be used as the basis of an expert-driven design. A desirability function is then used for quantifying the alternatives and supporting a reasonable choice. This methodology has been specifically designed to the slate splitter's task in the context of an improvement study (del Rio et al., 2009; Rego et al., 2010), although it can be adapted and extended to other kind of tasks.

On section 2, some initial considerations are made before explaining the general methodology on section 3. This methodology application and its results are described on section 4.

## 2. PROCESS AND STUDY CONSIDERATIONS

Spanish slate sector for roofing applications is mostly composed of small and medium enterprises (SMEs). The manufacturing process relies on highly labour-intensive activities and more specifically on the mastery of a specialized group of workers known as Splitters. Final product quality, plant's costs and productivity depend largely on their individual performance. Since a lot of repetitive and potentially hazardous movements have to be made there is a substantial risk of developing musculoskeletal disorders (MSDs).

Generally, when dealing with a workstation design some of the following restrictions have to be considered:

- *Financial.* Companies cannot stop their production in order to test new workplace alternatives. In that sense, SMEs are even more sensitive. A practical, quick, but effective approach is appropriate for SMEs which will lead to good results with affordable resources.
- *Technological.* Major changes in the structure of the task are not plausible unless a new automated concept approach is adopted.
- *Process Integration.* Upstream and downstream process parameters should not be altered. Changes

implemented under a myopic and local optimization perspective may decrease the global performance of the plant.

In the splitting case, all of them are found. Accordingly, a simple, quick and cheap video-based data collection is used in order to determine the set of postures and times which the worker usually performs. After a detailed input data analysis, the modelling process may start by reproducing a virtual environment assuring the geometric and operational similarity. The DHM-VE Delmia V5R19 allows a high level of detail in defining the exact sequence of movements even for hands and fingers.

Regarding the analysis stage, Delmia V5R19 provides with a Biomechanical Analysis model, reporting the spinal compression and joint shear forces (L4/L5), among others. Also RULA can be applied to each posture of which a task is made of. Cycle time can be obtained from the Delmia task time estimation and the statistical frequency analysis (from video-recording observation). Energy expenditure is also provided by the specific Delmia tool. More details of the modelling procedure can be found in Rego et al. (2010).

### 3. ANALYSIS METHODOLOGY

The analysis methodology is depicted in Figure 1. There are two parallel processes, the modelling and simulation typical stages (left part of the schema) and the results treatment (the right part). According to the study goals, quantitative KPIs have been defined. Usually, those goals are operational -improving the system productivity or reducing the cycle time- and ergonomic -reducing the injuries risk or improving the working environment-. In the splitters case, the aim is to improve the ergonomic conditions while maximizing the throughput rate. Accordingly, we have defined the following set of key performance indicators (Table 1).

Table 1: KPI definition

KPIs	Explanation
Postural Risk Assessment (P)	The worker's posture is assessed during the task according to RULA. It is especially thought for the assessment of tasks that mainly imply the upper limbs as is the splitters case
Biomechanical Risk For Injury (B)	The Spine Compression value is a complementary measure of risk of MSDs. According to NIOSH guidelines, compression force on the intervertebral disk over 3.4 kN may eventually lead to injuries
Energy Expenditure (E)	The metabolic energy consumption is included, according to Garg guidelines. It is a physiological measure to the amount of effort spent on the task
Cycle Time (T)	It is a measure of the worker's throughput rate

Those KPIs can be either local –a different value for each posture of which a subtask is composed of- or global –a single KPI value for each subtask-. For instance, in the case study, postural risk and L4/L5 spine compression are local KPIs, because they have an evolution over time. On the contrary, time and energy expenditure are global KPIs. Also the goals should be accounted in order to set priorities between KPIs.

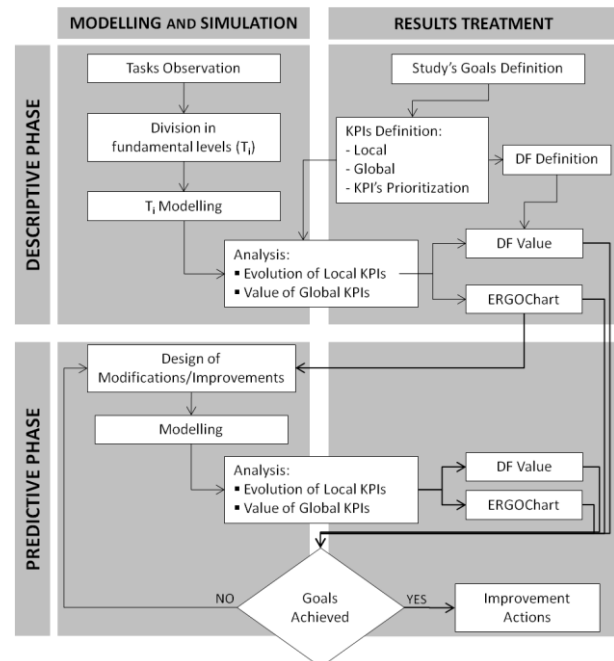


Figure 1: General Methodology for the Overall Analysis

Regarding the M&S part of the schema, present conditions are considered in the “descriptive model”. In the case of an early design, this descriptive model is just the initial scenario, but in the case of an existing task, it represents real events so that a validation phase is necessary. In any case, the work cycle will be divided in fundamental subtasks and then modelled. In the splitter's case, those subtasks consist on successively receiving the block, performing a rough initial cutting, executing three possible types of splitting and finally, three classifying subtasks. Each subtask will provide with a huge amount data (postural risk and spine load for each posture; time and energy expenditure). The resulting data will be graphically arranged into a fictitious time axis and tables and compared to some reference values (ERGOChart). An example is shown in Figure 2. We refer to a fictitious time because the fundamental tasks can occur more than once so that the global cycle time does not correspond with the addition of each single subtask. This chart shows the postural and biomechanical risk evolution over time, in comparison with the maximum recommended values. Also, as complementary information, an energy expenditure table is shown. The use of the overall chart makes it possible to know the ergonomic and operational features of every posture all

along the normal operations. This continuous analysis also enables finding “when” worst posture takes place.

As a result, it can be used for driving design efforts to reduce ergonomic impact and/or increase performance in specific harmful movements or operations. Designs coming from the first model are then implemented in a second model generating new data postures results that will shape a second ERGOChart. This iterative process finishes when the study goals are achieved. However, a local improvement in one single KPI could worsen other KPIs. The desirability function (DF) has been designed to assess the whole task by integrating the defined KPIs accounting for the impact of a specific movement in the complete work cycle. Besides, as opposite to the chart, it represents a quantitative general improvement so comparison between candidate solutions is allowed.

The concept of desirability, introduced by Harrington (1965), is a method for multicriteria optimization in industrial quality management. Via desirability functions (DFs), which allow for comparing different scales of the quality measures (QMs) by mapping them to a [0, 1] interval and the desirability index (DI), the multivariate optimization problem is converted into a univariate one (Trautmann and Weihs, 2005).

As suggested in Ben Gal and Bukchin (2001), a KPI-based desirability function can be developed for a multi-objective design of workstations. The desirability is a geometric average of the individual factors, given a particular weight (exponents,  $a$ ,  $b$ ,  $c$  and  $d$ ):

$$D = (P'^a \cdot B'^b \cdot E'^c \cdot T'^d)^{\frac{1}{a+b+c+d}} \quad (1)$$

$P'$ ,  $B'$ ,  $E'$  and  $T'$  are the [0, 1] normalised values of the KPI values for postural risk, biomechanical features, energy expenditure and cycle time.

Yet an important limitation of this overall methodology is that the models are not parameterized, the proposed ad hoc simulation-driven experimentation is adequate for assessing a whole set of design solutions for which parameterization would become unaffordable. Although a functional optimization is not reachable, important improvements may be attained. Were these exact models parameterized, an optimization approach could be adopted. Some examples in the literature go in that direction (Ambrosse et al., 2002 and Longo and Mirabelli, 2009). In these cases, by means of a DOE methodology different scenarios are automatically analysed. Ben Gal and Bukchin (2002) go a step further by optimising the best solution using Response Surface Methodology (RSM). Nevertheless, a biasing shortcoming arises as only geometrical factors adopting deterministic values are considered in all cases. As well as being fit-for-purpose in the case study, the non parameterized experimentation allows more comprehensive and tailored design alternatives (not only geometric variables).

#### 4. CASE STUDY: METHODOLOGY AND RESULTS

As above defined, desirability is a geometric average of KPIs. In Table 2, they are defined and explained.

Table 2: KPIs for the Desirability Function.

KPI (Formula)	Explanation
<b>Postural Risk</b> $P = \frac{APR + MPR}{2}$	APR/MPR are the average/maximal postural risk over the whole cycle-composed set of postures.
<b>Biomechanical Results</b> $B = \frac{ABR + MBR}{2}$	ABR/MBR are the average/maximal biomechanical result of the spine compression
<b>Energy Expenditure</b> $E$ (kcal/min)	Garg Equation Result, obtained directly from the DHM energy expenditure analysis
<b>Cycle Time</b> $T$ (s)	Cycle time, addition of the different subtasks a complete cycle is made up.

The reason for considering not only the average postural risk / biomechanical results but also their maximum values is that the average is related to a cumulative exposure whereas the maximum provides the information of peak values, so the complete risk of reporting MSD is taken into account.

The values of  $a$ ,  $b$ ,  $c$  and  $d$  are set according with the type of task and the project's goals. In this case, we have considered the following priority factors:

- Reducing the postural risk ( $a=4$ )
- Reducing the cycle time ( $d=3$ )
- Reducing the biomechanical exposure ( $b=2$ )
- Reducing the energy expenditure ( $c=1$ )

In our study, the ergonomic improvement has slight priority over the throughput rate maximization. Also, we consider that the ergonomic improvement is better represented by the RULA postural risk. RULA is a well known and widely used ergonomic assessment method and it is especially thought for the assessment of tasks that mainly imply the upper limbs as is the splitters case. Since the throughput rate is also important, the following factor in priority is the cycle time. Finally, the other two factors have been ordered according to the same priority criteria, i.e., third the biomechanical exposure and finally the energy expenditure.

##### 4.1. SCENARIO H1. Present Conditions

###### 4.1.1. Modelling process

A typical Splitters' work cycle is made up by the following stages:

- Previous Operations: it includes the block reception and the preparation for the operation (T1).
- Rough Splitting: the block is divided into a variable number of pieces depending on its quality and size, by using chisel and hammer (T2).
- Splitting: a variable number of thin plates are obtained by dividing the pieces split in the previous task. Three main situations have been characterized and simulated, i.e., no turning (T3), turning (T4), and rejecting the final plate (T5).
- Sorting and cleaning: sorting depends on the number and quality of the plates. It usually involves lifting tasks. Three main situations have been characterized, i.e., classifying one tile (T6), a group of tiles (T7) and cleaning the workplace (T8).

Each subtask has been modelled by using videos and historical information. Also geometrical data has been incorporated into the models. Figure 1 shows a frame of the model simulation and its equivalent.



Figure 1. Frame of Previous Operations in H1 Model (up) and its Real Equivalent (down)

#### 4.1.2. ERGOChart

Figure 2 shows, the evolution of the postural risk (upper graph) and the L4/L5 compression and joint shear forces (lower graph) along the cycle. The black line represents the L4/L5 (left axis) and the grey line depicts the joint shear forces together with their respective recommended limits (NIOSH). These threshold values -3400 N and 500 N for the compression and the joint shear forces, respectively- are represented by the same red line on two different scales. On the right, the Energy Expenditure (EE) and the time required (CT) for each subtask are also shown.

#### 4.1.3. Interpretation

The RULA postural risk goes from 1 (green zone, the posture is acceptable) to 7 (red zone, changes are required immediately). As it can be noticed, the highest postural risk is achieved during the sorting tasks. In particular, task T7, carrying a group of plates (6 kg) to the rolling table, reaches score 7. The highest level during the splitting is found during the rejection movements.

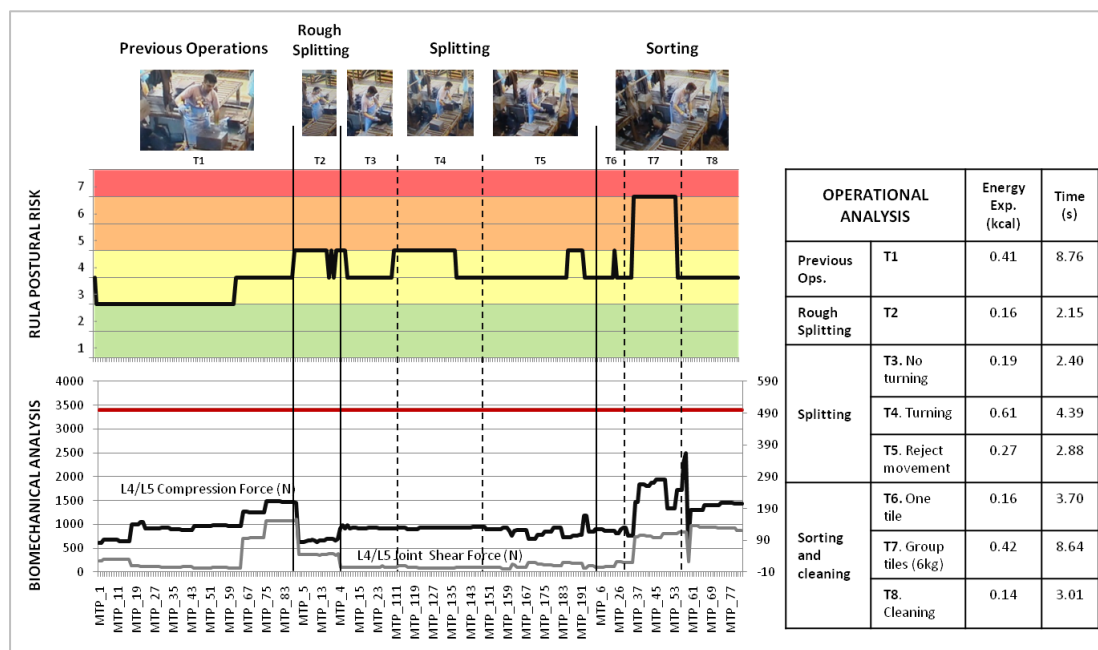


Figure 2. ERGOChart: Ergonomic and Operational Analysis Control Chart

The biomechanical parameters also increased their value during T7. It is remarkable that spine forces also increased at the end of T1, representing the time during the previous operations when the worker bends his back for picking his chisel and hammer from the work surface.

## 4.2. SCENARIO H2: Modified height

### 4.2.1. Motivation

While performing their tasks, splitters combine heavy, light and precision work in a continuous series of quick and repetitive movements so that it is very difficult to identify their corresponding shares and so determining a kind of ergonomically workbench height. Besides, as most of them use a pallet to separate from the wet floor, the effect of height on the ergonomic characterization of the actual operation is far from immediate.

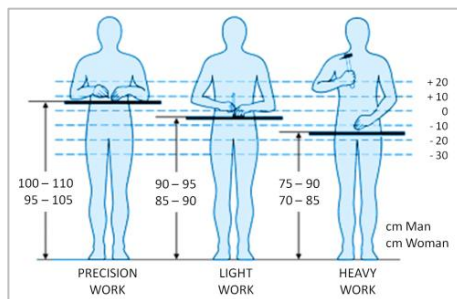


Figure 3: IBV (2011) recommended ranges of optimal workbench heights for precision, light and heavy works

### 4.2.2. Proposal

At present, the work surface height would fit with the heavy work recommended one. Every subtask has been re-modelled considering that the workplace is adequate for light work.

### 4.2.3. ERGOChart

As shown in Figure 3, none of the tasks achieved score 7 in the RULA analysis. Also, compression and joint shear forces decreased 6% and 3% in average (detailed results are shown in table 1). When comparing EE and CT, there is a 13% and 4% reduction, respectively.

### 4.2.4. Interpretation

In general, it would be better to work under those conditions so we assumed this scenario for further experimentation.

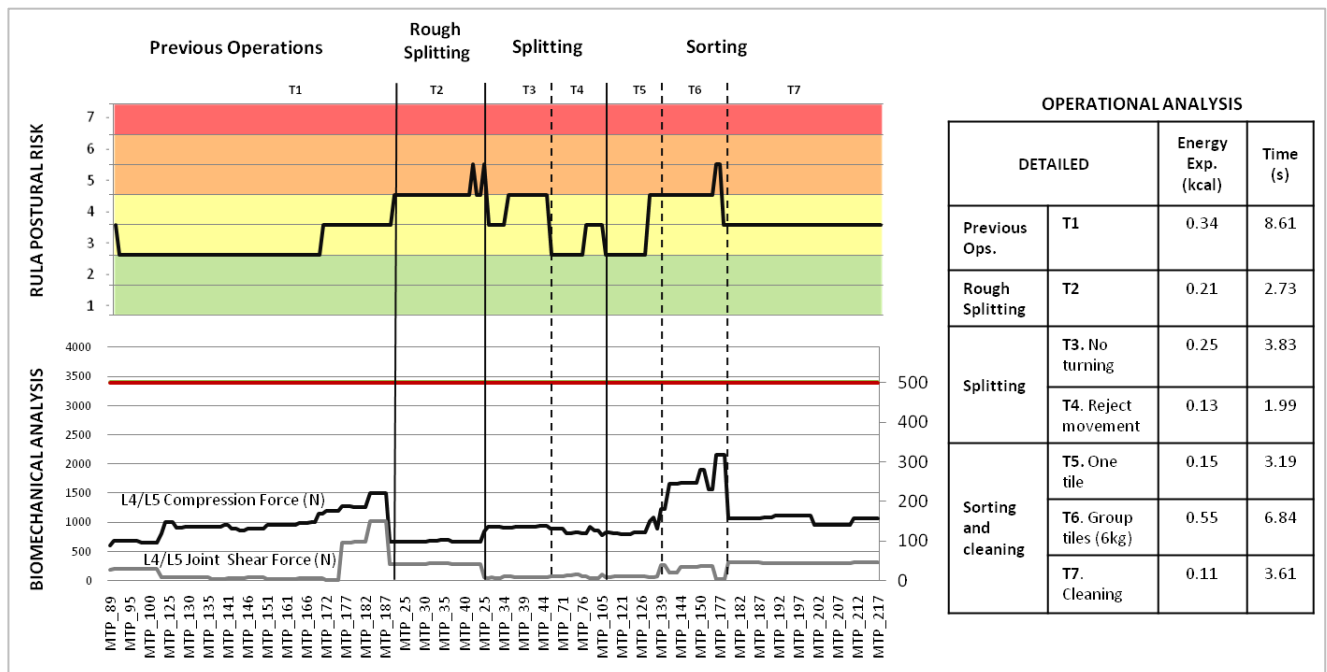


Figure 4. Ergonomic and Operational Analysis under H2 Conditions

### 4.3. SCENARIO H2 – A. Modified height and frontal rejection movement

### 4.3.1. Motivation

Both on H1 and H2 ERGOCharts, the postural risk of the reject movement reaches level 5 (changes are needed). In fact, every time the labour has to reject a block, part or plate, he is obliged to make a lifting and twisting movement.

### 4.3.2. Proposal

Were a connecting conveyor belt to the waste line provided, this highly undesirable movement could be changed for a frontal, safer and faster push movement.

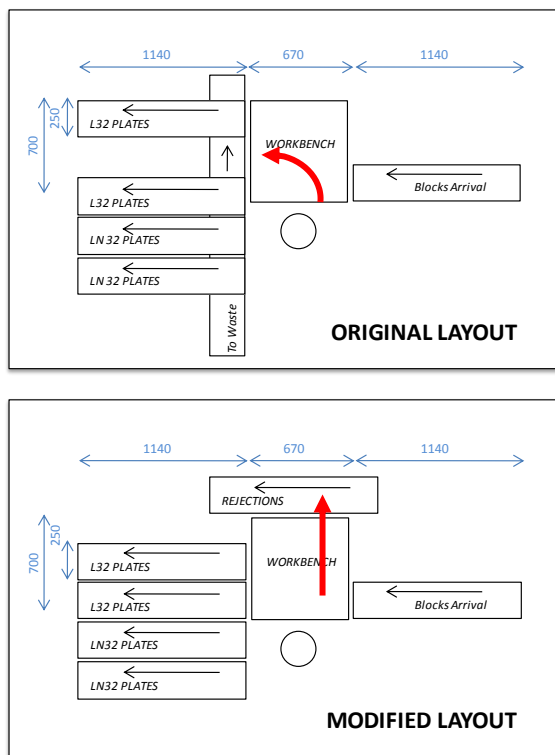


Figure 5: Original and Proposed Layouts.

#### 4.4. SCENARIO H2 – B: Modified height and tool-belt

#### 4.4.1. Motivation

Attending to the H1 and H2 ERGOChart, we can see that the value of postural risk increases from 3 to 4 in both cases at the end of the task. This is because splitters are forced to bend his back whenever they pick their tools from the table. These are unnecessary movements that should be avoided.

#### 4.4.2. Proposal

The employment of tool belts among the splitters, unlike other artisanal professionals, would be a brand new action, not only in this particular company, but in the whole sector. This simple accessory facilitates the work as it ensures comfortable reaching out to take the chisels without the need to bend forward or sideways.

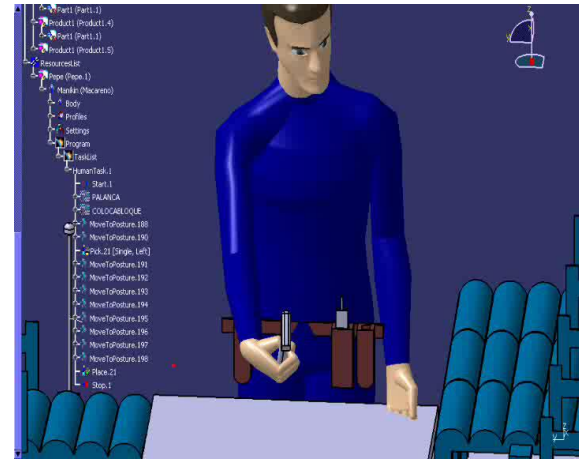


Figure 6. H2 – B Manekin with the toolbelt

#### 4.5. SCENARIO H2 – C: Modified height, radial distribution

### 4.5.1. Motivation

This initiative is the outcome of a specific design effort aimed at reducing the ergonomic impact and improving productivity of the sorting tasks, which show the highest risk levels.

### 4.5.2. Proposal

A radial distribution scheme is presented as an innovative conceptual design in this sector. The distribution conveyor belts have been located accordingly to their frequency of use; blocks arrive from the right side of the splitter and rejected materials are this way frontally pushed away, which is a much safer and faster operation.

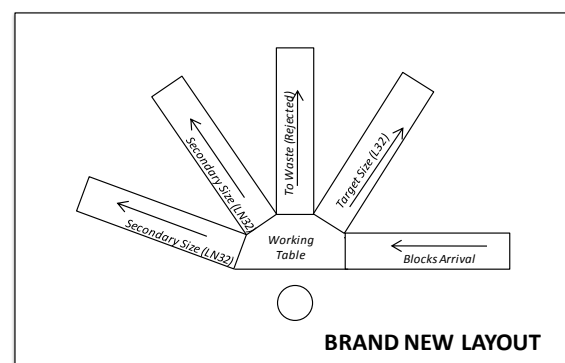


Figure 7. Proposed Layout.

## 5. DESIRABILITY ASSESSMENT

Table 3 summarizes descriptive (H1) results and the predictive improvement measures (H2, H2-A, H2-B, H2-C) with their corresponding KPI results. As it can be noticed all the proposals improve the ergonomic and the operational parameters. These values have been normalised and introduced in the DF. Results are presented on table 4.



Table 3. KPIs values for the Descriptive and Predictive Models

SUMMARY		H1	H2	H2A	H2B	H2C
T (s)		208.7	181.6	179.4	178.6	176.9
E (kcal/cycle)		13.47	12.88	12.24	12.08	11.70
P	Avg	4.19	3.92	3.90	3.81	3.79
	Max	7	6	6	6	5
B (N)	Avg	1029	993	967	915	894
	Max	2493	2159	2159	2159	1493

Table 4. Normalised KPIs and Desirability Values for each Scenario.

	H1	H2	H2A	H2B	H2C
<b>T'</b>	0,00	0,85	0,92	0,95	1,00
<b>E'</b>	0,00	0,33	0,69	0,79	1,00
<b>P'</b>	0,00	0,53	0,54	0,58	1,00
<b>B'</b>	0,00	0,33	0,35	0,39	1,00
<b>D*</b>	0,00	0,40	0,53	0,59	1,00

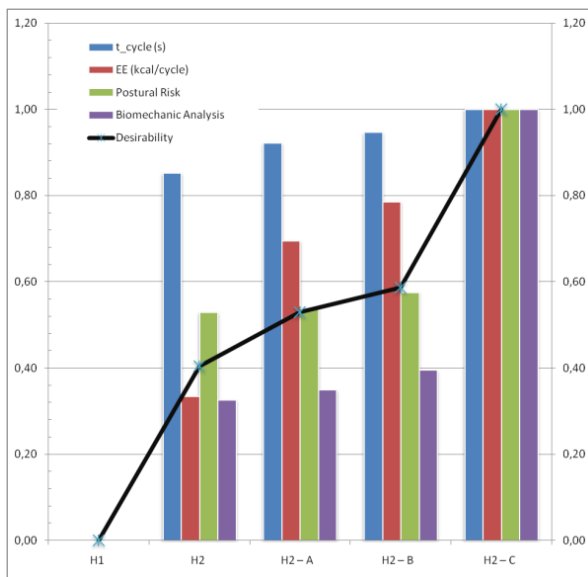


Figure 8: Final Results: KPI and Desirability Values for the Proposed Workplace Designs.

Finally, in Figure 8 the different proposals are assessed according to their desirability value. Every configuration is represented by a set of four bars representing T', E', P' and B' values respectively. Besides, the overall score is depicted by the continuous black line. It can be concluded that:

1. It is more convenient to perform the task under light work recommended height -H2- than within a heavy work height range-H1.
2. The employment of tool belts -H2B- would avoid unnecessary back bending whilst increasing productivity.

3. A simple ramp – H2A- would change the sequence of movements required every time a rejection has to be done to a better ergonomically and faster operation.
4. The best alternative is H2-C (radial distribution), especially in terms of biomechanical exposure. However, a feasibility economical analysis should be carried out since upstream and downstream factors would be affected.

## 6. CONCLUSIONS

A quantitative approach to a combined ergonomic and operational assessment of the slates splitters tasks has been presented by means of a DHMS study.

Under an engineering approach, we propose the employment of a control chart integrating biomechanical, postural and operational performance indicators in a continuous framework. The use of the overall chart allows knowing the ergonomic and operational features of every posture all along the normal operations. Its characterization would be the basis of an expert-driven experimentation stage.

The DF allows assessing the task by integrating both a local change into the whole of the work cycle and a set of KPIs including ergonomic and operational parameters. Such DF value quantifies the level of overall improvement.

This analysis methodology could be easily extended to other manufacturing tasks, especially when there are limited resources or time.

## REFERENCES

- Ambrose, H, Bartels J., Kwitowski A., Gallagher S., and Battenhouse T, 2005, Computer simulations help determine safe vertical boom speeds for roof bolting in underground coal mines. *Journal of safety research* 36, no. 4 pp. 387-97
- Ben-Gal, I., and Bukchin J., 2002, The ergonomic design of workstations using virtual manufacturing and response surface methodology. *IIE Transactions*. Vol. 34, pp.375-391
- Chang, Shao-wen, and Mao-jiun J Wang, 2007, Digital Human Modeling and Workplace Evaluation : Using an Automobile Assembly Task as an Example. *Human Factors* 17, no. 5 pp: 445-455.
- del Rio Vilas, D., Crespo Pereira, D., Crespo Marino, J.L., Garcia del Valle, 2009, A. Modelling and Simulation of a Natural Roofing Slates Manufacturing Plant, *Proceedings of The International Workshop on Modelling and Applied Simulation*, pp. 232-239
- Instituto de Biomecánica de Valencia, 2011. *Factores Humanos para el Desarrollo de Productos. Alturas de Mesas y Bancos para trabajar de Pie*. IBV, Universidad Politécnica de Valencia. Available from: <http://portaldisseny.ibv.org/factoreshumanos/verficha.asp?ficha=128> [accessed 1 July 2011]

- Fritzsche L, 2010, Ergonomics Risk Assessment with Digital Human Models in Car Assembly: Simulation versus Real Life. *Human Factors and Ergonomics in Manufacturing* 20, no.4, pp. 287-299
- Kazmierczak, K., and Neumann W-P, 2007. "A Case Study of Serial-Flow Car Disassembly Ergonomics, Productivity and Potential System." *Human Factors and Ergonomics in Manufacturing* 17, no. 4 pp. 331-351.
- Kharu, O., Harkonen, R., Sorvali, P., Vepsalainen, P., 1981. Observing working postures in industry: Examples of OWAS application. *Applied Ergonomics*, 12, pp.13-17.
- Lin R.T., Chan C.C., 2007. Effectiveness of workstation design on reducing musculoskeletal risk factors and symptoms among semiconductor fabrication room workers. *International Journal of Industrial Ergonomics*, 37, pp. 35-42.
- Longo F., Mirabelli G., 2009, Effective Design of an Assembly Line using Modeling & Simulation. *Journal of Simulation*, vol. 3; p. 50-60
- Santos, J, J Sarriegi, N Serrano, and J Torres., 2007, Using ergonomic software in non-repetitive manufacturing processes: A case study, *International Journal of Industrial Ergonomics* 37, no. 3 pp. 267-275.
- Temple R. Adams T., 2000. Ergonomic Analysis of a Multi-Task Industrial Lifting Station Using the NIOSH Method. *Journal of Industrial Technology*, 16, 1-6.
- Trautmann, Heike, and Claus Weihs. 2005. On the distribution of the desirability index using Harrington's desirability function. *Metrika* 63, no. 2 pp. 207-213
- Rego Monteil, N., del Rio Vilas, D., Crespo Pereira, D., Rios Prado, R., 2010, A Simulation-Based Ergonomic Evaluation for the Operational Improvement of the Slate Splitters Work, *Proceedings of the 22nd European Modeling & Simulation Symposium*, pp. 191-200
- Wright E. J. and Haslam R.A., 1999. E.J. Wright and R.A. Haslam, Manual handling risks and controls in a soft drinks distribution centre. *Applied Ergonomics*, vol. 30, pp. 311-318.

#### AUTHORS BIOGRAPHY

**Nadia Rego Monteil** obtained her MSc in Industrial Engineering in 2010. She works as a research engineer at the Integrated Group for Engineering Research (GII) of the University of A Coruna (UDC), where she is also studying for a PhD. Her areas of major interest are in the fields of Ergonomics, Process Optimization and Production Planning.

**David del Rio Vilas** holds an MSc in Industrial Engineering and has been studying for a PhD since 2007. He is Adjunct Professor of the Department of Economic Analysis and Company Management of the UDC. He has been working in the GII of the UDC as a

research engineer since 2007. Since 2010 he works as a R&D Coordinator for two different privately held companies in the Civil Engineering sector. He is mainly involved in R&D projects development related to industrial and logistical processes optimization.

**Diego Crespo Pereira** holds an MSc in Industrial Engineering and he is currently studying for a PhD. He is Assistant Professor of the Department of Economic Analysis and Company Management of the UDC. He also works in the GII of the UDC as a research engineer since 2008. He is mainly involved in the development of R&D projects related to industrial and logistical processes optimization. He also has developed projects in the field of human factors affecting manufacturing processes.

**Rosa Rios Prado** works as a research engineer in the GII of the UDC since 2009. She holds an MSc in Industrial Engineering and now she is studying for a PhD. She has previous professional experience as an Industrial Engineer in an installations engineering company. She is mainly devoted to the development of transportation and logistical models for the assessment of multimodal networks and infrastructures.

# A COMPUTATIONALLY EFFICIENT AND SCALABLE SHELF LIFE ESTIMATION MODEL FOR WIRELESS TEMPERATURE SENSORS IN THE SUPPLY CHAIN

Ismail Uysal<sup>(a)</sup>, Jean-Pierre Emond<sup>(b)</sup>, Gisele Bennett<sup>(c)</sup>

<sup>(a,b)</sup> College of Technology and Innovation, University of South Florida Polytechnic, Lakeland, FL, USA

<sup>(c)</sup> Georgia Tech Research Institute, Georgia Institute of Technology, Atlanta, GA, USA

<sup>(a)</sup> [iuysal@poly.usf.edu](mailto:iuysal@poly.usf.edu), <sup>(b)</sup> [jpemond@poly.usf.edu](mailto:jpemond@poly.usf.edu), <sup>(c)</sup> [gisele.bennett@gtri.gatech.edu](mailto:gisele.bennett@gtri.gatech.edu)

## ABSTRACT

Radio frequency identification (RFID) enabled temperature tracking technologies are used to monitor perishables such as fresh produce and pharmaceuticals during storage and transportation to validate the temperature integrity of the supply chain. With the help of RFID readers, the data stored in the memory of an RFID tag can be up-linked to a computer for further information processing. In this study, we develop a computationally-efficient, quality index based shelf life estimation model which operates on the stored temperature data in an RFID tag's sensor memory to predict the remaining shelf life using a parametric matrix. The advantages of the proposed model over conventional approaches like Arrhenius equation include: multi-component quality analysis, scalability to higher dimensions with additional environmental parameters such as humidity, greater control over the trade-off between accuracy and complexity and finally adaptability to application requirements and sensory device capabilities such as memory capacity and sampling speed.

Keywords: RFID, shelf life estimation, food quality, food safety, temperature estimation

## 1. INTRODUCTION

The shelf life estimation model discussed in this paper is developed as part of a project which employs an RFID system with handheld readers and temperature tags to estimate the remaining shelf life of army rations to increase food quality and safety. There are multiple approaches to shelf life modeling and estimation for perishable products such as the Arrhenius equation which formulates how the rate of a chemical reaction changes with temperature as shown below:

$$k = Ae^{\frac{-E_a}{RT}}$$

where the rate constant  $k$  of a reaction depends on the temperature  $T$ , whereas the pre-exponential factor  $A$ , the activation energy  $E_a$  and the gas constant  $R$  do not change with temperature (Petrou et. al. 2002). The coefficients in this equation such as  $k$  and  $E_a$  depend on the type of product and are determined experimentally by observing the speed of degradation at different

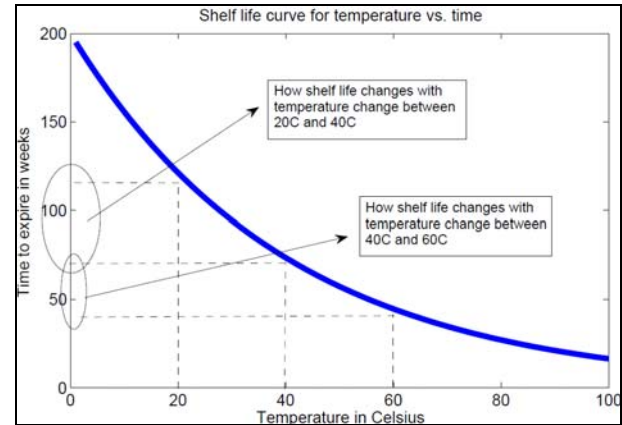


Figure 1: A typical shelf life curve for a product based on Arrhenius equation.

temperatures. Based on such experimental observations and the exponential form factor of Arrhenius equation, a typical shelf life curve which shows the time to expire vs. temperature is shown in figure 1 for the army rations used in this project. At normal temperatures, the army rations have a minimum shelf life of two years as can be observed from the above figure and the time to expire falls with increasing temperature similar to other perishable products. Given the average temperature during a shipment and storage cycle, one can approximate the remaining shelf life on a pallet of army rations by looking at the plot in figure 1.

However, in a typical sensory monitoring application, the tracked environmental variables, such as temperature points, are stored as pairs of value & timestamp instead of a single average temperature value. For example, if the sensory device is sampling the temperature every  $n$  minutes, in  $N$  minutes of total sampling time there are  $N/n$  pairs of temperature & timestamp data points which provide better insight to the temperature distribution within the supply chain. However, in order to utilize such information with higher resolution, unlike Arrhenius model, one needs to utilize more data driven and scalable models which can accommodate the capabilities of the sensory device. In addition, with advance of better sensory devices, other environmental variables, such as humidity can be tracked efficiently and should be used for more accurate shelf life estimation. Finally, today's customer experience oriented approach to perishable supply

chain, such as food qualities ranging from taste to color to odor, should be accounted for in a robust model. In this paper we propose a computationally efficient, quality index and parametric matrix based shelf life estimation model which is built open the multiple quality curve approach to perishables (Nunes 2008).

## 2. SHELF LIFE ESTIMATION MODEL

The quality models used in this project for different items in a package of army rations are the following five qualities: appearance, flavor, odor, texture and overall quality. Each of these quality factors has a quality index (QI) which ranges between 9.0 and 1.0 where 9.0 indicates the highest (initial) quality and 4.0 indicates the lowest acceptable quality. In order to find the QI values, periodic taste panels are performed where different batches of products which are exposed to different temperatures equally distributed within 27°C and 60°C, are sampled and their qualities assessed by trained participants. Through these taste panels, multiple sets of time-temperature data are obtained for each of the five qualities. An example data set is shown in table 1 for a storage temperature of 49°C.

Table 1: Taste panel results for a packaged item from army rations stored at 49°C and sampled at every 2 weeks

Weeks	App.	Aroma	Flavor	Texture	Overall
0	7.9	7.9	8.0	8.0	7.9
2	5.8	6.0	5.8	5.8	5.8
4	4.7	4.9	4.6	4.3	4.4
6	3.8	4.0	3.8	3.6	3.9
8	3.1	3.6	3.3	3.1	3.2
10	2.8	3.3	3.1	3.0	3.0

These results are averaged over all the people who participated in the taste panel and the colored entries show when the QI first drops below the threshold value of 4.0 and the product is declared inconsumable.

### 2.1. Formulation of Quality Index Change

The QI value of a specific product-quality pair depends on three important factors: the previous QI and the time and the temperature between the last sampling instant and the present. The following equation summarizes the relation between the past and present QI.

$$Q_{I_{current}} = Q_{I_{previous}} - f(Q_{I_{previous}}, t, T)$$

where  $Q_{I_{current}}$  and  $Q_{I_{previous}}$  indicates the current and previous QI values after time  $t$  and average temperature  $T$ . As shown in the equation, the drop in the QI is a function of the previous QI, time and temperature. Hence the problem of constructing a shelf life model with this approach can be reduced to finding where the current QI drops below the acceptable threshold of 4.0, or in other words, finding the function  $f$ .

Unlike Arrhenius equation, the function,  $f$ , needs to be defined and is different for each quality value which requires the use of multiple quality models, or in other

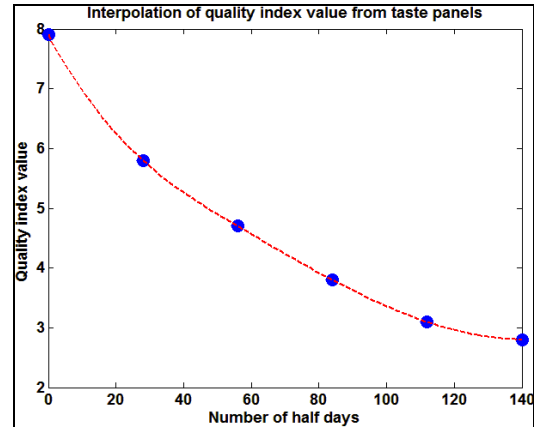


Figure 2: The interpolation of quality index value for the appearance quality of the product in table 1.

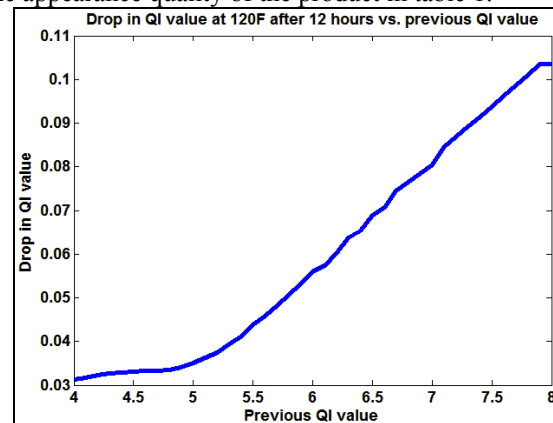


Figure 3: The plot that shows how much QI will drop depending on previous QI value, temperature and time (note 120°F = 49°C)

words, a parametric matrix. For example, for the product in table 1, we only have 6 sampling instants in time which are 2 weeks apart. However, the RFID temperature sensor used in the project samples twice every day. In order to construct a shelf life model which can make full use of this information, the first interpolation will take place in time, as shown in figure 2. The next step is to find how QI changes with time at a specific temperature and previous QI value. Since the time between sensor sampling instants will be fixed for an application (in this case 12 hours) one can use the values in figure 2 to find for each possible QI value between 4.0 and 8.0, how much the QI drop will be after 12 hours at 49°C as shown in figure 3. However, since there are only 4 experimental temperature points, a second 2-dimensional interpolation is performed between the QI drop vs. previous QI value curves for different temperatures. In order to make the algorithm run on a handheld RFID reader such as the one used in the project, this interpolation can be sampled at discrete temperature points (such as every 1°C) to create a set of parametric 2x2 matrices for each quality. Given a previous QI value and temperature, such a matrix provides the corresponding quality drop in the previous equation, thereby serving as the estimation function  $f$  as shown in figure 4.

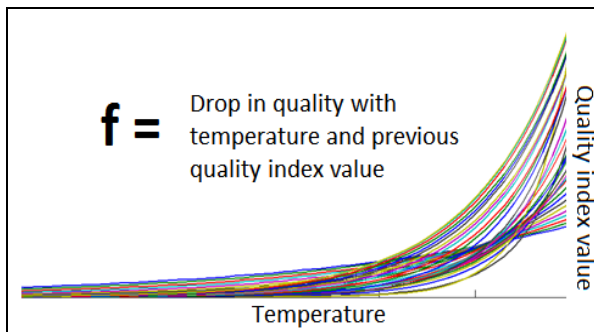


Figure 4: The function,  $f$ , which outputs the drop in quality with temperature and previous QI value given initial QI value (as indicated by lines with different colors)

Hence, the proposed model is not only computationally efficient, as it only uses look-up tables for shelf life estimation, but also has complete control over the trade-off between accuracy and performance. In addition it is scalable to other parameters such as humidity and its parametric approach can be modified based on sensory device capabilities and specific requirements of the application.

### 3. TEMPERATURE ESTIMATION MODEL

Temperature tracking and shelf life estimation of perishable products such as fresh produce or temperature sensitive pharmaceutical drugs during their transportation has been vital to ensure the quality degradation remains acceptable when the items reach their end destination. RFID enabled temperature loggers take this to the next level by adding the capability of wireless data transfer to remotely monitor the temperature inside a shipping container (Opasjurnskit et. al. 2006). Such information would pave the way for intelligent distribution practices such as first-expired-first-out (FEFO) instead of the more commonly used first-in-first-out (FIFO) based on the differences in the temperature profiles witnessed during transportation of individual shipments.

However, due to the dynamics of heating/cooling cycles, it is well known that the temperatures in close vicinity of the products may be different than the temperatures measured inside the shipping container itself (Faghri et. al. 2010). For instance, the temperature inside the container can change more rapidly than the temperature inside a pallet of tightly packed food products. In some cases this might be helpful; such as when the temperature inside the pallet is low and the temperature inside the container increases rapidly. In such a scenario, it will take longer for the pallet temperature to rise to the level of container temperature, which might have a positive impact on the remaining shelf life. However, in the opposite case, even when the temperature inside the container cools down, it will take longer for the pallet temperature to come down as well that will negatively affect the remaining shelf life. In summary, it is crucial to measure the temperature inside the pallet rather than the

container to have a more accurate representation of the remaining shelf life.

The limitations of RFID technology, such as reduced performance near metals and liquids might prevent placing RFID temperature tags inside pallets with significant metal and liquid content. One example to this would be First Strike Rations (FSR) as they are shipped and stored in tightly packed pallets where the shipping or storing temperatures can exceed 150-160°F, which results in reduced shelf life. Since the algorithm developed in this paper deals with the estimation of remaining shelf life based on the temperatures measured by RFID temperature loggers during shipment and storage, in order to best monitor the temperature inside the pallet, the trivial solution is to place the temperature tags inside for more accurate measurement. However, this causes serious problems during the interrogation of these tags due to all the metal and liquid content of the rations inside the pallet. Hence, the tags are placed outside the pallet, measuring only the temperature inside the container. One can use this information as an approximation to the actual temperature inside the pallet and calculate remaining shelf life based on this data, however there is a more accurate way of estimating the temperature in close vicinity of the products inside the pallet.

The following study was performed to find the correlation between the temperatures inside and outside a pallet of army rations during heating/cooling of the environment. Five different experiments were carried out where a pallet of army rations was exposed to different heating/cooling cycles. One temperature sensor was placed outside the FSR pallet and two sensors inside the pallet, labeled as *Alpha* and *Prime*. Based on the heating/cooling intervals, the tests can be divided as follows: 6 hours/6 hours, 18 hours/18 hours, 24 hours/24 hours, 2 days/2 days and 4 days/4 days.

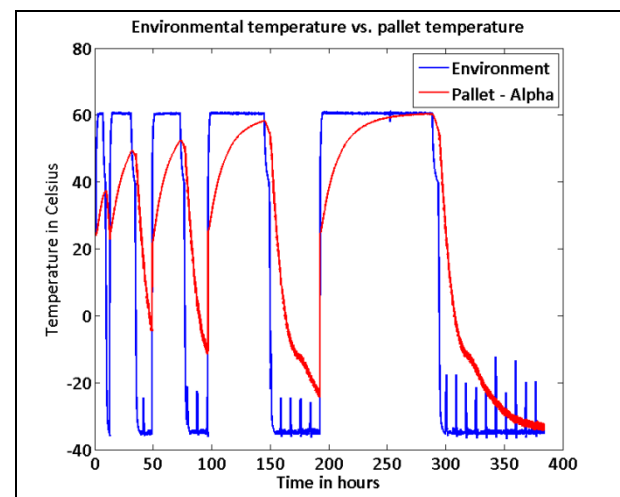


Figure 5: The temperature change inside the pallet at point labeled *Alpha* and how it fluctuates with the change in environmental temperature



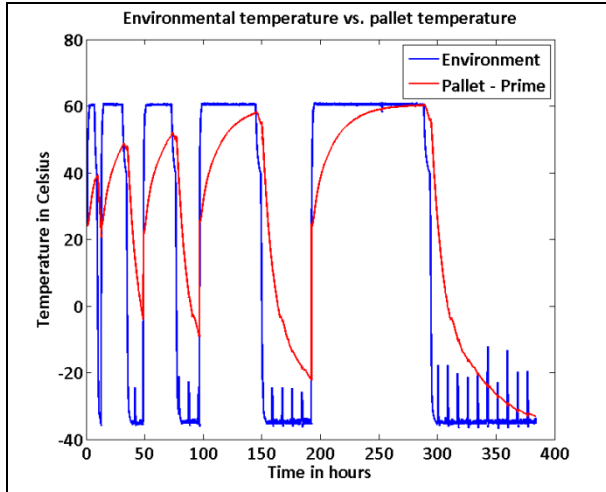


Figure 6: The temperature change inside the pallet at point labeled *Prime* and how it fluctuates with the change in environmental temperature

Figure 5 shows a concatenated plot for the measured temperature levels inside and outside the rations pallet during the entire study for the point labeled *Alpha* inside the pallet. Similarly, figure 6 shows the same plot for the point labeled *Prime*. As expected, the temperatures inside the pallet show a capacitor effect to rapidly changing temperature by slow heating/cooling cycles. In other words, the temperatures rise and decay with a time constant (though slightly different for the two points) that can be determined from these five experiments and can subsequently be modeled to estimate the temperature inside the pallet given the environmental temperature.

In order to better explain this phenomenon, let's take a look at a strikingly similar analogy where the environmental temperature is modeled by the potential difference between the terminals of a voltage source and the pallet temperature is modeled by the potential difference between the terminals of a capacitor as shown in the electronic circuit of figure 7 (Paul 2001).

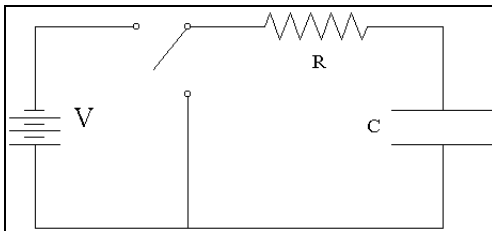


Figure 7: A typical resistor-capacitor circuit to simulate the behavior of the pallet temperature in the presence of changing environmental temperature

In this figure,  $V$  represents the environmental temperature and  $V_c$  represents the pallet temperature. The relation between the two temperatures can be explained by the following equation:

$$V_c = V_c^{initial} + (V - V_c^{initial}) \left(1 - e^{\frac{-t}{RC}}\right)$$

where  $V_c^{initial}$  is the initial pallet temperature,  $R$  is the resistance of the resistor and  $C$  is the capacitance of the capacitor. In other words, the pallet temperature will rise or fall with a speed determined by the time constant ( $RC$ ) and the difference between the temperature of the pallet and the environmental temperature. The bigger the difference the faster temperature will change inside the pallet.

In order to find the time constant empirically, one has to change the potential  $V$  and observe how the potential  $V_c$  changes with time. Both figures 5 and 6 provide enough information on how to find the time constant for both rising and falling temperatures. If we rearrange the terms in the above equation to find  $\tau$  ( $RC$ ), we arrive at the following equation:

$$\tau = \frac{-t}{\ln \left( \frac{V - V_c}{V - V_c^{initial}} \right)}$$

Hence, if one knows the temperature inside the pallet at a given time,  $t$ , the initial temperature inside the pallet and the environmental temperature, it becomes trivial to calculate the time constant. Unlike the electronic circuit described above, it is possible to have a different time constant for heating and cooling cycles and the way experiments are designed will allow for separate calculation of the two.

Let's take a look at the last experiment to calculate  $\tau_{rising}$  for *Alpha* point. In this experiment, the average environmental temperature sits at 60.5°C. If we define  $t = 0$  as the time the pallet temperature started to increase from 24.6°C, we can then choose a 2<sup>nd</sup> temperature point-time pair to calculate the time constant. For this example, let's look at the last study (4 days/4 days) where it takes  $t = 76$  hours for the pallet temperature to reach 60°C, and this point will be used in the calculations below where  $T$  represents temperature and  $t$  represents time.

$$T_{pallet}(t = 0) = 24.6C$$

$$T_{environmental} = 60.5C$$

$$T_{pallet}(t = 76 \text{ hours}) = 60C$$

Thus;

$$\tau_{rising} = \frac{-76}{\ln \left( \frac{60.5 - 60}{60.5 - 24.6} \right)} = 17.8 \text{ hours}$$

Similarly, to find  $\tau_{falling}$ , one only need define two time-temperature points where the pallet temperature slowly approaches the environmental temperature. We choose the same study using a 4 day cooling period to determine the falling time constant.



$$T_{pallet}(t = 0) = 60.5C$$

$$T_{environmental} = -35C$$

$$T_{pallet}(t = 67.5 \text{ hours}) = -33C$$

$$\tau_{falling} = \frac{-67.5}{\ln\left(\frac{-35 - (-33)}{-35 - 60.5}\right)} = 17.5 \text{ hours}$$

Based on these time constants, estimating the temperature inside the pallet can be modeled by the following two equations.

$$\begin{aligned} &\text{If } T_{environmental} > (T - 1)_{pallet}, \text{ then...} \\ &T_{pallet} = (T - 1)_{pallet} + (T_{environmental} - (T - 1)_{pallet}) \left(1 - e^{\frac{-t}{\tau_{rising}}}\right) \\ &\text{If } T_{environmental} < (T - 1)_{pallet}, \text{ then...} \\ &T_{pallet} = (T - 1)_{pallet} + (T_{environmental} - (T - 1)_{pallet}) \left(1 - e^{\frac{-t}{\tau_{falling}}}\right) \end{aligned}$$

where  $T_{pallet}$  is the current estimated pallet temperature and  $(T-1)_{pallet}$  denotes the previously estimated temperature sample.

Remember that this model is only an approximation of the actual temperature inside the pallet as measuring this temperature directly is difficult with RFID technology. In order to gauge the performance of this model, let us compare the model output with the actual measured temperature inside the pallet. Figure 8 shows the estimated pallet temperature against the actual pallet temperature measured by the sensor. As clearly observed from this figure, the estimated temperature is much closer to the pallet temperature than the environmental temperature and thus would be a much better candidate to be the temperature profile used in shelf life calculations.

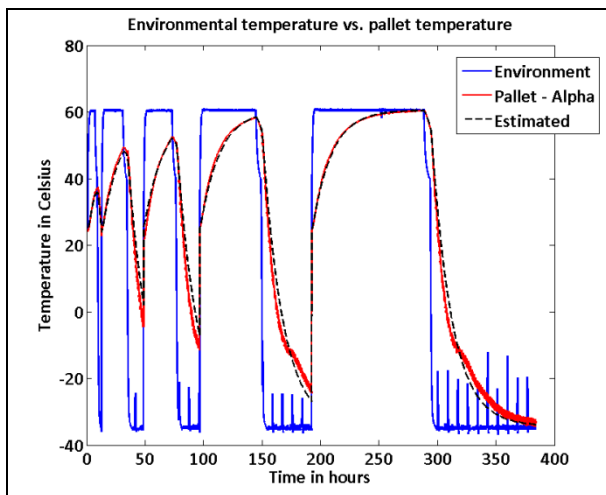


Figure 8. Comparison of environmental temperature, measured pallet temperature and the estimated pallet temperature for *Alpha*

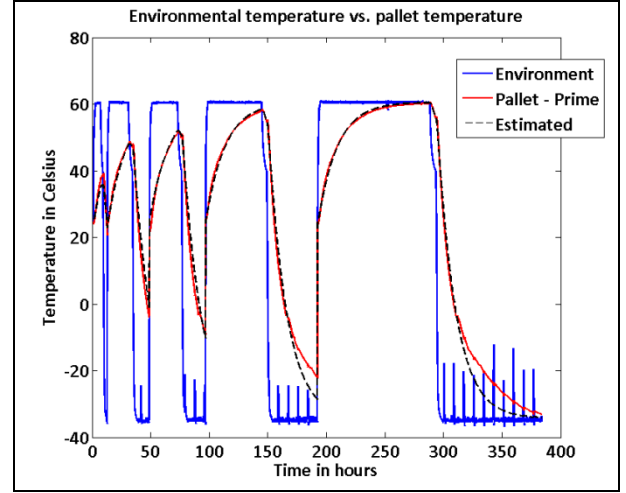


Figure 9. Comparison of environmental temperature, measured pallet temperature and the estimated pallet temperature for *Prime*

In terms of numerical evaluation, the root-mean-square-error and standard deviation between the environmental temperature and the pallet temperature are as follows:

$$\text{RMS}(e)_{\text{environmental-pallet}} = 26.2^{\circ}\text{C}$$

$$\sigma_{\text{environmental-pallet}} = 25.1^{\circ}\text{C}$$

In contrast, the mean error and standard deviation between the estimated temperature and the pallet temperature are much lower.

$$\text{RMS}(e)_{\text{estimated-pallet}} = 2.6^{\circ}\text{C}$$

$$\sigma_{\text{estimated-pallet}} = 2.5^{\circ}\text{C}$$

Even though these calculations were performed for the point *Alpha*, it's similar for the other point *Prime* as well. Figure 9 shows the estimation results for *Prime*.

$$\text{RMS}(e)_{\text{estimated-pallet}} = 2.9^{\circ}\text{C}$$

$$\sigma_{\text{estimated-pallet}} = 2.8^{\circ}\text{C}$$

The slight discrepancy in performance for points *Alpha* and *Prime* can be attributed to the fact that *Prime* shows a different temperature roll-off behavior at subzero temperatures than *Alpha*. This non-uniform behavior cannot be approximated accurately by the single order model. Future work will include piecewise modeling to account for such discrepancies in temperature time constants at different temperature intervals.

To summarize, where one needs to estimate the shelf life of a product based on an RFID tag attached outside the pallet, it is significantly better to estimate the pallet temperature using this model and then use the estimated temperature to calculate the remaining shelf life for much accurate results.

A simple capacitive model, though was shown to be quite effective, can be outperformed by more complex estimation models such neural network or time series estimations. As future work, we will explore these possibilities and more importantly integrate this type of temperature estimation inside the shelf-life

model described in this paper for a comprehensive software approach.

#### 4. CONCLUSIONS

In this paper a computationally efficient and scalable shelf life estimation algorithm is described, specifically designed for wireless temperature sensors in the supply chain which lack the computational and storage capabilities of computers. Furthermore, the complexity of the model can be adjusted for a minimal trade-off in accuracy. Considering the physics of how wireless sensors communicate with their base unit, a temperature estimation algorithm is developed to model the changes in pallet temperature with environmental temperature. Our results show a 90% improvement in root-mean-square-error performance between the pallet temperature and estimated pallet temperature when compared to environmental temperatures.

#### ACKNOWLEDGMENT

The authors would like to thank the U.S. Army Natick Soldier Research, Development and Engineering Center, Combat Ration Research and Development for their support in this research.

#### REFERENCES

- Nunes, M. C. do N., 2008. Impact of environmental conditions on fruit and vegetable quality. *Steward Postharvest Review*, 4(4), 1-14.
- Petrou A. L., Roulia M., Tampouris K., 2002. The use of Arrhenius equation in the study of deterioration and of cooking of foods-some scientific and pedagogic aspects. *Chemistry Education: Research and Practice in Europe*, 3(1), 87-97.
- Opasjumruskit K., Thanthipwan T., Sathusen O., Sirinamarattana P., Gadmanee P., Pootarapan E., Wongkomet N., Thanachayanont A., Thamsirianunt M., 2006. Self-powered wireless temperature sensors exploit rfid technology. *IEEE Pervasive Computing*, 5, 54-61.
- Faghri A., Zhang Y., Howell J., 2010. Advanced Heat and Mass Transfer. *Columbia, MO: Global Digital Press*, ISB 978-0984276004
- Clayton R. P., 2001. Fundamentals of Electric Circuit Analysis. *John Wiley & Sons*. ISBN 0-471-37195-5.

# SIMULATION AS A DECISION SUPPORT TOOL IN MAINTENANCE FLOAT SYSTEMS – SYSTEM AVAILABILITY VERSUS TOTAL MAINTENANCE COST

Francisco Peito<sup>(a)</sup>, Guilherme Pereira<sup>(b)</sup>, Armando Leitão<sup>(c)</sup>, Luís Dias<sup>(d)</sup>

<sup>(a)</sup>Industrial Management Dpt., Polytechnic Institute of Bragança

<sup>(b)</sup>Research Centre ALGORITMI, University of Minho

<sup>(c)</sup>Industrial Management Dpt., Polytechnic Institute of Bragança

<sup>(d)</sup>Research Centre ALGORITMI, University of Minho

<sup>(a)</sup>[pires@ipb.pt](mailto:pires@ipb.pt), <sup>(b)</sup>[gui@dps.uminho.pt](mailto:gui@dps.uminho.pt), <sup>(c)</sup>[afleitao@ipb.pt](mailto:afleitao@ipb.pt), <sup>(d)</sup>[lsd@dps.uminho.pt](mailto:lsd@dps.uminho.pt)

## ABSTRACT

This paper is concerned with the use of simulation as a decision support tool in maintenance systems, specifically in MFS (*Maintenance Float Systems*). For this purpose and due to its high complexity, in this paper the authors explore and present a possible way to construct a MFS model using Arena® simulation language, where some of the most common performance measures are identified, calculated and analysed. Nevertheless this paper would concentrate on the two most important performance measures in maintenance systems: system availability and maintenance total cost. As far as these two indicators are concerned, it was then quite clear that they assumed different behavior patterns, specially when using extreme values for periodic overhauls rates. In this respect, system availability proved to be a more sensitive parameter.

Keywords: Simulation, Discrete Event Simulation, Maintenance, Preventive Maintenance, Waiting Queue Theory, Float Systems.

## 1. INTRODUCTION

In production areas and service systems such as transport companies, health service systems and factories, the main goal is to achieve high levels of competitiveness and operational availability. In this environment the need for equipment to work continuously is very likely in order to maintain high levels of productivity. This is why MFS has an important role on equipment breakdown and production stoppage has a high and direct impact on production process efficiency and, as a consequence, on their operational results. Therefore, maintenance control and equipment use optimization become not only an important aspect for the mentioned reasons, but also for personnel security matters and to prevent negative environmental impact.

In general, preventive maintenance implementation increases equipment control and avoids unexpected stoppages. However, to overestimate these actions makes the maintenance costs too high for the required availability.

In production systems involving identical equipments such as the float systems it is an advantage to integrate maintenance management with materials and human resources. An example of this is to have spare equipment to replace those that fail or need review. Then, the direct and indirect costs due equipment stoppage are minimized and the level of production or service requirements fulfilled. Although the existence of spare equipment is important to maintain the production process working it is recommended to keep the number of spare equipment in an optimal level for economic reasons.

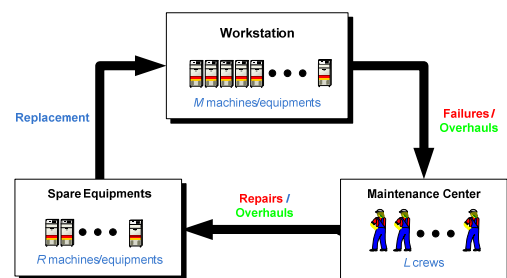


Fig. 1 –Typical Maintenance Float System

A typical Maintenance Float System is composed of a workstation, a maintenance center with a set of maintenance crews to perform overhauls and repair actions and a set of spare machines (Fig.1). The workstation consists of a set of identical machines and the repair center of a limited number of maintenance crews and a limited number of spare machines. The maintenance crew is responsible for the repair and overhaul actions and also responsible for:

- the transportation of the spare machines from the maintenance centre warehouse to the workstation;
- the removal and transportation of the machines needing repair or overhaul action to the maintenance centre;
- the installation of the spare machines in the workstation, replacing machines removed for repair or overhaul action.

After having described the maintenance float system under consideration, the next section of this paper focus on the literature review on analytical models for this type of maintenance system, thus allowing some type of validation for the simulation results achieved.

The following section describes the simulation model developed, based on the purpose of analysing system availability and total maintenance cost.

Next, this paper includes a section for output analysis, in order to evaluate sensitivity, precision and robustness of both performance measures under consideration.

Conclusions and Further Developments are the closing sections for this paper.

## 2. RESEARCH BACKGROUND

As far as float systems maintenance models is concerned, (Lopes 2007) refers some studies where simulation has been used to produce results based on specified parameters. Due to the fact that these simulation models were only concerned with the input/output process, without dealing with what is happening during the simulation data process, some metamodels have emerged (Madu and Kuei 1992a; Madu and Kuei 1992b; Madu and Lyeu. 1994; Kuei and Madu 1994; Madu 1999; Alam et al. 2003). The metamodels express the input/output relationship through a regression equation. These metamodels can also be based on taguchi methods (Madu and Kuei 1992a; Kuei and Madu 1994) or on neuro networks (Chen and Tseng 2003). These maintenance system models were also recently treated on an analytical basis by (Gupta and Rao 1996; Gupta 1997; Zeng and Zhang 1997; Shankar and Sahani 2003; Lopes 2007). However, the model proposed by (Lopes 2007) is the only one that deals, simultaneously, with three variables: number of maintenance teams, number of spare equipments, and time between overhauls, aiming the optimization of the system performance. Although this proposed model already involves a certain amount of complexity it may become even more complex by adding new variables and factors such as: a) time spent on spare equipment transportation, b) time spent on spare equipment installation; c) the introduction of more or different ways of estimating efficient measures; d) allowing the system to work discontinuously; e) speed or efficiency of the repair and revision actions; f) taking into account restrictions on workers timetable to perform the repair and revision actions; g) taking into account the workers scheduling to perform the repair and revision actions; h) taking into account the possibility of spare equipment failure; etc. Anyway these mentioned approaches would aim at ending up with MFS models very close to real system configurations. In fact, the literature review showed that most of the works published, involving either analytical or simulation models, concentrate on a single maintenance crew, or on a single machine on the

workstation or even considering an unlimited maintenance capacity – thus overcoming the real system complexity and therefore not quite responding to the real problem as it exists.

As far as the model presented by (Lopes et al. 2005; Lopes et al. 2006; Lopes 2007) is concerned it is assumed that systems works continuously, its availability is not calculated and the system optimization is only based on the total maintenance cost per time unit. Moreover, it considers that the total system maintenance cost is the same without taking into account the number of machines unavailable, which in many real situations it is not the best option. Finally the referred analytical model only allows that its failures occur under an homogeneous Poisson process (HPP).

Another important aspect on the companies management strategic definition is to have their tasks correctly planned. To help this planning procedure it is important to know different indicators such as: machine availability, equipment performance and maintenance costs, among others. Therefore one should consider new factors that affect these float systems indicators: possibility of some machine failure, efficiency, repair time.

Moreover, when preventive maintenance policy is used, the time for individual replacement is smaller than time for group replacement. It means that the latter situation requires more machine on the process to be stopped, and also implies an increase for a certain time, on the maintenance crews.

In general companies policy lies on using economic models to define their best strategies. Profits maximization or costs minimization are the most frequent goals used. However, strictly from the maintenance point of view availability is frequently used as an efficient measure of the system performance, and sometimes more important than the cost based process. In this work availability is calculated dividing the time the system is up ( $T_{up}$ ) by the time the system is up plus the time the system is down ( $T_{down}$ ) for maintenance reasons. Some authors, however, calculate availability through the ratio between  $MTBF$  and  $[MTBF + MTTR]$ . Being,  $MTBF$  the-Mean Time Between Failures and  $MTTR$  the Mean Time To Repair.

## 3. THE SIMULATION MODEL

The Arena® simulation language environment was chosen for the development of the simulation model for this MFS (Kelton 2004; Pidd 1989; Dias et. al 2006 and Pidd 1993).

This simulation model (see details on Peito et. Al 2011) follows the general assumptions of the analytical model proposed by (Lopes et al. 2005; Lopes et al. 2006; Lopes 2007) when considering a Maintenance Float System with 10 active and identical machines ( $M$ ), 5 spare machines ( $R$ ) and 5 maintenance crews ( $L$ ). The active machines considered operate

continuously. Machines that fail are taken from the workstation and sent to the maintenance park waiting queue, where they will be assisted according to arrival time. Machines that reach their optimal overhaul time are kept in service until the end of a period  $T$  without failures. However they will be also kept on a virtual queue to overhaul. If the number of failed machines plus the number of machines requiring overhaul is lower than the number of maintenance crews available, machines are replaced and repaired according to FIFO (*First In First Out*) rule. Otherwise if it exceeds the number of maintenance crews, the machines will either be replaced (while there are spare machines available) or will be sent to the maintenance queue. The machines that complete a duration period  $T$  or time between overhauls in operation without failures are maintained active in the workstation, where they wait to be assisted, and they are replaced when they are retired of the workstation to be submitted to a preventive intervention. Its replacement is assured by the machine that leaves the maintenance center in the immediately previous instant. If an active machine happens to fail it awaits for the accomplishment of an overhaul, then it will be immediately replaced, if a spare machine is available or as soon it is available.

Furthermore the time to replace a machine that needs overhaul or has failed is also included in our model and this is a parameter that could be adjusted during a simulation run.

In the MFS here analyzed, it is assumed that the  $M$  active machines of the workstation have a constant failure rate (Francisco et al. 2011). Moreover time between failures are assumed as independent and identically distributed following an Exponential Distribution for all machines (failures occur under a *Homogeneous Poisson Process*). However, during a simulation run, this value could be adjusted based on time between overhauls. Obviously a smaller time between overhauls implies greater time between failures.

In this first version of the simulation model the time between overhauls could be adjusted during a simulation run.

As far as time to overhaul and time to repair are concerned, we have assumed the *Erlang-2* distribution, even though considering overhaul time significantly lower than the repair time.

For this MFS, the following three parameters and ten relevant variables are identified:

#### Parameters

1. Number of active machines ( $M$ );
2. Number of maintenance crews ( $L$ );
3. Number of spare machines ( $R$ );

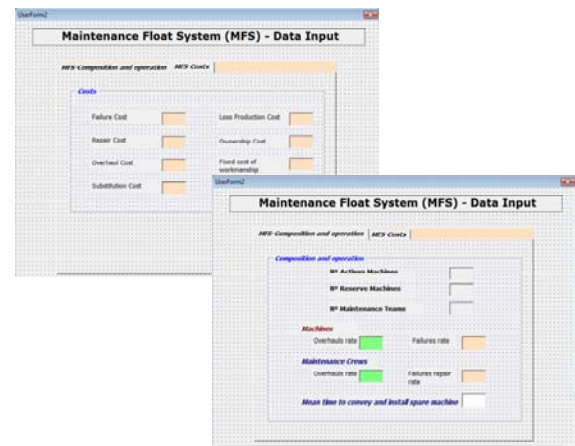
#### Variables

4. Machine- Overhauls rate ( $\lambda_{rev}$ )\*;
5. Machine-Initial Failures rate ( $\lambda_f$ )\*;

6. Crews-Repair rate ( $\mu_{rep}$ )\*;
7. Crews-Overhaul rate ( $\mu_{rev}$ )\*;
8. Failure cost ( $C_f$ );
9. Repair cost ( $C_{rep}$ );
10. Overhaul cost ( $C_{rev}$ );
11. Replacement cost ( $C_s$ );
12. Cost due to loss production ( $C_{lp}$ );
13. Holding cost per time unit ( $h$ );
14. Labour cost per time unit ( $k$ );
15. Time to convey and install spare machine ( $T_{ConvInst}$ ).

(\*) This variable can be adjusted during the simulation run.

The different types of input mentioned above occur in specific developed input menus – Visual Basic for Applications within Arena model.



**Fig. 2** - Data input area sample screenshot

Figures 2 and 3 highlight both input variables window and output updates – numerically and graphically. Fig. 4 shows an application screenshot including simulation animation.

The developed simulation application for a *Maintenance Float System* allows the estimation of the following global efficiency measures:

- a) Average system availability ( $AvgSA_v$ );
- b) Total maintenance cost per time unit [ $(AvgTC_u)$  and  $AvgTC_u(*)$ ].

Total Maintenance Cost per Time Unit would be estimated in two different ways:

- Fixed cost, independent of the number of available machines in the system;
- Variable cost, proportionally dependent on the number of available machines in the system.



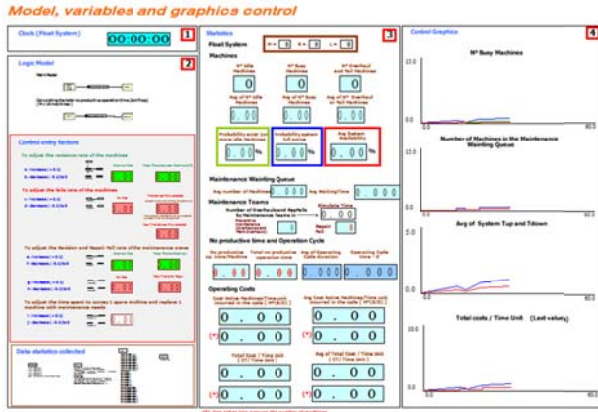


Fig. 3 - Variables and graphics control

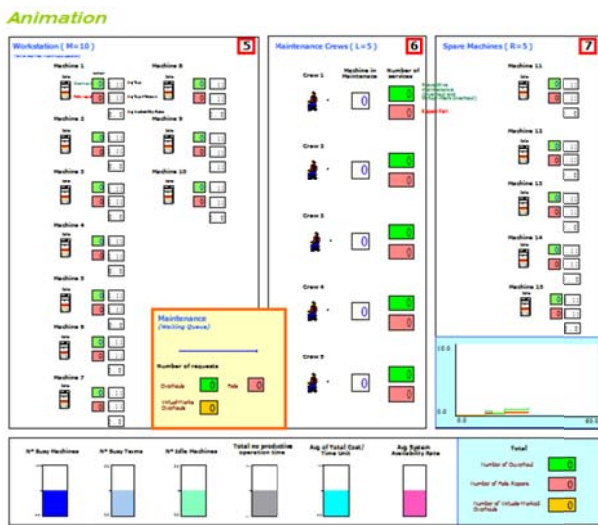


Fig. 4 – Animation area sample screenshot

However, some other performance measures can be estimated, such as:

- c) Average number of missing machines at the workstation ( $AvgM_{eq}$ );
- d) Average number of machines in the maintenance waiting queue ( $AvgLq$ );
- e) Average waiting time in the maintenance waiting queue ( $AvgWt$ );
- f) Average operating cycle time ( $AvgD$ );
- g) Probability of existing 1 or more idle Machines ( $Prob_{im}$ );
- h) Probability of the system being fully active ( $Prob_s$ );

And, finally, the simulation model also computes some individual efficiency measures per machine or maintenance crew, i.e.:

- i) Utilization rate per machine;
- j) Utilization rate per maintenance crew;

- k) Number of overhauls and repair actions performed per maintenance crew;
- l) Average availability per machine.

This paper will focus on the two general performance measures mentioned above – system availability and total maintenance cost per time unit.

Simulation length was set to 9.000 hours (approximately one year) – warm-up period was set to 3.500 hours.

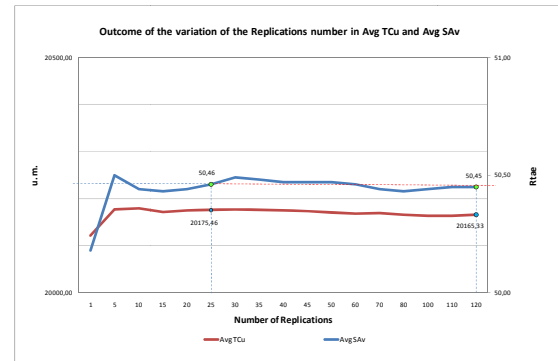


Fig. 5 – Outcome of the variation of the Replications number in  $AvgTCu$  and  $AvgSAv$  variables

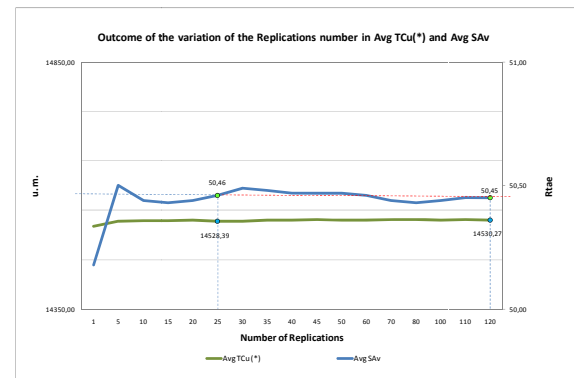


Fig. 6 – Outcome of the variation of the Replications number in  $AvgTCu(*)$  and  $AvgSAv$  variables

For each set of input parameters and pattern for variables, the simulation output variables  $AvgSAv$ ,  $AvgTCu$  and  $AvgTCu(*)$  were estimated based on 25 replications – for an adequate system stabilization and results robustness for both performance measures (Figures 5 and 6) and also due to computational time required to run the model.

#### 4. SIMULATION RESULTS AND DISCUSSION

Bearing in mind the twelve variables of MFS previously referred, simulation models were used to test and estimate the behaviour of the two global efficiency measures mentioned on the previous section. Simulation models were carried out for (1-60) hypothetical scenarios with different review rates ( $Arev$ ). These different review rates are associated with different times between review ( $T$ ) which are defined



accordingly to the preventive maintenance policy aiming the best option.

**Table 1-** Global efficiency measures outcomes in the MFS model after 25 replication

(Values estimated by simulation after 25 replication)

Scenario	$\lambda_{rev}$ (/hour)	$T$ (hour)	AvgSAv (%)	AvgTCu (m.u./hour)	AvgTCu(*) (m.u./hour)
1	0,10	10,000	29,25	21064,60	17304,37
2	0,20	5,000	34,75	21127,96	16646,80
3	0,30	3,333	40,79	21065,24	15870,35
4	0,40	2,500	45,26	20915,39	15287,79
5	0,50	2,000	48,12	20751,46	14908,68
...	...	...	...	...	...
52	9,00	0,111	48,71	20195,00	14776,87
53	15,00	0,067	48,52	20212,24	14807,21
54	20,00	0,050	48,54	20228,63	14815,57
55	25,00	0,040	48,56	20223,63	14814,18
56	30,00	0,033	48,54	20230,24	14816,98
57	35,00	0,029	48,62	20226,82	14811,16
58	40,00	0,025	48,54	20229,16	14820,31
59	45,00	0,022	48,50	20228,80	14822,78
60	50,00	0,020	48,52	20226,42	14815,55

(\*) Considers that the cost of lost production changes in function of the number of active machines lacking in the system.

**Table 2 –** Observe percentual change in the global efficiency measures after 25 replication

MPO -Observe percentual change			
Scenario	AvgSAv	AvgTCu	AvgTCu(*)
1-2	15,83%	0,30%	-3,95%
2-3	14,80%	-0,30%	-4,89%
3-4	9,87%	-0,72%	-3,81%
4-5	5,95%	-0,79%	-2,54%
...	...	...	...
52-53	-0,14%	0,06%	0,10%
...	...	...	...
58-59	-0,07%	0,00%	0,02%
59-60	0,04%	-0,01%	-0,05%

Max.	15,83%	-0,79%	-4,89%
Mean	0,80%	-0,07%	-0,27%

A first global analysis of the values presented in tables 1 and 2 indicate that the precision obtained on the three efficient measures analysed is different. An individual analysis of each measure indicates that AvgTCu shows the smaller variation (MPO lower). In Table 1 it can also be observed that when  $T$  takes very small values ( $T \leq 0,111$  or  $\lambda_{rev} \geq 9$ ) the three efficient measures [AvgTCu, AvgTCu(\*) and AvgSAv] are kept practically unchangeable. This fact can be confirmed in Fig. 7 or in Table 2 where the MPO for these values of  $T$  is extremely low, almost zero. On the other hand, when  $T$  assumes very high values ( $T \geq 2,5$  or  $\lambda_{rev} \leq 0,4$ ) the efficiency measures AvgTCu(\*) and AvgSAv present high MPO values in opposition to AvgTCu that shows very small values. In Table 2 it

can also be observed that AvgTCu presents the lowest MPO average value of the three efficiency measures and that AvgSAv has the highest value.

In order to simplify the interpretation and analysis of these global efficiency measures, figures 7, 8 and 9 pinpoint the maximum and minimum values (table 2 and 3) as well as other points considered relevant for the analysis.

**Table 3 -** Maximum values of the main efficiency measures

	Statistics (Maximum)	$\lambda_{rev}$ (/hour)	$T$ (hour)
AvgSAv (%)	50,70%	0,90	1,111
AvgTCu (m.u./hour)	21127,96	0,20	5,000
AvgTCu (*) (m.u./hour)	17304,37	0,10	10,000

Note: Red points in the graphics

**Table 4 -** Minimum values of the main efficiency measures

	Statistics (Minimum)	$\lambda_{rev}$ (/hour)	$T$ (hour)
AvgSAv (%)	29,25%	0,10	10,000
AvgTCu (m.u./hour)	20096,90	1,80	0,556
AvgTCu (*) (m.u./hour)	14518,77	1,20	0,833

Note: Yellow points in the graphics

Tables 3 and 4 show that the  $T$  value corresponding to the minimum value of AvgSAv corresponds the maximum value, as expected, of AvgTCu(\*). When compared with the minimum of AvgTCu, there is a significant  $T$  gap ( $\approx 5$  hours), although, its remains practically the same when the value of  $T$  changes from 5 to 10 hours (Fig. 9).

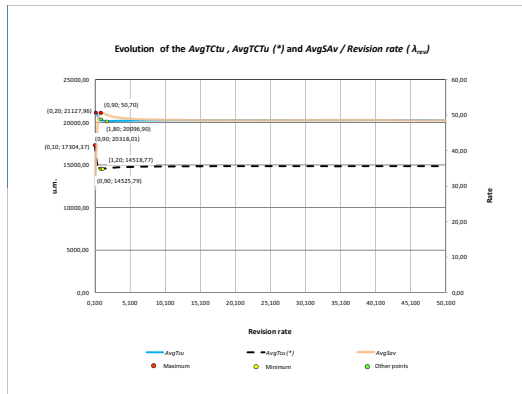
When comparing the  $T$  value corresponding to the maximum value of AvgSAv with the  $T$  value corresponding to the minimum value of AvgTCu(\*), there is only a small gap, which is clearly higher in the case of the  $T$  value corresponding to the minimum of the AvgTCu (Fig. 9)

**Table 5-** Correlation coefficients

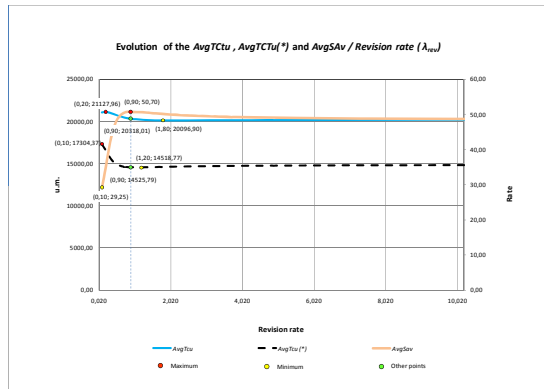
	$T$	AvgSAv	AvgTCu	AvgTCu(*)
$T$	1	-0,9021	0,8279	0,9017
AvgSAv	-0,9021	1	-0,7980	-0,9986
AvgTCu	0,8279	-0,7980	1	0,8237
AvgTCu (*)	0,9017	-0,9986	0,8237	1

A careful analysis of the correlation coefficients of three efficiency measures, table 5, shows that  $T$  variations are better explained by AvgSAv and AvgTCu(\*) ( $\approx 90\%$ ). It is also verified that there is a high inverse correlation between AvgSAv and

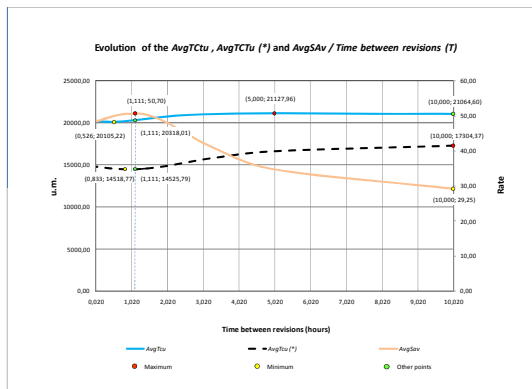
$AvgTCu(*)$  ( $\approx 99,8\%$ ). However when  $AvgTCu$  is compared with  $AvgTCu(*)$  or with  $AvgSav$ , the correlation coefficient decreases to 82,37% and to 79,80%, respectively. This partially explains why in tables 3 and 4 the  $T$  value corresponding to the maximum of  $AvgSav$  does not correspond exactly to the  $T$  value corresponding to the minimum of the  $AvgTCu$  and  $AvgTCu(*)$  and that difference being higher in the case of  $AvgTCu(*)$  (Fig. 9).



**Fig. 7 - Evolution of the  $AvgTCtu$ ,  $AvgTCTu(*)$  and  $AvgSav$  / Revision rate ( $\lambda_{rev}$ )**



**Fig. 8. Evolution of the  $AvgTCtu$ ,  $AvgTCTu(*)$  and  $AvgSav$  / Revision rate ( $\lambda_{rev}$ ) [Zoom Fig.7]**



**Fig. 9- Evolution of the  $AvgTCtu$ ,  $AvgTCTu(*)$  and  $AvgSav$  / Time between revisions ( $T$ )**

As it can be observed in Fig.7 and more clearly in figures 8 and 9, for the MFS analysed, the three global measures of efficiency being studied only present small variations for values of  $\lambda_{rev}$  between 0,10 and 9,00 (or  $T$  between 10,000 and 0,111 hours). For values of  $\lambda_{rev}$  higher than 9.00 the three global measures of efficiency remain practically unchanged.

**Table 6 - Comparison among the  $AvgTCu$  values estimates by the simulation model and analytic model (Lopes, 2007)**

$T$	Model	$AvgTCu$ (m.u./hour)	$AvgTCu(*)$ (m.u./hour)	$A_1$ (%)	$A_2$ (%)	$A_3$ (%)
1,66	Simulation	20381,74	14554,5	-40%	-12%	-25%
	Analytic	18130,56	----	----		
3,33	Simulation	20126,65	14555,62	-38%	-19%	-17%
	Analytic	16968,39	----	----		
1,66	Simulation	20111,24	14628,02	-37%	-16%	-18%
	Analytic	17303,65	----	----		
3,33	Simulation	21065,24	15870,35	-33%	-16%	-14%
	Analytic	18167,34	----	----		
mean				-37%	-16%	-18%

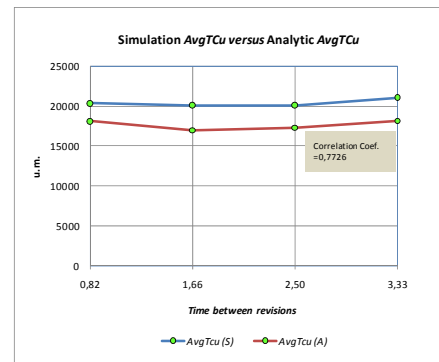
Note:  $M=10$ ;  $R=5$ ;  $L=5$ .

(\*) Considers that the cost of lost production changes in function of the number of active machines lacking in the system.

$A_1$  - Difference among Simulation  $AvgTCu$  and Simulation  $AvgTCu(*)$

$A_2$  - Difference among Simulation  $AvgTCu$  and Analytic  $AvgTCu$

$A_3$  - Difference among Analytic  $AvgTCu$  and Simulation  $AvgTCu(*)$



**Fig. 10 - Comparison among the  $AvgTCu$  values estimates by the simulation model and by the analytic model (Lopes, 2007)**

In Table 6 there is a comparison between the values obtained from the simulation model developed by the authors in a former (Peito et al 2011) and the analytical model developed by (Lopes 2007). The sample size of the results presented and compared in this case was limited by the number of results presented by the author in her work (Lopes, 2007). In this table it can be verified that when the two global efficiency measures are both estimated from the simulation model the difference ( $A_1$ ) is on average -37%, presenting  $AvgTCu$  always higher values. However when  $AvgTCu$  is estimated through the analytical model that difference ( $A_3$ ) is on average -18%. When the same efficiency global measure based on the analytical model is compared with the one calculated based on the simulation model,  $AvgTCu$ , this

if calculated from the analytical model presents lower values, on average, of 16%. It is also observed that the analytical model always presents for its efficiency measure values that lie between the two efficiency measures estimated from the simulation model.

Finally, through Fig. 10 it can be verified that the behaviour of *AvgTCu* is identical in both models. However this results analysis lacks confirmation due to small sample size dimension.

## 5. CONCLUSIONS

This paper shows similar estimated values for the performance measures analysed – system availability and total maintenance cost per time unit, for both simulation model and analytical model, as far as a *Maintenance Float System* with  $M=10$ ,  $R=5$  and  $L=5$  is concerned. Nevertheless a difference of nearly 15% is noticed. Also, it is quite clear that variance is different for both global efficiency measures analysed, specially when using extreme values for periodic overhauls rates. In this respect, *AvgSav* is the most sensitive parameter. As expected, the least sensitive parameter is *AvgTCu*, as it does not take into consideration the number of available machines, i.e., the cost for production loss is constant, irrespective of the number of available machines in the system.

## 6. FUTURE DEVELOPMENTS

The simulation model here presented, incorporating analysis of usual performance measures, also drives its concern towards new efficiency measures, enabling new trends for the analysis and discussion of the best decisions as far as a specific *Maintenance Float System* is concerned. Nevertheless the authors are now aiming to the development of an advanced simulation model, incorporating flexibility. This target would be reached by developing and incorporating new modules in our simulation tool, following past experiences found on literature (Luis S Dias, 2005, 2006 and Vilks, P., 2009, 2010) where the automatic generation of simulation programs enables desired model flexibility, i.e., making the model generating specific simulation programs for specific *Maintenance Float Systems*. These mentioned future developments also intend to potentiate the known capability of simulation to efficiently communicate with managers and decision makers, even if they are not simulation experts.

## REFERENCES

- Alam, Fasihul M.; McNaught, Ken R. and Ringrose, Trevor J. (2003), "Developing Simulation Metamodels for a Maintenance Float System", AMORG, Engineering Systems Department, Cranfield University, RMCS Shrivenham, Swindon SN6 8LA, UK
- Chen, M. C. and Tseng, H. Y. (2003), "An approach to design of maintenance float systems," *Integrated Manufacturing Systems*, vol. 14, pp. 458-467.
- Dias, Luis M. S.; Rodrigues, A. J. M. Guimarães; Pereira, Guilherme A. B. (2005), "An Activity Oriented Visual Modelling Language with Automatic Translation to Different Paradigms, Proceedings of the 19th European Conference On Modelling And Simulation (ECMS 2005), Riga, Letónia. Ed. Yury Mercuriev et al. Junho de 2005. pp. 452-461
- Dias, Luis S.; Pereira, Guilherme A. B. and Rodrigues, A. J. M. Guimarães (2006), "A Shortlist of the Most 'Popular' Discrete Simulation Tools", ASIM 2006 - 19th Symposium on Simulation Technique. SCS Publishing House. Ed. M. Becker and H. Szczerbicka. Hanover, Alemanha. pp. 159-163.
- Dias, Luis S.; Pereira, Guilherme A. B. and Rodrigues, A. J. M. Guimarães (2006), "Activity based modeling with automatic prototype generation of process based arena models", EMSS 2006 - 2nd European Modeling and Simulation Symposium. Barcelona, Espanha. pp. 287-296
- Gupta, V. and Rao, T. (1996), "On the M/G/1 machine interference model with spares," *European Journal of Operational Research*, vol. 89, pp. 164-171.
- Gupta, S. M. (1997), "Machine interference problem with warm spares, server vacations and exhaustive service," *Performance Evaluation*, vol. 29, pp. 195-211.
- Ingalls, Ricki G. (2001), "Introduction to Simulation", in *Proceedings of 2001 Winter Simulation Conference*, B.A. Peters, J.S. Smith, D.J. Medeiros, and M. W. Rohrer, eds.
- Kelton, W. David; Sadowski, Randall P. and Strurrok, David T. (2004), "Simulation With Arena", (3rd edition), McGraw-Hill, (1998-2004).
- Kuei, C. H. and Madu, C. N. (1994), "Polynomial metamodelling and Taguchi designs in simulation with application to the maintenance float system," *European Journal of Operational Research*, vol. 72, pp. 364-375.
- Lopes, Isabel S. (2007), "Técnicas Quantitativas no Apoio à Decisão em Sistemas de Manutenção", Tese de Doutoramento, Universidade do Minho.
- Lopes, Isabel S.; Leitão, Armando L. F. and Pereira, Guilherme A. B. (2005), "Modelo de Custos de Manutenção para um Sistema com M Unidades Idênticas", in C. G. Soares, A. P. Teixeira e P. Antão (eds), *Análise e Gestão de Riscos, Segurança e Fiabilidade 2*: p 603-620, Lisboa: Edições Salamandra (ISBN 972-689-230-9).
- Lopes, Isabel S.; Leitão, Armando L. F. and Pereira, Guilherme A. B. (2006), "A Maintenance Float System with Periodic Overhauls", in Guedes Soares & Zio (eds), *Safety and Reliability for Managing Risk 1*: p 613-618, London: Taylor & Francis Group (ISBN 0-415-41620-5).
- Madu, C. N. and Kuei, C. H. (1992a), "Group screening and Taguchi design in the optimization

- of multi-echelon maintenance float simulation metamodels," *Computers and Operations Research* vol. 19, pp. 95-105.
- Madu, C. N. and Kuei, C. H. (1992b), "Simulation metamodels of system availability and optimum spare and repair units," *IIE Transactions*, vol. 24, pp. 99-104.
- Madu, C. N. and Lyau, P. (1994), "On the use of simulation metamodeling in solving system availability problems," *Microelectronics Reliability*, vol. 34, pp. 1147-1160.
- Madu, I. E. (1999), "Robust regression metamodel for a maintenance float policy," *International Journal of Quality & Reliability Management*, vol. 16, pp. 433-456.
- Mamede, Nuno (1984), "Simulação Digital de Processos", Tese de Mestrado em Engenharia Electrotécnica e de Computadores (Telecomunicações e Computadores), Instituto Superior Técnico.
- Pegden, C.D.; Shannon, R. E. and Sadowski, R. P. (1990) "Introduction to Simulation Using SIMAN, McGraw-Hill", New York, USA. v. 2. 1990.
- Pidd, Michael (1989), "Computer Modelling for Discrete Simulation", (Editor), Wiley.
- Pidd, Michael (1993), "Computer Simulation in Management Science", Third Edition, Wiley.
- Peito, Francisco; Pereira, Guilherme; Leitão, Armando; Dias, Luís (2011), "Simulation as a Decision Support Tool in Maintenance Float Systems", 17<sup>th</sup> European Concurrent Engineering Conference (ECEC'2011). London, England. pp.68-74.
- Rodrigues, Guimarães and Carvalho, Valério (1984), "CAPS - ECSL, Experiência de modelagem e simulação aplicada a um sistema de elevadores", Relatório Técnico, Universidade do Minho.
- Rubinstein, Reuven Y. and Melamed, Benjamin (1998), "Modern Simulation and Modeling", Wiley Series in Probability and Statistics, Applied Probability and Statistics Section, A Wiley-Interscience Publication, John Wiley & Sons, INC., ISBN 0- 471-17077-1.
- Shannon, Robert E. (1998), "Introduction to the Art and Science of Simulation", in *Proceedings of 1998 Winter Simulation Conference*, D.J.Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.
- Shankar, G. and Sahani, V. (2003), "Reliability analysis of a maintenance network with repair and preventive maintenance," *International Journal of Quality & Reliability Management*, vol. 20, pp. 268-280.
- Vik, P.; Pereira, G. and Dias, L. (2009), "Software Tools Integration for the Design of Manufacturing Systems", *Proceedings of the Industrial Simulation Conference*, Loughborough, eds. Diganta Bhusan Das, Vahid Nassehi and Lipika Dekka, pp 127-131
- Vik, P.; Dias, L. and Pereira, G. (2010), "Automatic Generation of Computer Models Through the Integration of Production Systems Design Software Tools". *ASMDO2010 - Third International Conference on Multidisciplinary Design Optimization and Applications*. Paris (France)
- Vik, P.; Dias, L.; Pereira, G. and Oliveira, J. (2010), "Improving Production and Internal Logistics Systems – An Integrated Approach Using CAD and Simulation". *ILS2010 - 3rd International Conference on Information Systems, Logistics and Supply Chain - Creating value through green supply chains*. Casablanca (Morocco), April 14-16.
- Zeng, A. Z. and Zhang, T. (1997), "A queuing model for designing an optimal three-dimensional maintenance float system", *Computers & Operations Research*, vol. 24, pp. 85-95.

## AUTHOR BIOGRAPHY

**FRANCISCO PEITO** was born in 1966 in Macedo de Cavaleiros, Portugal. He graduated in Mechanical Engineering-Management of Production in Polytechnic Institute of Porto. He holds an MSc degree Industrial Maintenance in University of Porto. His main research interests are Simulation and Maintenance.

**GUILHERME PEREIRA** was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management at 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.

**ARMANDO LEITÃO** was born in 1958 in Porto, Portugal. He graduated in Mechanical Engineering at University of Porto, Portugal. He holds an MSc degree in Production Engineering and a PhD degree in Production Engineering from the University of Birmingham, UK. His main research interests are Reliability and Quality Maintenance.

**LUÍS DIAS** was born in 1970 in Vila Nova de Foz Côa, Portugal. He graduated in Computer Science and Systems Engineering at 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, Operational Research and Systems Visual Modeling.

# ANP APPROACH FOR IMPROVING PUBLIC PARTICIPATION IN STRATEGIC ENVIRONMENTAL MANAGEMENT PLANNING

Fabio De Felice<sup>(a)</sup>, Antonella Petrillo<sup>(b)</sup>

<sup>(a)</sup> D.i.M.S.A.T - Department of Mechanism, Structures and Environment - University of Cassino, Italy

<sup>(b)</sup> D.i.M.S.A.T - Department of Mechanism, Structures and Environment - University of Cassino, Italy

<sup>(a)</sup>defelice@unicas.it, <sup>(b)</sup>a.petrillo@unicas.it

## ABSTRACT

Environmental challenges decisions are often characterized by complexity, irreversibility and uncertainty. Much of the complexity arises from the multiple-use nature of goods and services, difficulty in monetary valuation of ecological services and the involvement of numerous stakeholders. From this point of view multicriteria techniques are considered as a promising framework to take into account conflictual, multidimensional, incommensurable and uncertain effects of decisions explicitly. Also principles and practice of public participation can serve to promote environmental equity for disadvantaged social groups.

Thus the objective of this paper is to propose a new methodological approach based on the Analytic Network process (ANP) and to examine the scope and feasibility of the ANP in incorporating participatory approach.

Keywords: ANP, Public participation; Multi-criteria decision-making, Environmental

## 1. INTRODUCTION

As we said, environmental challenges decisions are often characterized by complexity, irreversibility and uncertainty. Under these circumstances, conventional methods such as cost-benefit analysis are not adapt to evaluate environmental decisions (Ananda, 2003).

From this point of view multicriteria techniques are considered as a promising framework for evaluation since they have the potential to take into account conflictual, multidimensional, incommensurable and uncertain effects of decisions explicitly (Carbone et al., 2000; Munda, 2000; Omann, 2000). The most widely used multicriteria methods include the ANP – Analytic Network Process, multiattribute utility theory, outranking theory and goal programming, developed by LT. Saaty (Saaty, 1980). The ANP does not separate intangible from tangible factors. It is a useful tool to analyse decisions in complex social and political problems. The ANP is also useful when many interests are involved and a number of people participate in the judgement process (Saaty, 2005).

So this method can be useful in environmental challenges planning as it can accommodate conflictual, multidimensional, incommensurable and incomparable sets of objectives. On the other hand principles and

practice of public participation can serve to promote environmental equity for disadvantaged social groups.

The effectiveness of this practice in preventing or reducing environmental inequity depends upon the use of participation methodology which caters to the cultural and social needs of such groups. These methods need to provide appropriate forms of information, suitable venues for participation, and access to expertise and education which enable the public to understand policy issues and formulate preferences. The extent to which public preferences are incorporated in policy decisions determines the worth of public participation programs in promoting environmental equity (Hampton, 1999).

From this point of view we noted that some of the participatory methods developed so far have often been criticized as lacking efficacy because of poor rigor and need of better structuring and analytical capabilities.

In spite of this criticism, several studies applying the AHP (and its generalization ANP) to incorporate public participation have concluded that the AHP/ANP method is worth pursuing (Kangas, 1994, 1999; Ananda and Herath, 2003, Mau-Crimmins *et al.* 2005).

Thus the objective of this paper is to propose a new methodological approach based on the Analytic network process (ANP) and to examine the scope and feasibility of the ANP in incorporating participatory approach and stakeholder preferences into environmental challenges planning (De Felice *et al.* 2010).

To overcome environmental challenges a holistic approach and effective co-management strategies are required. Co-management means all stakeholders, and especially local communities, participate in management: analyzing, proposing actions, taking decisions, applying them, assessing their results, following up on the actions, giving feedback, etc.

Our project is based on the assumption that the barriers to effective decision-making that exist between local communities and other stakeholders cannot be broken down by one party acting alone. We propose to avoid modeling the conflicts of interests on the basis of a win-lose approach, or a top-down flux. The project will take the approach of discussing and proposing alternatives and choosing the most appropriate one from various points of view. The focus is more on active participation and debate, and less on who is in command or whose priorities and preferences are more important.

## 2. THEORETICAL BACKGROUND OF ANP MODEL

The Analytic Network Process is the generalization of AHP that is Analytic Hierarchy Process. AHP was developed by Thomas Saaty (Saaty, 1980) in the early 1970s. The strength of the AHP approach lies in its ability to structure a complex, multiattribute, multiperson and multiperiod problem hierarchically.

In addition, it can also handle both qualitative (through representing qualitative attributes in terms of quantitative values) and quantitative attributes. The general approach followed in AHP is to decompose the problem and to make pairwise comparisons of all the elements (attributes, alternatives) at a given level with respect to the related elements in the level just above.

AHP usually involves three stages of problem solving: the principles of decomposition, comparative judgments, and synthesis of priority.

In particular, details on the Analytic Network Process (ANP) model can be found in Saaty (1999); the fundamentals are summarized here for completeness. The ANP model consists of the control hierarchies, clusters, elements, interrelationship between elements, and interrelationship between clusters. The modeling process can be divided into four steps for the ease of understanding which are described as follows:

- **Step I:** pairwise comparison and relative weight estimation. The determination of relative weights in ANP is based on the pairwise comparison as in the standard AHP. Pairwise comparisons of the elements in each level are conducted with respect to their relative importance towards their control criterion based on the principle of AHP. Saaty suggested a scale of 1–9 when comparing two components. The score of  $a_{ij}$  in the pairwise comparison matrix represents the relative importance of the component in row (i) over the component in column (j), i.e.,  $a_{ij} = w_i / w_j$ . The score of 1 represents equal importance of two components and 9 represents extreme importance of the component i over the component j. The reciprocal value of the expression (1 /  $a_{ij}$ ) is used when the component j is more important than the component i. If there are n components to be compared, the matrix A, is defined as:

$$\begin{array}{c|cccc}
 & A_1 & A_2 & \dots & A_n \\
 \hline
 A_1 & w_1/w_1 & w_1/w_2 & \dots & w_1/w_n \\
 A_2 & w_2/w_1 & w_2/w_2 & \dots & w_2/w_n \\
 \dots & \dots & \dots & \dots & \dots \\
 A_n & w_n/w_1 & w_n/w_2 & \dots & w_n/w_n
 \end{array}$$

$$A = \begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{bmatrix}$$

After all pairwise comparison is completed the priority weight vector (w) is computed as the unique solution of:  $Aw = \lambda_{\max} w$  where  $\lambda_{\max}$  is the largest eigenvalue of matrix A.

The consistency index (CI) of the derived weights could then be calculated by Equation (1):

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (1)$$

In general, if CI is less than 0.10, satisfaction of judgments may be derived.

- **Step II:** formation of initial supermatrix. Elements in ANP are the entities in the system that interact with each other. The determination of relative weights mentioned above is based on pairwise comparison as in standard AHP. The weights are then put into the supermatrix that represents the interrelationships of elements in the system. The general form of the supermatrix is described in Figure 1 where CN denotes the Nth cluster, eNn denotes the nth element in the Nth cluster, and  $W_{ij}$  is a block matrix consisting of priority weight vectors (w) of the influence of the elements in the ith cluster with respect to the jth cluster.



		C <sub>1</sub>				C <sub>2</sub>				...	C <sub>N</sub>				
		e <sub>11</sub>	e <sub>12</sub>	...	e <sub>1n<sub>1</sub></sub>	e <sub>21</sub>	e <sub>22</sub>	...	e <sub>2n<sub>2</sub></sub>		e <sub>N1</sub>	e <sub>N2</sub>	...	e <sub>Nn<sub>N</sub></sub>	
C <sub>1</sub>	e <sub>11</sub>	W <sub>11</sub>				W <sub>12</sub>				...	W <sub>1N</sub>				
	e <sub>12</sub>														
	...														
	e <sub>1n<sub>1</sub></sub>														
C <sub>2</sub>	e <sub>21</sub>	W <sub>21</sub>				W <sub>22</sub>				...	W <sub>2N</sub>				
	e <sub>22</sub>														
	...														
	e <sub>2n<sub>2</sub></sub>														
⋮		...				...				...		...			
C <sub>N</sub>	e <sub>N1</sub>	W <sub>N1</sub>				W <sub>N2</sub>				...	W <sub>NN</sub>				
	e <sub>N2</sub>														
	...														
	e <sub>Nn<sub>N</sub></sub>														

Figure 1: Supermatrix form

- **Step III:** formation of weighted supermatrix. The initial supermatrix consists of several eigenvectors each of which sums to one. The initial supermatrix must be transformed to a matrix in which each of its columns sums to unity.
- **Step IV:** calculation of global priority vectors and weights. In the final step, the weighted supermatrix is raised to limiting power to get the global priority vectors as in Equation (2):

$$\lim_{n \rightarrow \infty} (W)^n \quad (2)$$

Once the matrix of pairwise comparisons has been developed, one can estimate the relative priority for each of the alternatives in terms of the specific criteria. Preferences derived from a criteria or sub criteria matrix are used to calculate a composite weight for each alternative. This part of ANP is referred to as synthesis. This enables ANP to obtain not only the rank order of the alternatives, but also their relative standings measured on a ratio scale. The alternative with the highest overall rating is usually chosen as a final solution.

### 3. ANP APPROACH TO IMPROVE PUBLIC PARTECIPATION IN ENVIRONMENTAL MANAGEMENT

The main objective of our work is to develop a participatory decision-making model which will be used when dealing with key environmental decisions together with local communities and other important stakeholders. To achieve this participatory decision-making model, the following objectives are envisaged:

1. Balance the starting knowledge level of all partners.
2. Analyze and parameterize conflicts of interest in natural resource management.
3. Identify the reference cases.
4. Model the decision-making processes helping the local communities.

5. Model the participation processes.
6. Improve decision-making procedures.
7. Develop the proper support for participation, discussion, learning, evaluation, prioritization, communication, traceability, etc.
8. Improve the capability of local communities to become a partner when defining natural resource management policies.
9. Develop procedures for collective working on line.
10. Construct an Analytic Network Model to enhance participatory approaches.

To structure the decision problem we identified and structured objectives which required careful empirical and literature investigations (De Felice and Petrillo, 2010). They provide the basis for quantitative modeling. Keeney (1992) classifies objectives as fundamental objectives and means objectives. The fundamental objectives are the issues or attributes that stakeholders genuinely care about, and means objectives are ways to accomplish the fundamental objectives. Objective hierarchies can be constructed using this classification. For example, ecologically sustainable development could be the fundamental objective and economic, social and environmental objectives could be the means objectives in case of forest decisions. Attributes to measure these objectives too should be identified. In appendix (Figure 2) we illustrate research framework.

Definitely our model can be used to clarify public preferences in a more rigorous manner. The most natural controversies hinge on disagreements about values. Such disagreements are usually about the degree and not the kind. The model can highlight such value tradeoffs in a useful way. It can also help construct and evaluate options, providing credibility and transparency to the process apart from its educational value. In this context, the model can be effectively used to obtain second best solutions or compromise solutions.

#### 3.1. The research line

The problem is based on principles stated by the European Community (see Appendix in Figure 3 - Research Line Framework). In particular our aims are:

1. Identify where and when the decisions are made and identify managerial procedures;
2. Identify who makes decisions and how they are carried out;
3. Identify a model to manage the environmental challenges or problems:
  - Definition of procedure for modeling and solving problems;
  - Database of sustainable problems and their on-going solutions (Mining, Fishing, Forest Management, Tourism, Waste);
4. Develop a platform to:

- Facilitate/improve the decision procedure (E-Democracy, Participation of all stakeholders);
- Optimization use of natural resources;
- Improve environmental and climate changes;
- Optimization ecosystem services.

Platform should work on-line accessible, as for example a real “decision centre” (see Appendix in Figure 4- the ANP-IT Platform Framework).

### 3.2. The proposal

The aim of the research project consists in building-up an ANP theoretical framework and the IT internet-based platform allowing stakeholders to actively participate in the decision making process and facilitate the achievement of public consensus.

Stakeholders differ according to the nature of the problem. Principally, referring to classical site location problems of point/linear/areal facilities (e.g. waste-to-energy plants, landfills, production facilities, road/electric/pipelines infrastructures,...) they are people living in the surroundings of site options. Stakeholders also include technical experts and public administrators involved in decision making. The logical structure of a collective decision making will contribute to:

- Point out decision makers and procedures of decision processes;
- Increase public awareness of environmental/social/economic effects of alternatives;
- Increase the e-participation (e-Democracy) of people in the decision making process to achieve public awareness consensus;
- Spread environmental information;
- Facilitate discussions on the environmental matter.

### 3.3. Description of methodological approach phases

Here below is the description of methodological phases:

#### Phase 1. Problem and public institution identification.

Aim of this phase is the identification with the local community of the environmental problem.

#### Phase 2. Structuring decision making model.

The decision making process will be structured by ANP techniques; problem components as well as tangible/intangible decision variables will be defined and clustered. Relations among components will be defined as well as the definition of the scale of preferences. Problem structuring will be carried out by considering scientific literature as well as judgments of experts and public decision makers. Definitely the aim of the decision making model will be:

- The strengthening of the local and regional system planning;

- The definition of an integrated framework for sustainable development, climate and energy to promote research for prevention and environmental protection;
- The individualization of tools for mapping of regional and local actors and relations system;
- The availability of specific set of indicators;
- The integration of the evaluation and monitoring procedures;
- The promotion and support for consolidated planning tools.

The approach will be:

1. Multilevel;
2. Multisector;
3. Multiactor.

As we said the use of ANP involves several steps.

- Structuring the decision problem;
- Identifying management options;
- Identifying criteria;
- Identifying stakeholders;
- Weighting schemes.

We note that participatory tools such as In-depth Groups (De Marchi et al., 1998), Negotiation Forums (Eastman et al., 1998), Focus Groups (Keeney et al., 1990; McDaniels and Roessler, 1998), and Citizen’s Juries (Crosby, 1996) can be effectively employed to elicit the most important criteria for a particular forest region. Thus the main criteria could be historic, aesthetic, environmental conservation, recreation, economic, social and cultural and educational values (see Appendix in Figure 5 - a Simplified ANP decision model based on the representative participatory approach is shown).

#### Phase 3. Building ANP-IT Platform for Internet-based Collective Decision Making.

The model will be implemented by an Internet-based IT platform.

#### Phase 4. Offline Simulation of the ANP model.

Collective decision making implies preferences are defined as random variables. The IT platform will be able to simulate offline collective evaluation providing statistics of interest for decision making. Outputs (e.g. alternatives rankings and sensitivity analysis, integrated objectives for climate and energy protection, specific set of indicators, specific evaluation and monitoring environmental procedures) will be treated as random variables.

Use of ANP Model will assure:

- Take operating challenges into account;
- Define and prioritize criteria;
- Evaluate decision alternatives;
- Justify those decisions.

#### Phase 5. Developing a Full Scale Case Study.

The model and AHP IT platform will be tested on a full scale case study of interest for the public institution involved.

#### 4. CONCLUSIONS

Quantifying stakeholder preferences in environmental management is a complex task. From this point of view the methodologies of public participation can be judiciously selected and modified to promote equity. Participation can promote equity if the methodology adopted is culturally sensitive and accommodating.

The provision of information which can be readily processed by the public and education for enabling understanding of issues is critical for the promotion of equity through public participation.

The most critical aspect of promoting equity through participation is the extent to which public preferences are incorporated in policy decisions which govern environmental quality. Limited incorporation reduces participation programs to an inconsequential democratic drama.

On the other hand ANP allows for participation of more than one person as a decision maker, which is important in dealing with several stakeholder groups. Another advantage of the AHP is the ability to include many decision makers in an electronic meeting environment.

Therefore we decided to use the ANP in this study for the following reasons: (1) the ANP is a structured decision process quantitative process which can be documented and replicated, (2) it is applicable to decision situations involving multi-criteria, (3) it is applicable to decision situations involving subjective judgment, (4) it uses both qualitative and quantitative data, (5) it provides measures of consistency of preference, (6) there is an ample documentation of ANP applications in the academic literature, (7) the ANP is suitable for group decision-making.

The results of this study could provide valuable information regarding decision-making tools for strategic environmental management. The ANP could be applied to several problems in a wide range of fields, it has enjoyed relatively little use for natural resources planning. Definitely the ANP could successfully be applied to the experimental problem of wilderness area siting, thus offering potentials for actual application as a public participation tool.

#### REFERENCES

Ananda J., Herath, G., (2003). The use of Analytic Hierarchy Process to incorporate stakeholder preferences into regional forest planning. *Forest Policy and Economics* 5 (2003) 13–26.

Carbone, F., De Montis, A., De Toro, P., Stagl, S., 2000. MCDA methods comparison: environmental policy evaluation applied to a case study in Italy. *Third International Conference of the European Society for Ecological Economics*, Vienna, Austria, May 3–6, 2000.

Crosby, N., (1996). *Creating an authentic voice of the people. Deliberation on Democratic Theory and Practice*, Midwest Political Science Association, Chicago.

De Felice, F., Petrillo, A., (2010). A new multicriteria methodology based on Analytic Hierarchy Process: the “Expert” AHP *International Journal of Management Science and Engineering Management*. 5(6): 439–445, 2010.

De Felice, F., Petrillo, A., Falcone, D., (2010). Multi-criteria analysis for improving public participation in strategic environmental and social planning. *Multicriteria and multiagent decision making. Social and Economical Applications*, Naples, 26–27 November 2010.

De Marchi, B., Funtowicz, S., Gough, C., Guimaraes Pereira, A., Rota, E., (1998). The ULYSSES Voyage: The ULYSSES Project at the JRC. Report EUR 17760EN, European Commission, Joint Research Centre, Ispra, <http://yalba.jrc.it/ulysses.html>.

Eastman, R.J., Jiang, H., Toledana, J., (1998). Multi-criteria and multi-objective decision making for land allocation using GIS. In: Beinart, E., Nijkamp, P. (Eds.), *Multicriteria Analysis for Land-Use Management*. Kluwer Academic Publishers, Dordrecht.

Hampton, G., (1999). Environmental equity and public participation. *Policy Sciences* 32: 163–174, 1999.

Kangas, J., 1994. An approach to public participation in strategic forest management planning. *Forest Ecology Management* 70, 75–88.

Kangas, J., 1999. The Analytic Hierarchy Process (AHP): standard version, forestry application and advances. In: Helles, F., Holten-Anderson, P., Wichmann, L. (Eds.), *Multiple Use of Forests and Other Natural Resources*. Kluwer Academic Publishers, Amsterdam, pp. 96–105.

Keeney, R.L., Winterfeldt, D.V., Eppel, T., (1990). Eliciting public values for complex policy decisions. *Management Science* 36, 1011–1030.

Keeney, R.L., 1992. *Value-Focused Thinking: A Path to Creative Decision analysis*. Harvard University Press, Cambridge.

Munda, G., 2000. Conceptualising and responding to complexity. In: Spash, C., Carter, C. (Eds.), *Environmental Valuation in Europe*. Cambridge Research for the Environment, Cambridge, Policy research brief, No. 2.

Mau-Crimmins, T., de Steiguer, J.E., Dennis, D., (2005). AHP as a means for improving public participation: a pre-post experiment with university students. *Forest Policy and Economics* 7 (2005) 501–514.

McDaniels, T.L., Roessler, C., (1998). Multiattribute elicitation of wilderness preservation benefits: a constructive approach. *Ecological Economics* 27, 299–312.

Omann, I., 2000. How can multi-criteria decision analysis contribute to environmental policy

making? a case study on macro-sustainability in Germany. Third International Conference of the European Society for Ecological Economics, Vienna, Austria, May 3–6, 2000.

Saaty, T.L., 1980. The Analytic Hierarchy Process. third ed. McGraw-Hill, New York.

Saaty, T.L., 1999. Fundamentals of the analytic network process. Proc of International Symposium on Analytical Hierarchy Process, Kobe, Japan.

Saaty, T.L., (2005). Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks, RWS Publications, 4922 Ellsworth Ave., Pittsburgh, PA, 2005, p. 15213.

### **AUTHORS BIOGRAPHY**

**Fabio De Felice**, Researcher at the Faculty of Engineering of the University of Cassino, board member of several international organizations and responsible for scientific research and training in industrial plants. Holder of the course planning and management of industrial plants.

The scientific activity developed through studies and researches on problems concerning industrial plant engineering. Such activity ranges over all fields from improvement of quality in productive processes to the simulation of industrial plants, from support multicriteria techniques to decisions (Analytic Hierarchy Process, Analytic Network Process), to RAMS Analysis and Human Reliability Analysis. The main courses, both curricular and specialistic for territorial businesses in which he is involved are: Safety of Industrial Plants, Industrial Production Management, Industrial Simulation, Human Reliability Analysis. General Secretary of the Analytic Hierarchy Process – A.H.P. Academy - International Association for the promotion of multi-criteria decision making methods.

**Antonella Petrillo**, degree in Mechanical Engineering, now PhD at the Faculty of Engineering of University of Cassino where she conducts research activities on Multi-criteria decision analysis (MCDA) at the Department of Mechanism, Structures and Environment.

## APPENDIX

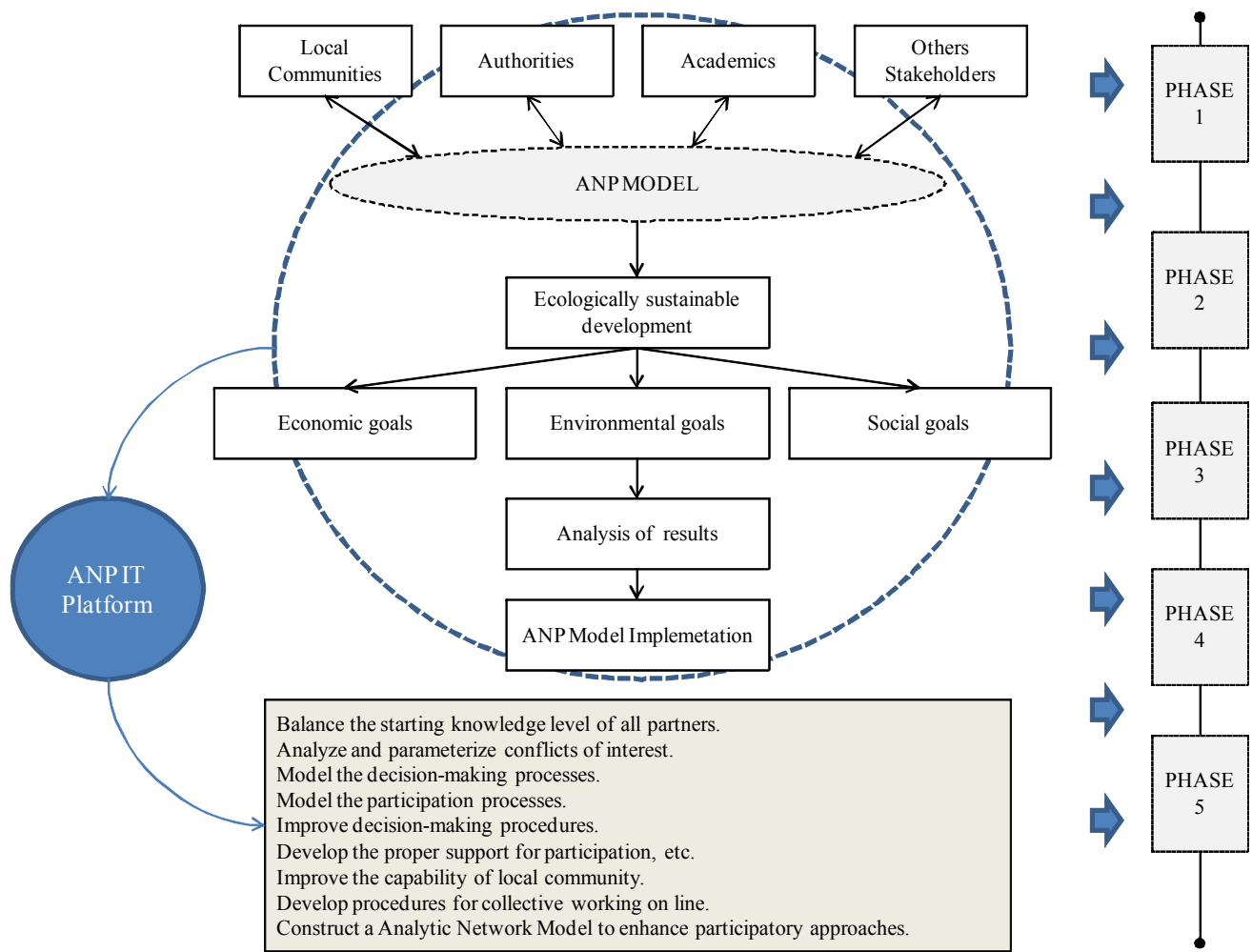


Figure 2: Research framework

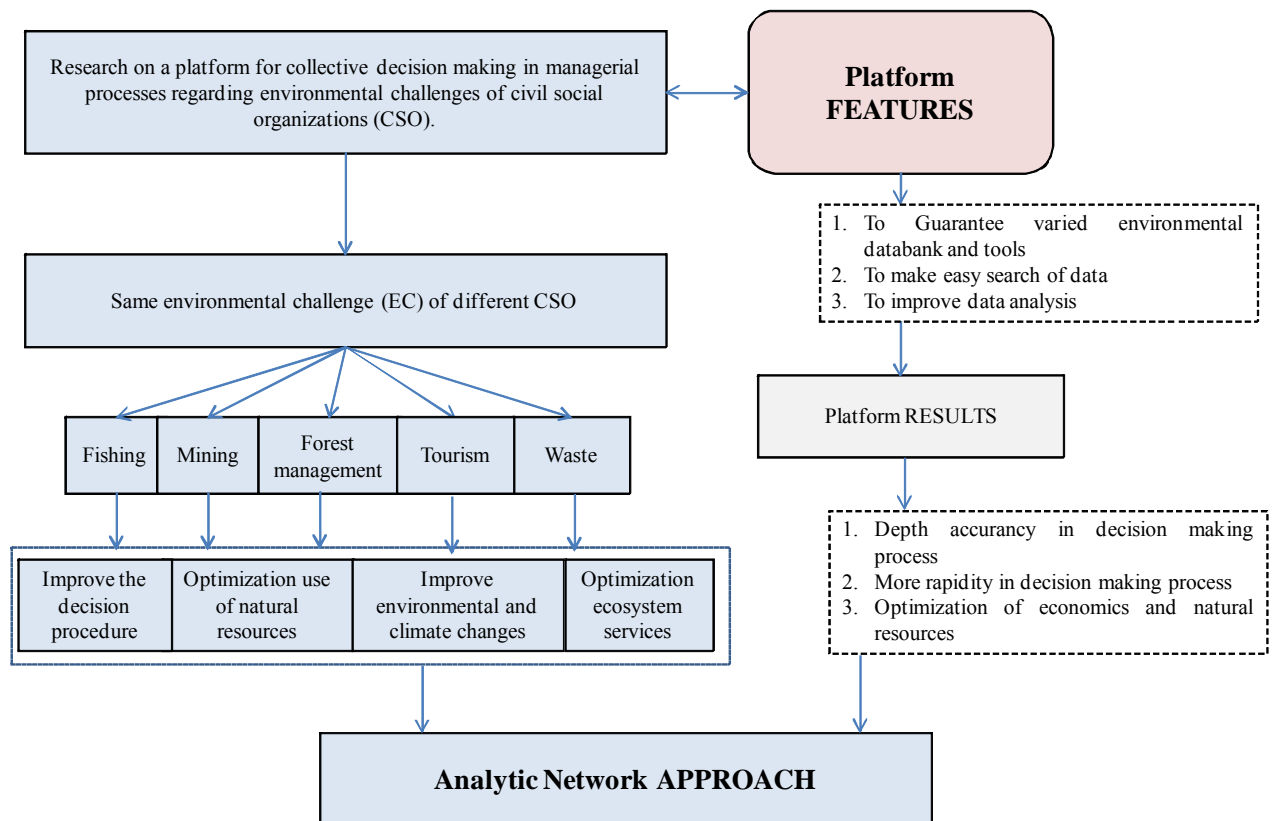


Figure 3. Research Line Framework

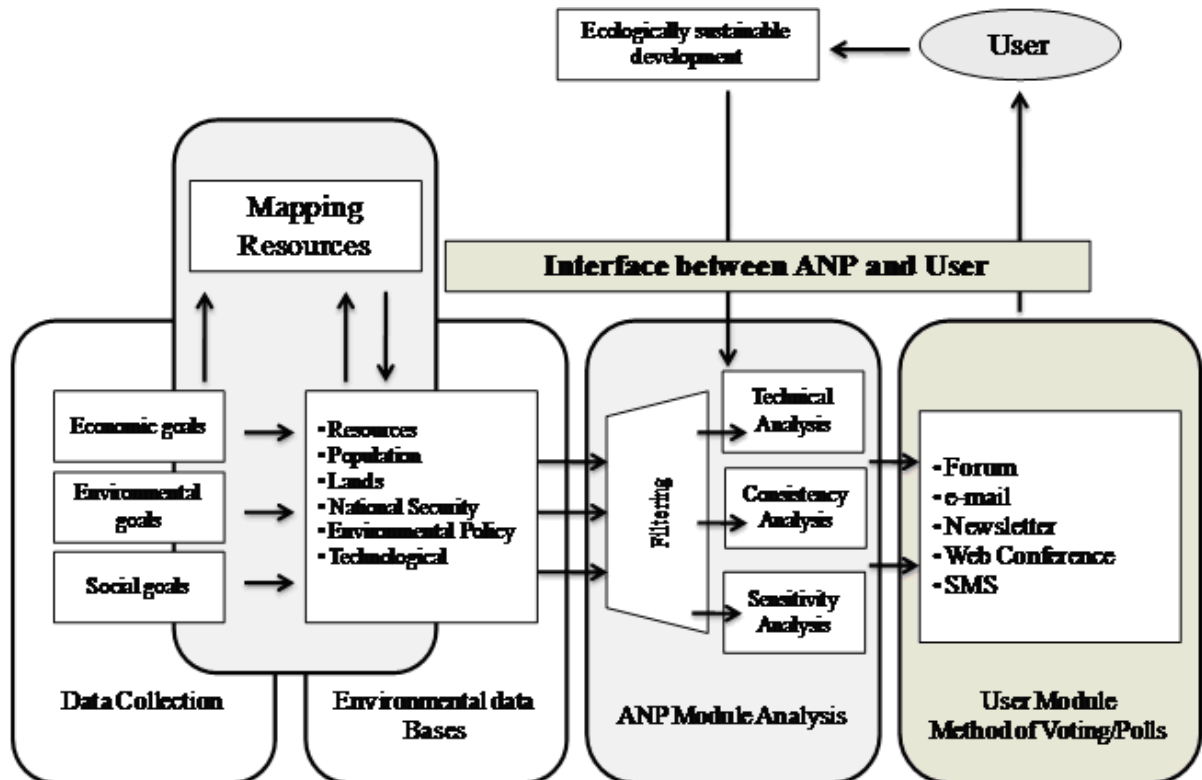


Figure 4. ANP-IT Platform Framework



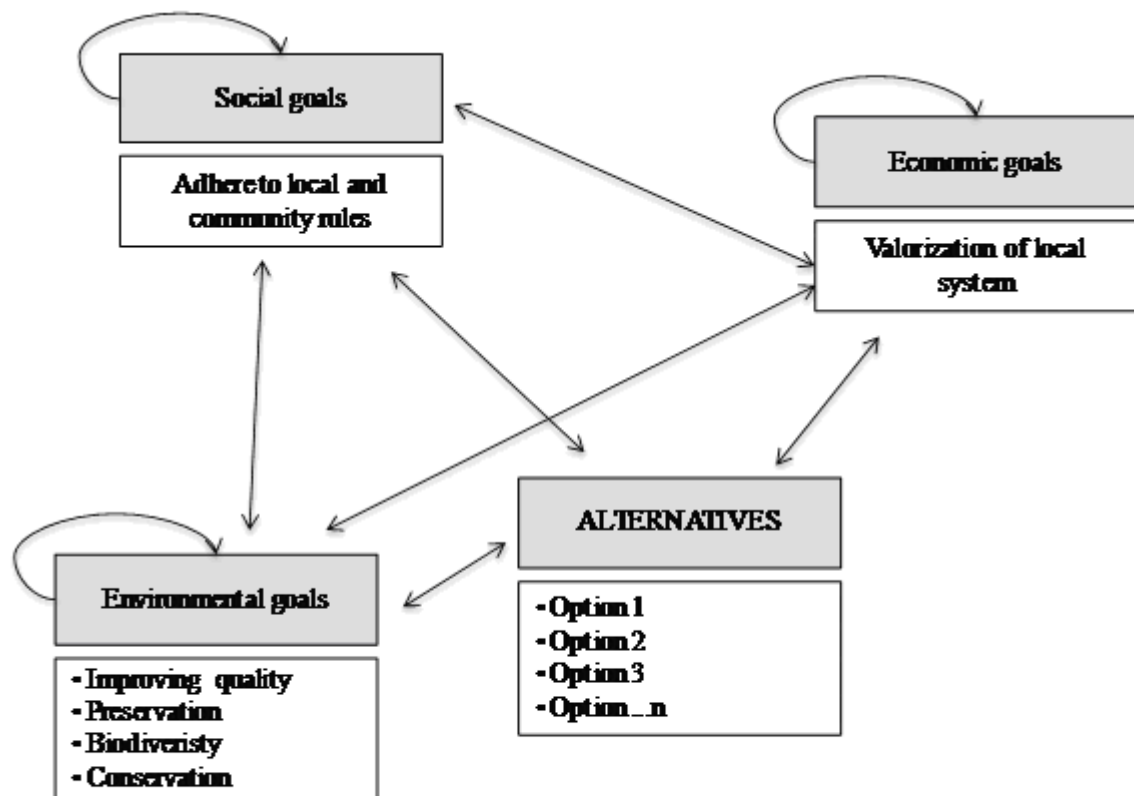


Figure 5. Simplified ANP decision model based on the representative participatory approach

# DISTRIBUTED M&S FOR PRODUCT SERVICE SYSTEM

Thecle Alix<sup>(a)</sup>, Gregory Zacharewicz<sup>(b)</sup>

<sup>(a), (b)</sup> Univ . Bordeaux, IMS, UMR 5218, F-33400 Talence, France.

<sup>(a)</sup> [theacle.alix@ims-bordeaux.fr](mailto:theacle.alix@ims-bordeaux.fr), <sup>(b)</sup> [gregory.zacharewicz@ims-bordeaux.fr](mailto:gregory.zacharewicz@ims-bordeaux.fr)

## ABSTRACT

A challenge for many companies is to propose product service systems to foster customer loyalty and gain a competitive advantage. The objective can be reached under the condition that systems are designed regarding their usage to avoid waste and to increase the system provider profitability. Furthermore, services must be developed jointly with products to benefit from synergies between them and to reduce resource consumption. Even if these conditions of success are known, the way to do is not formalized and guided by a clear methodology. This paper proposes a service model in toys industry to be simulated with different service scenario before development to validate and satisfy both provider and client. A service-based modeling environment for service specification and simulation under distributed G-DEVS/HLA is presented. The interoperability required for simulation viability is addressed by complying with the distributed simulation standard HLA.

Keywords: Product service system, simulation, G-DEVS, HLA federation

## 1. INTRODUCTION

Sustainable production and consumption is an issue of current international concern. Many different approaches and concepts have been developed over the last decades to address environmental problems, such as cleaner production, cleaner technologies, waste minimization and recycling approaches, eco-design and design for sustainability (Maxwell, Sheate and Van der Vorst, 2006). The expectation that these strategies would solve the problems has not been realized, mainly because of increasing levels of consumption stimulated by continuous population growth and rising levels of affluence. Economic growth is seen as a driver of development and the sign of a healthy economy by major economic models. It is sustained by increased consumer demand and is usually measured in material terms. But, there is an apparent conflict between the goals of economic growth and the goal of preserving a clean environment.

The “Product-Service System” (PSS) or servicing concept has been suggested as a way to reduce the conflict by exploring possibilities to sustain economic growth and consumer demand by creating more value without using more resources and causing more pollution. PSS challenge the traditional economy based

on the sale of tangible goods and shift to use selling. The optimization of the functions provided by a product to its user is part of a new business strategy whose transactions must:

- Involve the use of a capital asset rather than its sale,
- Expect positive effects on sustainable development by reducing consumption of material and/or energy.

Despite, some companies have already shifted their business and gained a competitive advantage clearly assessing a reduced environmental impact, others are feebleness to follow this innovative way of generating profit and addressing environmental burden. The main reason is that the new business strategy requires managerial and organizational changes. New business models and business plan are to be developed; a close relationship with the customers is required to clearly understand its needs and use of products; networks of partnership are to be settled. Moreover, it is important to understand how the system for delivery of use value can be organized to ensure its environmental soundness in comparison to traditional business models of selling products. Practically, it is unclear that new business is low resource consumption than traditional one and scenario simulation can be interesting to compare the two strategies.

Although the research on PSS is abundant (Tukker and Tischner, 2006; Meier, Roy and Seliger, 2010) no research propose clear guidance for companies that are planning to shift from selling products to providing use value and there is no model to guarantee a high value offer. For instance, the research lacks a clear and commonly recognized methodology to support PSS design and realization fulfilling the environmental soundness and expected profit.

The purpose of this paper is to contribute to PSS design. As such, the paper proposes systematical customer-oriented PSS development process that enables companies to develop offers use and environment oriented. The idea is to propose service models to be simulated in services scenario before implementation. The simulation will permit to validate the desired properties and anticipate undesired effects. In other words, the simulation results will provide decision makers with a business decision support to guide their actions and know whether the development strategy where they tend to, is win-win.

The paper is organized as follows. First, we review the PSS specificities and problematic. Then, a PSS business model is presented. Focusing on the design phase, we propose a service-based modeling environment with a graphical language for modeling service specification and running distributed DEVS simulation before concluding.

## 2. PSS INS AND OUTS

Many PSS definitions exist in the literature assessing the difficulty to have a non-ambiguous idea of the coverage of the term (Goedkoop, van Halen, Riele and Rommens1999).

### 2.1. PSS concept

The accepted definition was proposed by Mont (2001) who defines a product-service system as “a system of products, services, networks of actors and supporting infrastructure that continuously strives to be competitive, satisfy customer needs and have a lower environmental impact than traditional business models”.

The definition emphasizes both the relationship with the customer and the environmental impact insinuating the socio-economic character of the PSS innovation. The innovation can take three different shapes: product-oriented, use-oriented and result-oriented and has accordingly, a specific impact on the environment and on the customer satisfaction (Kanda and Kakagami, 2006)

Despite the two pillars, the research on PSS mainly rests on the innovative relationship between provider and customer. The concept is not used to policy development as the measure of the load of the lifecycle of product and service is complex. Therefore, it is possible to avoid waste and provide a sustainable solution by comparing on the one hand, the user-product utilization to the user need and, on the other hand, the provider offer value (based on quantitative and qualitative criteria) to the resource consumption. Higher considerations on the environment should integrate reflections on resource consumption and effect for human being on the long term. Finally, a PSS requires to clearly defining the offer architecture and associated resources, the company position during value creation and the revenue model. These elements are constituent of a business model.

### 2.2. PSS business model fundamental

The fundamental of a model business adapted to servicing would be the following ones:

- the value proposition which is offered to the market: product / service combination;
- the customer interface management:
  - The customers segments concerned by the value proposition: similar in a combination product / services, different in a formal logic of “serviticising” (Vandermerwe and Rada, on 1988);
  - The communication and the distribution channels used to reach the customers and offer them the value proposition: this mix

can strongly change since the offer leans on intangible elements.

- The management infrastructure: this element is central to the division of the structures responsible for the product manufacturing and service delivery:
  - The skills and the fundamental resources making the model business feasible;
  - The key activities necessary for the model business realization.
- The financial aspects:
  - The cash flows generated by the business model;
  - The costs of structure for the business model functioning;
  - The business model control methods and tools.

That view of the fundamentals of a PSS business model is aligned with Muller-Stewens and Lechner business model proposition. The authors defined a PSS Business Model composed of five sub-models (Muller-Stewens, and Lechner, 2005) in (Aurich, Mannweiler and Schweitzer, 2010). Despite objectives of each sub-models are described, the design and realisation phase of the PSS life cycle are given less attention. The next sub-section presents a general overview of a PSS life cycle-oriented business model.

## 3. PSS DEVELOPMENT METHODOLOGY

The proposed business model leans on a project process and thus follows the different phases of the project life cycle (Figure 1).

### 3.1. Global overview

In that specific case, the project aims at proposing a high value system composed of physical and non physical elements to a customer. As such, the design and realisation of the PSS are the main concerns. The project process is divided into four main phases: a definition, design, realization and closing phase. We do not specifically focus on the PSS use phase as it is another process, orthogonal to the project realization. Considerations on client-PSS use are obtained during: (i) the definition phase that stems from a marketing analysis and a customer-system utilization analysis and, (ii) the closing phase allowing a feedback on the solutions proposal (lacks and advantages). During each phase, operational and support activities are performed.

Operating activities are activities that make the PSS evolve from the status of "idea" to the status of system “delivered” and “evaluated” " by a client. These activities ensure that the functional requirements of the PSS are met according to the customer need, the business environment influence and the company itself.

Support activities allow the business process control. During these activities costs, time and risks are controlled to ensure the delivery on time of the solution and anticipate problems; knowledge and communication management enables to control the proposition value. Finally, the organization management tasks are equivalent to the management infrastructure above-

mentioned. Main aspects of the support activities concern are illustrated in Figure 1.

### 3.2. Detailed business model

Each phase of the business model is detailed hereafter.

#### 3.2.1. Definition phase

The first step of the methodology consists in a strategic analysis whose aim is to list the less risky and most profitable PSS innovations a company can propose. To make the list, the company will compare the customer needs, refined regarding product utilization, to the results of an internal and an external analysis. The internal analysis focuses on the company market position, organization, skills, knowledge, resources and ability to mobilize the resources. The external analysis is based on the macro-environment analysis (PESTEL model) and on the microenvironment analysis (Porter, 1985). Environmental incentives can be taken into account during that phase thanks to the PESTEL model.

A list of approximately forty criteria aggregated into ten factors is proposed to make that internal and external analysis. The quantification of the criteria allows their positioning into strengths or weaknesses, opportunities and threats according to their nature (internal or external, positive or negative). Two types of aggregators are considered: the "average" aggregator and the "weighted average" aggregator. The list of criteria and factors may be increased or decreased depending on their relevance to the case studied. It is possible to represent these quantified factors in a SWOT matrix chart and to compare, regarding the different charts, the PSS innovations against each other on a common basis in order to retain only the most advantageous. The different charts make it possible for a strategist to choose between several scenarios by analyzing the potential of

the company, the environment and the customer demand (Alix and Vallespir, 2010a).

#### 3.2.2. Design phase

The second step is an engineering phase that allows the determination of the PSS to design regarding the value of the system for any stakeholder. Currently, only the customer and the provider of the system are considered. Based on the value engineering approach, the determination rests on the distribution of the costs necessary to realize the PSS over the functions expressed as benefits expected for the provider and expressed as needs to satisfy for the customer (Goyhenetche, 1999).

Among the benefits are included : (i) the construction of a customer loyalty, (ii) the search for differentiation, (iii) the increase and stabilizing of firms' turnover, (iv) the corporate image reinforcement, (v) the occupation of an existing or new market, (vi) the possibility to create alliance with service providers and to share risks, (vii) the possibility to increase the quickness of a design or production process using product-service based on information and communication technologies, (viii) the possibility to shorten sales delay or negotiation phase using financial services and (ix) the search for a product-service system that is designed to have a lower environmental impact than traditional business models.

The needs of the customer include implicit and explicit needs get from a consumer panel. It also takes into account user-product utilization.

Costs are divided in direct and indirect costs. They encompass component costs, labour cost, and overheads. The value engineering analysis is done for both stakeholders and thus, two ratios are obtained. The first one represents the value of the PSS for the provider while the other represents the value of the PSS for the customer.

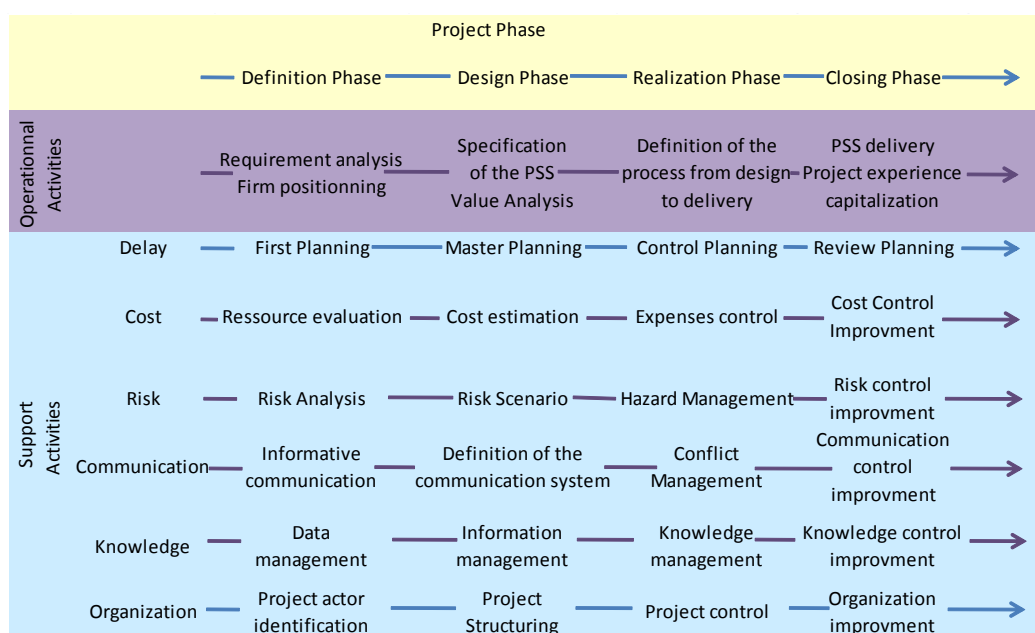


Figure 1: Business model of a PSS based on a project life cycle

The two values represented in a square matrix enables to analyze the relevance of a PSS innovation against the costs of design, implementation, delivery and mercantile strategy of the company before realization (Alix and Vallespir, 2010b).

The value analysis can be performed at a macroscopic level or at a more microscopic level considering that the system is made of elementary subsets whose value can be analyzed separately. The underlying idea is to simulate the process of coupling these components to see the combinations that increase the system value for both the customer and the provider. Modeling methods such BPMN, SADT/IDEF0 (Morelli, 2006), or blueprints (Chekitan and Schultz, 2005) can be used to represent the activities and their sequence. Work on validation of distributed simulation models coupling product-service are detailed in the following section.

### 3.2.3. Realization and closing phase

The realization phase aims to operationalize the solution defined in the previous phase. The production and delivery processes will be set up during the phase; the availability of material and human resources required will be checked as well as the competence of the human resources. The closing phase is the PSS assessment phase during which user-product utilization will be capitalized for future low consumption development. Possible subset coupling could merge from that feedback leading to the possibility for new development to benefit from past experience and thus minimizing expenses.

## 4. SERVICE MODELLING AND SIMULATION

Service design and development are more than ever a key challenge for enterprises. However, the development of new service activities is not fully formalized and is not guided by a clear or commonly recognized methodology. These facts may cause a drawback from the definitions to the practical implementation of services that can penalize the enterprise, especially when finding lately the defects in the services development. The idea presented in (Zacharewicz, Alix and Vallespir, 2009) was to model and simulate the services associated with the products before they real implementation in order to validate the desired properties and anticipate a wrong behavior of services. However, the authors identified a limit to this proposal: the complexity of a full product-service system (actors, software, machines) excluded from catching this whole complex behavior into the model. The idea is to offer physical or software actors to act in the loop simulation for validation. In consequence, the simulation requires interacting and synchronizing with heterogeneous actors and distributed service. To meet these considerations, the authors proposed a Service M&S environment based on a graphical language for modeling service selecting the essential concepts from (Bell, 2008) service models and reusing the G-DEVS/HLA distributed simulation (Zacharewicz, Frydman and Giambiasi, 2008). The G-DEVS formalism (Giambiasi, Escude and Ghosh, 2000) was chosen

for its formal properties. This previous works also described a method for transforming service specification models to simulation models. The interoperability of simulation models with human-machine interfaces was addressed by complying with the HLA standard for distributed simulation that got the thirty years of experience in the distributed simulation. Finally, the modeling environment of distributed services and interfaces proposed service models with other actors in a compatible HLA federation, propose to validate some desired properties of service-couples before their course into production (realization). These previous works on service modeling have been extended to cover PSS specification; they are detailed in the following sections.

### 4.1. Modeling

As mentioned earlier, the Product Service System Modeling is a recent domain that has not adopted yet a unique and/or common standard for developing frameworks to manage services processes. The complexity of PSS Modeling specification comes from PSS may involve different software and / or actors components which are essential to its execution but heterogeneously specified. The literature reports several standards specification, the authors have chosen the graphical definition of the Service-Oriented Modeling Framework (SOMF) (Bell, 2008). The reasons are mainly the fine coverage of the domain and the user friendly design of the language. In detail, the service modeling does not require expertise on service specification and language coding (execution code can be eventually generated from a graphical specification). Nevertheless, the missing piece is the correctness checking or the scenario evaluation of the couple product-service models. Concretely, this field is demanding verification and validation methods; the authors believe that simulation can give projecting information on the correctness of the models. In more detail, one identified lack is the SOMF does not define an explicit simulation semantics associated to the Service Modeling. This fact may lead to difficulties to determine if simulation result errors have to be imputed to a wrong model or to an incorrect implementation of the Service Modeling engine.

Having selected the essential concepts from SOMF in the elaboration of PSS Modeling, the authors have defined a PSS Modeling process to represent the components involved in PSS. The authors added to this description the notion of Resource that is key issue to represent the PSS. The Resource will be part of the model if possible, if not, thanks to interoperability (introduced next section), it will be involved in the loop human or external machine or software. To sum up the approach, a XML PSS Modeling process model is composed of sub service components that treat in-progress service and controllers components that route in-progress service information between sub services. In-progress services pass over a sequence of these components.

Then, the SOMF models are using a XSLT based method to transform XML PSS Modeling specifications

into G-DEVS simulation models. The work is based on a previously introduced methodology to transform Service models to Simulation models (Zacharewicz, Frydman and Giambiasi, 2008). In depth, the PSS model is transformed into a coupled G-DEVS model by coupling G-DEVS atomic models representing Service Modeling basic components. This G-DEVS model takes advantage of formal properties and can be simulated.

#### 4.2. Distributed Simulation HLA based

Based on the experience of distributed simulation, the authors proposed to address the interoperability by conforming to the distributed simulation standard High Level Architecture (HLA). For a short recall, HLA is a software architecture specification that defines a common understanding to create a global simulation composed of distributed simulations (or other software components). In HLA, every participating simulation is called federate. A federate interacts with other federates within an HLA federation, which is in fact a group of distributed federates. The HLA set of definitions brought about the creation of Standard 1.3 in 1996, which then evolved into HLA 1516 in 2000 and HLA 1516 Evolved in 2010 (IEEE, 2010).

Finally, they defined a distributed PSS Modeling Environment (detailed in § 5.3) that interfaces components (implementation of subsets) of the PSS Models and others actors in HLA compliant Federation.

### 5. PSS M&S CASE STUDY

In this study, the authors modeled a PSS oriented Service where service is more important than material property. The example study is dedicated to the baby toys market. The targets of this market are parents of child from 0 to 1 year. The authors proposed a G-DEVS coupled model of different market approaches (buying versus renting toys). This model has been formalized according to several data reports from CRIOC realized in Belgium (Vandercammen, Warrant and Meirsman, 2010). The study describes Service Models in a formal approach and validates these models by simulation.

#### 5.1. PSS Toys industry G-DEVS Models

This section details the PSS G-DEVS model for toy renting systems.

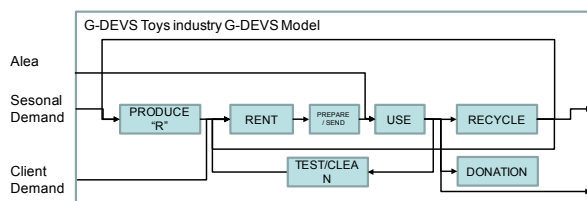


Figure 2: G-DEVS model PSS Service oriented

In figure 2, the G-DEVS coupled model is describing a PSS oriented service where the toy is rented instead of being sell. This model is also starting from a production phase that has been differentiated from a “classical” production on the fact it is supposed to integrate more environmental care in the choice of the raw materials, processes and recycling concerns. The second

sub model is the renting model it takes into account the demands from clients based on seasonal demands. Once the toy is rented it is prepared and expedited to the client. The use model of the toy is based on stochastic input on toys duration and child interest in the toy. In case the product is destroyed or damaged by the child, three solutions are envisaged. First, if too much damaged or destroyed the product is recycled as raw materials input for new production. Second, in case the product has been used more than a fixed amount of time or has been decided too much old-fashioned regarding criteria it is donated to charity organizations. Last, in case the toy is still functional and not degraded after a rent duration it is tested, cleaned and proposed again for renting.

#### 5.2. Other components

A random event generator has been created, based on Excel stochastic function to generate a file that schedules events to feed the toy breaking incidents, the toy interest for a child and old-fashioned. This file is charged in a scheduler for G-DEVS models in “alea” (hazard) input of the use model. The seasonal toy market selling and client demands have been elaborated in an Excel file to transpose the results from CRIOC to events planned as inputs of the PSS G-DEVS models (Client / Seasonal Demand input of Produce and Rent models).

#### 5.3. Distributed Simulation Architecture

The architecture presented in figure 3 has retained the HLA standard to solve interoperability problems between heterogeneous components introduced.

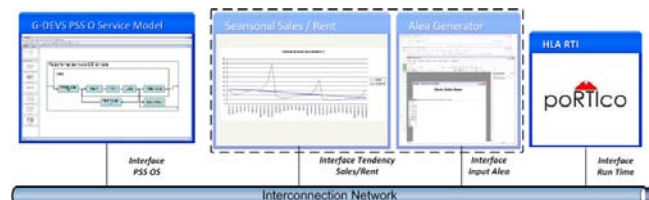


Figure 3: PSS M&S Distributed Architecture

The open source Run Time Infrastructure poRTIco takes in charge the distributed execution and message synchronized delivery. This architecture is based on previous works to handle message passing synchronization consideration (Zacharewicz, Frydman and Giambiasi, 2008; Tu, Zacharewicz and Chen, 2010). The simulation has been run on several scenarios.

One simulation scenario has leaded to compare from the client point of view the best valuable option regarding buying and renting toys for under one year child. The results conclude on the necessity of a fine tuning of the renting duration (2 months or 3 months for 5 toys dedicated to an under one year child) and price regarding buying (less than 30 €). The second measurement was about the impact on the environment, the study has shown that renting can divide by three the impact by reusing the toys regarding the age of the child, the interest duration to a toy from a child and toys non breaking duration.



## 6. CONCLUSION

The study has envisaged that profitability can be obtained under the condition PSS design and realization consider simultaneously the design of products and services, taking care of the use of the system and of its impact on the environment. A special attention has been paid to simulate PSS behaviour regarding classical selling systems to be sure that the realized PSS meets customer needs. This paper has presented premise of a Product-Service Modelling & Distributed Simulation Environment based on G-DEVS / HLA. Therefore, a toys industry innovative PSS scenario has been simulated and tested regarding specific criteria allowing a decision maker to choose an innovation baby toys distribution win/win strategy. Future research will include testing other scenarios and integration of other criteria for the PSS solution choice.

## REFERENCES

- Alix, T., Vallespir, B., 2010b. PSS design based on project management concepts. *Proceedings of the CIRP IPS2 Conference*, pp. 211-217. April 14-15, Linköping, Sweden.
- Alix, T., Vallespir, B., 2010a. Product-Service System development based on project management: the definition sequence. *Proceedings of the Advanced in Production Management Systems*. October 11-13 Cernobbio, Italie.
- Aurich, J.C., Mannweiler, E., Schweitzer, E., 2010. How to design and offer services successfully. *CIRP Journal of Manufacturing Science and Technology*, 2(3), 136-143.
- Bell, M., 2008. *Service-Oriented Modeling: Service Analysis, Design, and Architecture*. Hoboken, NJ: John Wiley & Sons, Inc.
- Chekitan, D. S., Schultz, D.E., 2005. In the Mix: A Customer-Focused Approach Can Bring the Current Marketing Mix into the 21st Century. *Journal of Marketing Management*, 14(1), 16-22.
- Giambiasi N., Escude B., Ghosh S., 2000. GDEVS: A generalized discrete event specification for accurate modeling of dynamic systems. *Transactions of the Society for Computer Simulation International*, 17(3), 120-134.
- Goedkoop, M.J., van Halen C.J.G., Harry R.M., Riele H.R.M., and Rommens P.J.M., 1999, Product Service systems, Ecological and Economic Basics, Available from [accessed June 2011]: <http://teclim.ufba.br/jsf/indicadores/holan%20Product%20Service%20Systems%20main%20report.pdf>
- Goyhenetche, M., 1999. *Le marketing de la valeur, Créer de la valeur pour le client*. Paris: INSEP.
- IEEE std 1516-2010, 2010. *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) - Framework and Rules*. New York, NY: IEEE.
- Kanda, Y., Kakagami, Y., 2006. *What is Product-Service Systems (PSS) – A Review on PSS research and Relevant Policies*. IGES Kansai Research Centre Discussion Paper, KRC-2006-No.1E
- Maxwell, D., Sheate, W., Van der Vorst, R., 2006. Functional and systems aspects of the sustainable product and service development approach for industry. *Journal of Cleaner Production*, 14, pp. 1466-1479.
- Mont, O., 2001. Clarifying the concept of product-service system. *Journal of Cleaner Production*, 10, pp. 237-245.
- Morelli, N., 2006. Developing new product service systems (PSS): methodologies and operational tools. *Journal of Cleaner Production*, 14(17), 1495-1501.
- Meier, H., Roy, R., Seliger, G., 2010. *CIRP Annals – Manufacturing technology*. Available from: Doi:10.1016/j.cirp2010.05.004
- Muller-Stewens, G., Lechner, C., 2005. *Strategisches Management*. Stuggart: SchaefferPoeschel.
- Porter, M.E., 1985. *The Competitive Advantage: Creating and Sustaining Superior Performance*. N.Y.: Free Press.
- Tu, Z., Zacharewicz, G., Chen, D., 2010. Unified Reversible Life Cycle for Future Interoperable Enterprise Distributed Information Systems. *Proceedings of the Interoperability for Enterprise Software and Applications Conference*. April, Coventry, UK.
- Tukker, A., Tischner U., 2006. Product-services as a research field: past, present and future. Reflections from a decade of research. *Journal of Cleaner Production*, 14, pp. 237-245
- Vandercammen M., Warrant C., Meirman A., 2010. *Le Marche du Jouet*. CRIOC, Edition Novembre 2010, NE 417541646.
- Vandermerwe, S., Rada, J.F., 1988. Servitization of Business-Adding Value by Adding Services, *European Management Journal*, 6(4), 314-324.
- Wise, R., Baumgartner, P., 1999. Quand l'industrie se met à l'heure du service. *L'Expansion Management Review*, December, 6-15
- Zacharewicz G., Frydman C., Giambiasi N., 2008. G-DEVS/HLA Environment for Distributed Simulations of Workflows. *Simulation*, 84(5), 197-213
- Zacharewicz, G., Alix, T., Vallespir, B., 2009. Services modeling and distributed simulation DEVS / HLA supported. *Proceedings of the Winter Simulation Conference*. December 13-16, Austin, USA.

## AUTHORS BIOGRAPHY

**Thecle ALIX** is Associate Professor at the University of Bordeaux 4. Her research activities focus on the design, modeling and control of PSS. She is interested in servicizing and sustainability. She was involved in European research projects and is currently involved RAUDIN that deals with the use for the development of numeric disposals (Feder project).

**Gregory ZACHAREWICZ** is Associate Professor in Bordeaux 1 University (IUT MP). His research interests include Discrete Event Modeling (e.g. DEVS, G-DEVS), Distributed Simulation, HLA, MDA, Short lived Ontologies. He recently focused on Enterprise Modeling and Interoperability. He served as reviewer in SCS Conferences and Simulation journals. He is involved in several French and European projects.

# DESIGN AND DEVELOPMENT OF REALISTIC FOOD MODELS WITH WELL-CHARACTERISED MICRO AND MACRO-STRUCTURE AND COMPOSITION.

Axelos M.<sup>(a)</sup>, Daudin J-D<sup>(a)</sup>, Della Valle G.<sup>(a)</sup>, Perrot N.<sup>(a)</sup>, Renard C.M.G.C.<sup>(a)</sup>, Sautot C.<sup>(b)</sup>, Sebedio J-L.<sup>(c)</sup>

<sup>(a)</sup> INRA, Division CEPIA, BP71627, Nantes 44316 France

<sup>(b)</sup> INRA-Transfert, BP71627, Nantes 44316 France

<sup>(c)</sup> INRA, Division AlimH, St-Genes Champanelle 63, France

<sup>(a)</sup> [monique.axelos@nantes.inra.fr](mailto:monique.axelos@nantes.inra.fr), [jean-dominique.daudin@clermont.inra.fr](mailto:jean-dominique.daudin@clermont.inra.fr), [dellaval@nantes.inra.fr](mailto:dellaval@nantes.inra.fr), [nathalie.perrot@grignon.inra.fr](mailto:nathalie.perrot@grignon.inra.fr), [catherine.renard@avignon.inra.fr](mailto:catherine.renard@avignon.inra.fr), <sup>(b)</sup> [caroline.sautot@paris.inra.fr](mailto:caroline.sautot@paris.inra.fr), <sup>(c)</sup> [jls@clermont.inra.fr](mailto:jls@clermont.inra.fr)

## ABSTRACT

This article presents an EU project, DREAM, started two years ago, that develops realistic, physical and mathematical food models to be used as standards that can be exploited across major food categories. We underline the need for physical models realistic enough to cope with the wide diversity of food products and we detail their selection criteria, keeping in mind that these models should be versatile enough to favour the development of common approaches to risk assessment and nutritional quality for food research and industry. We also summarize the different sources of knowledge (scientific from various disciplines, expert engineers ...) that have to be completed, the essential properties to be determined at every structural scale and the scientific models of different phenomena to be integrated. An example of reverse-engineering approach is also proposed to optimise technical pathways in food processing, as aimed to be promoted by DREAM project.

Keywords: model, multi-scale, cellular medium, plant, meat, dairy, cereals, properties, knowledge, integration.

## 1. INTRODUCTION

The physical and chemical properties of food constituents, their interactions and assemblies in food matrixes influence consumer's health in terms of bioavailability of nutrients, phytochemicals and toxicants as well as microbial food safety and quality. Beside macronutrients delivery, food may provide a wide variety of positively bioactive molecules like peptides in meat or milk products (Meisel, 2004), phytochemicals like polyphenols which have a contribution in preventing cardiovascular diseases (Das and Das, 2007), but also negatively active such as mycotoxins in cereal based products (Visconti et al., 2004). For solid foods, the release of these compounds, and thus their bioaccessibility, is largely affected by the complex multiscale structure of the matrix elaborated during processing. The safety and quality of food depend on its propensity to support microbial growth, spores germination and toxin synthesis. Despite the

amount of work on the fate of microorganisms in the food chain (Antwi et al., 2007), the importance of food structure on these phenomena is poorly understood and has not been exploited in food preservation, in food risk assessment as well as in the emerging field of the assessment of food functionality.

For long, food processing was mostly dedicated to product safety, stabilization and scaling-up operations in industry; process engineers applied concepts from chemical engineering and focused on time-temperature diagrams for predicting micro-organisms survival and growth in foods. Nowadays, the consumer's driven demand for products of high sensory and nutritional quality has brought together the interests of process engineers and food scientists on "how to build up the right structures" of foods (Aguilera, 2006). The need for a pluridisciplinary approach is illustrated by the emergence of food material science which considers food as a multi-scale object, with various possible interactions and dynamics (Donald, 2004). Indeed, changes in food product structure resulting from the modification of a single physical variable (moisture, temperature, residence time...) have been largely studied at a single structural level. However, neither changes in state nor their dynamics and spatial distribution have been thoroughly studied yet, and products are frequently analysed after being processed through experimental designs which consider processes as "black boxes". Conversely, strong progresses have been made in modelling food processes (Bimbenet *et al.*, 2007). But still, their extension to the design and processing of real foods encounters limits in the unavailability of material properties, the uncertainty linked to the variability of raw materials, the diversity of conceptual frame of existing models (Rodriguez-Fernandez *et al.*, 2007). As a consequence, the resulting models of food structure modification under processing can be used only over a tiny experimental domain and the links with their end-properties scarcely predicted.

In this context, the DREAM project (<http://dream.aaeuropae.org/>) aims (i) to propose realistic generalized model foods (GMF) to be used by

downstream partners for main properties assessment, and (ii) to integrate the basic knowledge models (BKM) available under various formalisms for designing them (Fig.1). To address the first objective, realistic, physical and mathematical food models (GMF) of the four major food categories (vegetable and fruit products, meat, dairy and bakery products) are developed in four workpackages (WP2-5). In relation with each WP, WP1 fulfills the second objective. WP6 and 7 assess the relevance of GMF and BKM respectively for toxicology, microbiology, nutritional tests (WP6) and for transfer to industry, specially SMEs (WP7).

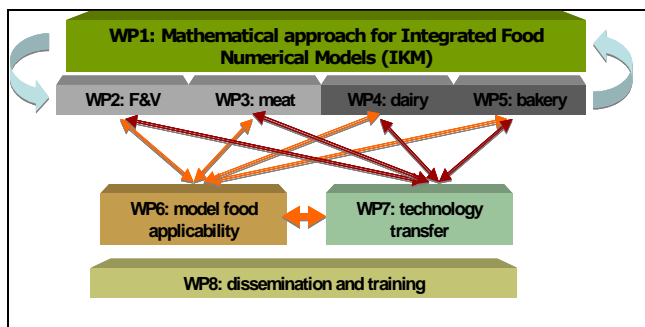


Figure 1: Organization of Dream project

In the following, these works are illustrated by a few examples of the design of GMF (section 2) and BKM building (section 3) after having stated their knowledge background at each step.

## 2. MODEL FOODS (GMF)

The four major food categories can be envisioned according to generic models of their structure which suggests according to their physical appearance and technical pathway listed in Table 1.

Table 1: first classification of model food structure

	Generic model structure	Example of food product
Raw and semi-processed	Filled cellular solids	Tomato, cabbage, apple
	Protein cellular networks	Pork, poultry, beef
Formulated and processed	Combined gelled/dispersed/aerated systems	Yoghurts, creams, cheeses, protein based foams
	Open solid foams	Bread, biscuits, snacks

The GMF definition includes the food product, its structural characterization, its processing pathway and the guidelines for determining them. After enhancing their generic feature, in the following, we illustrate those by two examples from Table 1.

### 2.1 The multi-scale cellular medium

The knowledge gained on the multi-scale structure of these model foods suggests a generic scheme of multi-scale cellular medium with different structural entities, in analogy with cellular solids, the properties of which

are modeled from the knowledge of the properties of these entities (Gibson & Ashby, 1997). The cells (or bubbles in foam, droplets in emulsion) are entrapped in a network of walls or fibres (meat) which may adhere to each other or be separated by an interface. The whole organization, represented in fig.2 as an equivalent effective medium, can be defined by an architecture where the preceding entities may be repeated and associated, according to more or less periodic and isotropic arrangements.

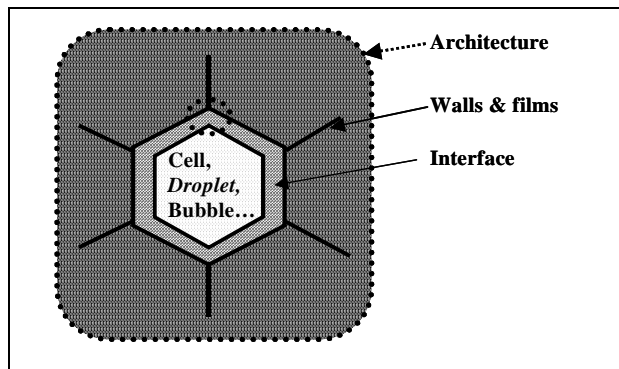


Figure 2: Multi scale cellular representation of model food structure

Depending on the information available and the scale observed, each entity may in turn defined as a cellular medium at a lower scale, which finally allows to achieve the necessary change of scale. Although schematic, this representation is based on all the observations performed by instrumental methods (Microscopies, Nuclear Magnetic Resonance, X-ray tomography) and quantitative inputs may be provided by image analysis. Its interest also relies on the possibility to infer the end-use properties (nutritional, microbiological, toxicological) of the product from the contribution of each structural entity, which in turn depends on the food category considered. For instance, many minor bioactive compounds may be stored close to the interface and released during digestion once the architecture broken down into pieces.

### 2.2 Application to the plant GMF

Most fruits and vegetables are eaten with limited processing and their microstructure is largely imparted by nature. They are a valuable source of a number of micronutrients, which show interesting activities related with human health. Because of the wide diversity of plant tissues, different models must be developed that are representative of leaf, root and fleshy fruit tissues, in order to relate the composition of fruit and vegetables to their functional and nutritional properties. Data accumulated on their composition, including influence of variety and growing conditions, may cover this area, but these tables give no indication on the actual amount available for physiological activity after absorption in the gut. Moreover, the bioavailability of micronutrients is extremely variable due to their chemical structure and to the plant matrix (Parada and Aguilera, 2007). The micronutrients are largely confined into a cellular

structure determined by cell walls (Fig.2), from which they need to be released for absorption.

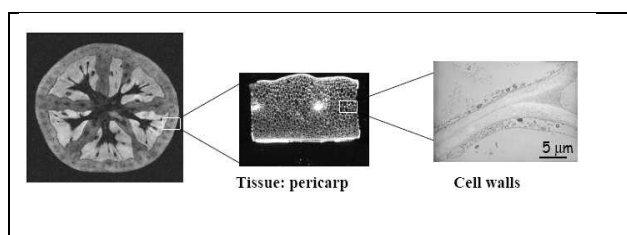


Figure 3: Multi-scale structure of a fleshy fruit selected as model of plant food product (Devaux et al., 2009).

We thus need to understand how the varying resistance of cell walls to mechanical disruption at the cellular level can be understood and modelled in raw plants, by using tissues with e.g. different histology and cell wall thickness (Fig.3); how processing and in particular thermal processing can change these properties, and create pores through which water-soluble molecules may diffuse, and gaps sufficient for migration of fat droplets and micelles; how this interacts with the form in which the nutrients are present in the food, and in particular their solubility. Carotenoid bioaccessibility from tomatoes is thus modulated by the sequence in which heat and grinding treatments are applied, in link with cell-wall degrading enzymes activities. The model of closed, or fluid-filled, cellular solids (Warner *et al.*, 2000) can be adapted to relate structure-mechanical properties by coping with turgor pressure and determine the various scale elements (cell and cell wall, tissue, organ...) to be used as input parameters.

## 2.2 The biomimetic meat GMF

Meat products are built of protein networks, a major source of essential amino-acids. Proteins undergo oxidations and conformational changes, especially during cooking, which could decrease their nutritional value. In order to better understand the mechanisms involved in these changes and their related kinetics under processing, whilst getting rid of biological variability, a mimetic model has been developed (Promeyrat et al., 2011). This model is based on a suspension of myofibrillar proteins extracted from 2 pure fiber type muscles (rabbit). Oxidants with various iron concentrations were added to enrich this basic model, then heat treated between 45 to 90°C, 5 to 120mn (Fig. 4). Protein oxidation and denaturation were assessed by measurement of carbonyl groups and protein surface hydrophobicity, respectively. Results show that heat treatment alone do not lead to amino-acids oxidation whereas there is a synergy between oxidants and heat. Comparison of the mimetic model with pork meat for the same iron concentration (Fig.4) underlines the protection by antioxidants in the latter case. Conversely, measurements of protein surface hydrophobicity showed that thermal denaturation occurred very rapidly ( $\leq 5\text{mn}$  for  $T \geq 75^\circ\text{C}$ ) for pork meat as for models.

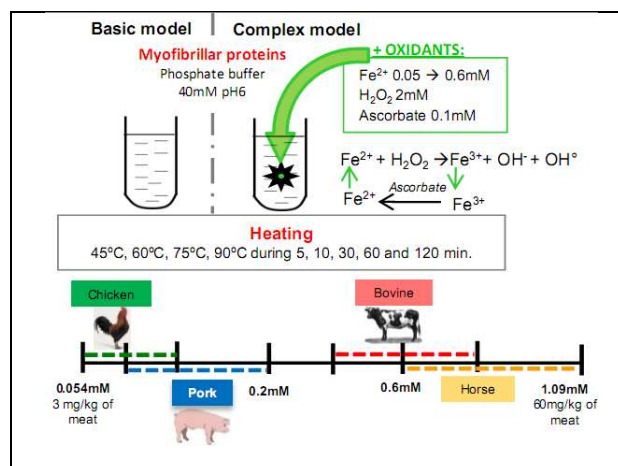


Figure 4: Design and process of meat GMF mimetic and iron content in different animal species (Promeyrat et al., 2011).

The latters will be further complexified by adding antioxidants in order to better mimick real meat products. Results will contribute to determine mathematical models for changes induced by oxidation and thermal denaturation of proteins in meat during cooking.

## 3. FOOD MODELS

In addition to the limits encountered in modelling food processes, the close interaction between continuous structural changes and transfer mechanisms impairs the complete modeling of coupled physical, chemical and microbiological phenomena. The fragmentation and incompleteness of our knowledge require integration of technological know-how in mathematical models (Perrot *et al.*, 2011). In this purpose, the whole food chain system can be viewed as a complex one, like in recent scientific issues in biology. The emerging science of complex systems proposes new ways to understand systems located in turbulent, instable and changing environments, which are not predictable within a conventional scientific framework.. The application of those concepts relies on the tools able to take explicitly into account the fragmented and heterogeneous knowledge available on the dynamics of the process with uncertainty on the global behavior of the system.

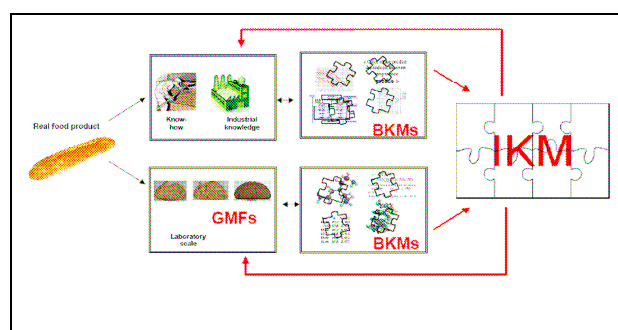


Figure 5: integrating model knowledge (IKM) of food processing; GMF= generalized model foods; BKM = basic knowledge model.



[illegible]

For each model food, available BKM<sub>s</sub> have to be validated and integrated into a mathematical model (IKM) for multistage dynamic reconstruction of food models (GMFs), which can in turn be implemented for reverse engineering. The following illustrates this approach taking example on two other GMFs from Table 1.

Open solid foams can be considered as valuable models for most cereal and bakery foods since their physical properties can be determined from (1) the intrinsic properties of the solid phase, (2) product density and (3) cellular structure. The contribution of product density can be derived from Gibson and Ashby's model (1997) for cellular solids, at least for mechanical properties, but the contribution of other factors is poorly known. Recently numerical finite element models (FEM) of solid foams have allowed assessing the effect of cellular structure described by images from X-ray tomography on product texture (Guessasma *et al.*, 2008). Besides extending Gibson & Ashby's model, these numerical simulations have shown that the more heterogeneous the cellular structure, the stiffer the foam, due to the regularity of cell wall size (Fig. 7). This result may have

Besides, FEM may be also apply to model the mechanical behavior of the solid phase of the foam, envisioned as a composite of starch and gluten (and other components), and the properties of which depend on composition and morphology (Fig. 3). Mechanical properties of such composite materials can then be used as inputs to compute the solid foam properties, which allows achieving the change of scale. For vitreous brittle products, the lack of adhesion between starch and proteins has been shown of great importance on composite behavior and its fragility, promoting rupture at phase interface (Chanvrier et al., 2006). Indeed, the fracture properties and water dynamics in the matrix have to be determined because they govern the sensory quality, the stability and also the digestibility, through the delivery of small molecular weight components in the gastrointestinal tract and thus envision their de-structuration during mastication.

By extending the use of numerical methods and improving experimental ones, it will be possible to better cope with the changes of composition (dietary fibres, fatty acids, enzymes...) at different structural levels during processing from the divided solid medium (flour), through mechanical and thermal actions, to a continuous matrix with multi-scale porosity.

Cheese manufacturing is certainly one of the most representative areas of food industry in Europe. In spite

of this industrial importance, soft cheese like Camembert is an ecosystem and a bioreactor difficult to assess in its entirety. Despite extensive research conducted on this product, knowledge remains fragmentary and incomplete and no model provides a comprehensive representation of the process. In this context, Dynamic Bayesian Networks (DBN) have been used to model the network of interactions occurring at different scales and reconstruct its dynamics (Baudrit et al., 2010). The concept of DBNs provides a practical mathematical formalism that enables to describe dynamical complex systems tainted with uncertainty. DBNs are an extension of classical Bayesian networks that rely on probabilistic graphical models in which nodes representing random variables are indexed by time. They are very useful tools for combining expert knowledge with data at different levels of knowledge, where the structure can be explicitly built on the basis of expert knowledge and conditional probability, quantifying dependence between variables, can be automatically learned without *a priori* knowledge on the basis of a dataset. From operational and scientific knowledge, Baudrit *et al.* (2010) defined the structure of a DBN providing a qualitative representation of the coupled dynamics of microorganism behaviour (*Kluyveromyces marxianus* (Km), *Geotrichum candidum* (Gc), *Brevibacterium aurantiacum* (Ba) with their substrate consumptions (lactose (lo), lactate (la)) influenced by temperature (T) and involving the sensory changes (Odour, Under-rind, Coat, Colour and Humidity) of cheese during ripening (Fig.8).

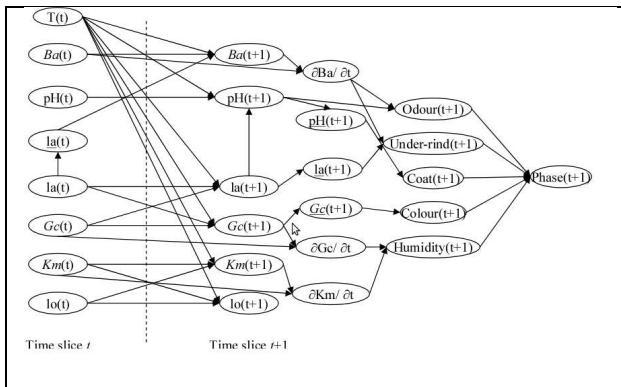


Figure 8: Dynamic Bayesian network representing the coupled dynamics of micro-organism growth with their substrate consumptions influenced by temperature and involving the sensory changes of cheese during the ripening process (from Baudrit et al., 2010).

After the learning step to define conditional probability distributions from experimental trials with various temperature and humidity, DBNs inferences can be carried out in order to simulate the behaviour of microbial activities associated with sensory development, for instance the beliefs of the possible trajectories of the yeast *Km* during ripening at 8°C (Fig.9a). This figure means, for instance, that at the 27th day of ripening, the concentration of *Km* has a probability of 39% to be  $\approx 10^7$ cfu/g FC and that it

cannot be lower than  $3.10^5$ . In addition the mean evolution of *Km* (Fig.9b) as well as the modal evolution of odor properties( Fig.9c) could be estimated by DBN simulations and compared to experimental data.

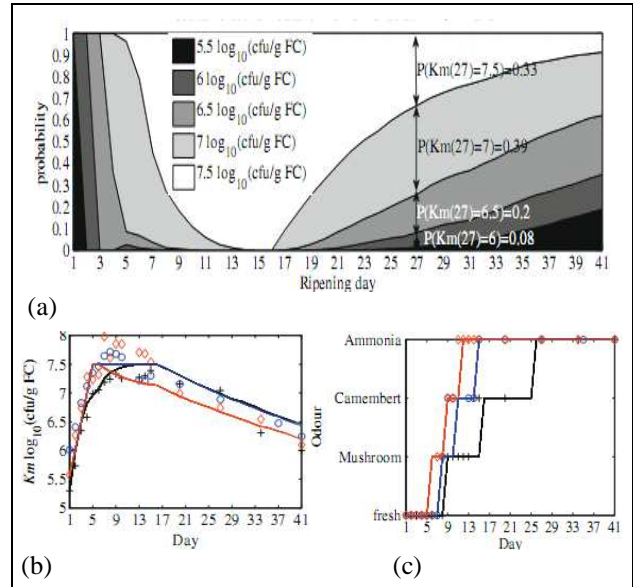


Figure 9: DBN results of (a) *Km*(t) probability distribution at 8°C, predictive evolutions of (b) *Km* microbial growth and (c) odour, versus experimental data for ripening performed at T= 8 (+), 12 (o) and 16°C (◇), RH=98%.

The model was thus shown to be able of (1) coupling and integrating heterogeneous knowledge at different scales; (2) predicting the evolution of microbial activities and sensory properties with an overall average adequacy rate of about 85% to experimental data.

### 3.3. Example of an IKM used for reverse engineering

Following the same practical application, a model of cheese mass loss (Helias et al., 2007) has been considered for optimizing the ripening process. In this purpose, a viability kernel representing a compromise between production costs and ripened cheese quality was computed (Sicard et al., 2009). Viability theory aims at controlling a dynamical system, here cheese during ripening, in order to maintain it in a given set of evolutions, namely the viability kernel. The viability kernel was defined by associating a target on cheese mass at the end of ripening ( $\approx 280$ g) and constraints on microorganisms respiration. It was then computed thanks to the classical heat&mass transfer model, proposed by Helias et al. (2007) that predicts cheese mass, surface temperature and the respiration of the microorganisms. Meanwhile, the cost trajectories, involving the number of control variations and the ripening time are computed to define the compromise between cheese quality and energy consumption saving (Sicard et al., 2009). Then, optimal trajectories, lowering cost, were found; among those, one reached the mass target after 8 days ripening, and the results of



its control variations are presented in Fig. 10a, to be compared to the conventional control performed for 12 days ripening (Fig. 10b), without quality loss. Whilst setting a higher humidity (94%), it imposes a daily change of temperature between 14 and 9°C. These controls have been applied to real ripening chambers; the analysis of processed cheese gave sensory results very close to those obtained under classical conditions (12°C, 92%).

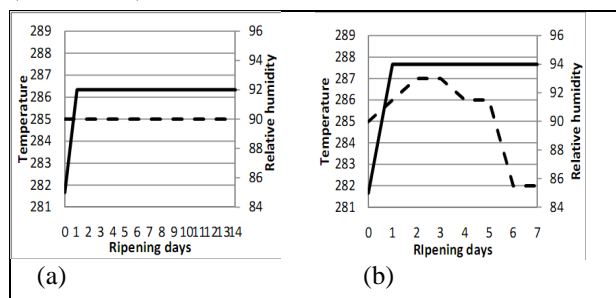


Figure 10: Control (T(K) ---, RH(%) —) of ripening chambers, (a) conventional in industry and (b) optimal computed applying viability approach.

This is an example of the application of the reverse engineering approach to a single food processing operation.

#### 4. CONCLUSION

Food engineering deals with complex systems in which knowledge is incomplete and which depend on many interacting factors. The needed information is not fully available from literature and partly relies on technological expertise. Managing such systems is a real challenge, that can be addressed by Knowledge Engineering. In order to design Model Foods, we have presented some of its recent applications to build Food Models by describing some tools involved in the EU-Dream project. Dream's goals will be achieved by integrating these models, in order to build an integrated mathematical model (IKM) for multistage dynamic reconstruction of foods which can in turn be implemented for reverse engineering. The application of such approach opens prospects for the virtual design of food products, which will be of help for the sustainable production of high quality foods.

#### ACKNOWLEDGMENTS

Results presented in this paper have been obtained in the frame of the EU-FPVI Dream project and ANR-ALIA-INCALIN.

#### REFERENCES

Aguilera, J.M. (2006). Seligman lecture: Food product engineering: building the right structures, *J. Sci. Food Agric.* 86, 1147-1151.

Antwi, M., Geeraerd, A.H., Van Impe J.F., Geeraerd A.H., 2007. Modelling the combined effects of structured food model system and lactic acid on *Listeria innocua* and *Lactococcus lactis* growth in

mono- and coculture. *Int. J. Food Microbiol.*, 120, 71-76.

Baudrit C., Sicard, M., Willemin, P.H., Perrot, N. 2010. Towards a global modelling of the Camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic Bayesian networks. *Journal of Food Engineering*. 98: 283-293.

Beckerman, L.P., 2000. Application of complex systems science to systems engineering. *Systems engineering*. 3, 96.

Bimbenet, J.J., Schubert, H., Trystram, G., 2007. Advances in research in food process engineering as presented at ICEF9. *Journal of Food Engineering*. 78, 390.

Chanvrier, H., Della Valle, G., Lourdin, D. 2006. Mechanical behaviour of corn flour and starch-zein based materials in the glassy state: A matrix-particle interpretation. *Carbohydrate Polymers*, 65, 346-356.

Cohen, I.R., Harel, D., 2007. Explaining a complex living system: dynamics, multi-scaling and emergence. *Journal of the royal society interface*. 4, 175.

Das, S., Das, D.K., 2007. Resveratrol: a therapeutic promise for cardiovascular diseases. *Recent Patents Cardiovasc. Drug Discov.* 2, 133.

Desiere, F., German, B., Watzke, H., Pfeifer, A., Saguy, S., 2001. Bioinformatics and data knowledge: the new frontiers for nutrition and foods. *Trends in Food Science & Technology*. 12, 215.

Devaux, M.F., Legland, D., Bouchet, B., Lahaye, M., Guillon F., 2009. Cartography of the cellular structure of tomato pericarp tissue within the whole fruit using macrovision and confocal microscopy. *Invited Lecture: BioPhys Spring (BPS) The Institute of Agrophysics, PAS, Lublin, Poland, May 21-22<sup>nd</sup>*.

Donald, A.M., 2004. Comment: Food for thought. *Nature Materials*. 579-580.

DREAM project (<http://dream.aeurope.org/>)

Gibson, L.J., Ashby, M.F. 1997. Cellular solids, structure and properties. Cambridge Press University, 510 p.

Guessasma, S., Babin, P., Della Valle, G., Dendievel, R., 2008. Relating cellular structure of open solid food foams to their Young's modulus : finite element calculation. *Int. J. Solids Struct.* 45, 2881-2888.

Guessasma, S., Chaunier, L., Della Valle, G., Lourdin, D., 2011 Mechanical modelling of cereal solid foods. *Trends in Food Science and Technology*, 22, 142-153.

Helias, A., Mirade, P. S., & Corrieu, G. 2007. Modeling of camembert-type cheese mass loss in a ripening chamber: Main biological and physical phenomena. *Journal of Dairy Science*, 90: 5324-5333.

Meisel, H., 2004. Multifunctional peptides encrypted in milk proteins. *Biofactors*. 21, 55.

- Ndiaye, A., Della Valle, G., Roussel, P., 2009. Qualitative modelling of a multi-step process: the case of *French breadmaking*. *Expert Syst. with Appl.*, 39, 1020-1032.
- Parada J., Aguilera J.M., 2007. Food microstructure affects the bioavailability of several nutrients. *J. Food Sci.* 72, R21.
- Perrot, N., Baudrit, C., Trelea, I.C., Trystram, G., Bourguine, P. (2011). Modelling and analysis of complex food systems: state of the art and new trends. *Trends in Food Science and Technology*, 22, 304-314
- Promeyrat, A., Daudin, J-J., Santé-Lhoutellier, V., Gatellier, P. 2011. Study of protein physico-chemical changes during heating in mimetic meat model. *ICEF 11*, Athens, May 21-24<sup>th</sup>.
- Rodriguez-Fernandez, M., Balsa-Canto, E., Egea, J.A., Banga, J.R., 2007. Identifiability and robust parameter estimation in food process modelling: application to a drying model. *J. Food Eng.* 83, 374.
- Segui L., Barrera C., Oliver L., Fito P. Practical application of the SAFES (systematic approach to food engineering systems) methodology to the breadmaking process. *J. Food Eng.* 83, 219.
- Sicard M., Perrot, N., Baudrit, C., Reuillon, R., Bourguine, P., Alvarez, I. & Martin, S. 2009. The viability theory to control food processes. *Eur. Conf Complex Systems*, Univ. Warwick (UK).
- Turbin, A., Della Valle G., Fernandez C., Ndiaye, A., 2010. Knowledge book on cereal processing, available from Dream: [https://workspaces.inra-transfert.fr/LotusQuickr/dream/PageLibraryC12576E7003B99D8.nsf/h\\_Toc/E275B6801F8D60F7C12577FF00571E52/?OpenDocument](https://workspaces.inra-transfert.fr/LotusQuickr/dream/PageLibraryC12576E7003B99D8.nsf/h_Toc/E275B6801F8D60F7C12577FF00571E52/?OpenDocument)
- Visconti, A., Haidukowski, E.M., Pascale, M., Silvestri, M., 2004. Reduction of deoxynivalenol during durum wheat processing and spaghetti cooking. *Toxicology Letters*. 153, 181.
- Warner, M., Thiel, B.L., Donald, A.M., 2000. The elasticity and failure of fluid-filled cellular solids: theory and experiment. *PNAS*. 97, 1370.
- meat products processing at QuAPA research unit, near Clermont-Ferrand.
- Guy Della Valle, head of a research group on materials processing and properties at BIA, Nantes, has a great experience in modelling the mechanisms which govern the creation and behaviour of cereal based solid foams.
- Nathalie Perrot develops methodological guidelines to manage the expert-operator knowledge for modelling food processes and has coordinated the ANR-Incalin project that led to the Dream workpackage on knowledge integration.
- Catherine Renard has published more than 70 papers on pectins and polyphenols, 30 of them dealing with apples; she is head of a group working on processing and qualities (nutritional, sensory) of fruit, notably impact of process on composition and availability of anti-oxidant micronutrients and use of infrared for fast analysis of inner quality..
- Caroline Sautot, holds a Master of Science in Biotechnology and a Master in Project Management. She has 5 years experience in the setting up and management of large collaborative research projects. Currently, she manages several FP7 collaborative research projects including the DREAM project.
- Jean-Louis Sebedio., also from INRA Clermont-Ferrand, has published more than 90 articles concerning fatty acid studies.

## AUTHORS BIOGRAPHY

Monique Axelos, coordinator of DREAM, is a physico-chemist at INRA-Nantes, France. She is the head of the Science and Engineering of Agricultural Products division (French acronym: CEPIA) which provides knowledge on raw materials of animal or plant origin and on their transformations for foods and non food uses (chemistry, energy, materials), and contributes to the development of new products and new technologies in the food and green chemistry sectors in the context of sustainable development.

Jean-Dominique Daudin has worked long on modeling of heat and mass transfers in food processing and on

# FROM EXPERT KNOWLEDGE TO QUALITATIVE FUNCTIONS: APPLICATION TO THE MIXING PROCESS

Kamal Kansou<sup>(a,b)</sup>, Guy Della Valle<sup>(a)</sup>, Amadou Ndiaye<sup>(b)</sup>

<sup>(a)</sup>INRA, UR 1268 Biopolymères Interactions et Assemblages, Nantes, 44316 France

<sup>(b)</sup>INRA, CNRS, Univ. Bordeaux 1, UMR 927, Institut de Mécanique et d'Ingénierie, Talence, 33405 France

<sup>(a)</sup>[\[kamal.kansou, Guy.Della-valle\]@nantes.inra.fr](mailto:[kamal.kansou, Guy.Della-valle]@nantes.inra.fr), <sup>(b)</sup>[Amadou.ndiaye@bordeaux.inra.fr](mailto:Amadou.ndiaye@bordeaux.inra.fr)

## ABSTRACT

We present a procedure of knowledge representation based on a qualitative algebra, to predict the wheat flour dough behaviour from mixing settings. The procedure guarantees the consistency of the knowledge base and provides a concise and explicit representation of the knowledge. The qualitative model is implemented as a knowledge-based system (KBS) accessible and understandable by scientists and technologists in breadmaking. The KBS is a record of the domain knowledge, mainly know-how, and a tool to confront predictions of the dough condition with real observations. An example of such a confrontation about the wheat flour dough mixing process is shown; the results gives insight into ill-known relations between the process settings and the dough condition.

Keywords: qualitative modelling, know-how, breadmaking, expert knowledge.

## 1. INTRODUCTION

The general idea that domain know-how can do a lot for the production of scientific knowledge becomes more concrete, for example in the domain of knowledge management (Van de Ven and Johnson 2006) and agronomy (Girard and Navarette 2006). Indeed, in the domain of food industry, the management of production still relies on know-how, while scientific knowledge explains a part of the phenomena occurring during a process. Therefore know-how can help to point out the lack of knowledge with questions like why this practice fundamentally works or how can we improve it? Answering such interrogations will drive the production of an operational scientific and technical knowledge that in turn will support the improvement of practices.

The challenge is to elicit and represent the know-how so to make it accessible for any food scientists and technologists involved in production. As a matter of fact, the knowledge related to the implementation of food processes is partly tacit as shown by Nonaka and Takeuchi (1995) in their work on the bread making machine. Moreover know-how is valid in a given context of production, seldom defined or even known.

To address this issue we chose to work with a group of technologist experts and domain researchers to

build a knowledge-based system (KBS) on a given topic in food science so to select operational knowledge with scientific consistency. To structure the knowledge of different sources we adopted a systemic approach. Both choices seem equally important to hand on the knowledge of the knowledge base to the different professional communities.

The domain of application is the French breadmaking. Breadmaking is a multistage processing chain, the management of which relies on professional know-how, without direct input from the large scientific literature available on wheat flour dough rheology and structure.

The KBS should ultimately predict the dough or bread condition from the inputs and processing conditions. At the moment, the modelling phase of the mixing process is over and two models were developed, the pre-mixing operation model (Kansou et al. 2008) and the texturing operation model (Ndiaye et al. 2009). The knowledge on the mixing process is mostly a domain know-how consistent with scientific principles. One of the first challenge of this work was the following:

- building an explicit representation of the expert knowledge and maintain the consistency of the KB when it is updated or refined. Indeed, because the KB ought to be a repository of knowledge about a given topic, it needs to be updated along with the advance of the domain knowledge.

In this article, we present a knowledge representation procedure that addresses this point through the building of a qualitative model of the expert reasoning for the mixing process.

In the background section we present features of the expert knowledge in breadmaking and the qualitative algebra, *Q*-algebra, which is the formalism developed to represent this knowledge. Then, we illustrate the knowledge representation procedure with an example taken from the qualitative model of the texturing operation, the second stage of the mixing process. Then we describe briefly the KBS about the mixing process and the result of the validation stage.

Finally we will present results from the confrontation of the KBS' predictions against real observations of the dough condition.

## 2. KNOWLEDGE AND TECHNOLOGICAL BACKGROUND

French breadmaking has long been a traditional activity relying on craftsman's manual skills. Today, part of the baker's work is automated and it is also an industrial activity, which demonstrates a good level of knowledge about the ingredients and the baking process. From a physico-chemical point of view, the successive stages of the breadmaking bring about an initially dispersed granular medium (flour) to become a visco-elastic homogeneous mass (mixed dough), aerated (after the fermentation) and finally fixed (after the baking) (Bloksma 1990). However the causal relations between the physico-chemical properties of the components and the sensory and nutritional characteristics of bread, according to the sequence of unit processes (mixing, proofing, laminating... cooking), remain ill known; so the development of scientific models of the whole process or even unit operations is still challenging.

Among the operations of breadmaking, mixing is crucial because it covers the formation of the dough, and yet one of the most ill-known stage of the process (Stauffer 2007). This motivates the building of a qualitative model of the mixing operation, to state what the know-how says about the relationship between the dough behaviour and the setting of mixing.

## 3. KNOWLEDGE IN BREADMAKING, HOW EXPERTS PREDICT THE DOUGH CONDITION

### 3.1. Descriptors and evaluation grid to assess dough condition

An important resource of this work is the standard procedure for the evaluation of the French breadmaking process (including ingredients quality) (NF V03-716, 2002). The procedure includes an evaluation grid of the dough and the bread (Tab. 1). This grid is now widely used in French baking technical centres and training institutes. After each step of the breadmaking process a trained baker describes the dough condition according to a set of sensory descriptors, which reflect important properties of the dough. Sensory descriptors are assessed following a scale of notation of 7 levels centred around the normal level considered as the reference value for a standard French bread. Assessment of a deficiency in a property ranges over 3 levels (very insufficient, insufficient, slightly insufficient) as well as the excess of a property (very excessive, excessive, slightly excessive). Some descriptors can only be assessed in excess or in insufficiency, for example a dough can be excessively sticky but never insufficiently sticky.

Table 1. Standard scoring of the dough mixing operation according to the standard of the French breadmaking process (NF V03-716, 2002)

The scale of expert		Very insufficient	Insufficient	Slightly insufficient	Normal	Slightly excessive	Excessive	Very excessive
Notation		1	4	7	10	7	4	1
<b>MIXING</b>								
SMOOTH					x			
ASPECT								
STICKINESS						x		
CONSISTANCY					x			
EXTENSIBILITY						x		
ELASTICITY					x			
STABILITY					x			

### 3.2. Type of expert knowledge

The domain expert knowledge is expressed first as simple rules "If-Then", which are punctual knowledge such as:

*"If the protein content of the flour is high (>12%) then the dough consistency at the end of the first mixing will be excessively firm"*

or

*"If the protein content of the flour is high (>12%) and the water content of the dough slightly high then the dough consistency at the end of the first mixing will be normal"*

In these examples the dough consistency is a sensory descriptor, whereas the protein and the dough water contents are measurements.

Besides this punctual knowledge, the expert knowledge consists also in functional relationships between two variables, called functional knowledge. This kind of knowledge is expressed with assertions such as "the more X, the more Y". For example:

*"The higher the flour protein content the more the firmness of the dough"*

and also

*"The higher the dough water content the more the softness of the dough"*

Note that these two functional knowledge are consistent with the two previous If-Then rules.

According to Dieng et al. (1995), the functional knowledge plays an important role in the expert reasoning. It results from the abstraction of a scientific knowledge, a physical law for example, or from the synthesis of many observations or trials. Such knowledge expresses the way a variable behaves along with another and therefore should be captured in the knowledge base. A classical rule-based representation is inadequate in this regard.

## 4. BACKGROUND

### 4.1. Qualitative algebra ( $Q, \approx, \oplus, \otimes$ )

This work is based on the  $Q$ -algebra defined in Ndiaye et al., (2009). Here are summarized the basic elements of the formalism:

- A quantities space ( $Q$ ) of seven symbolic elements strictly ordered ( $vvl < vl < l < m < h < vh < vvh$ ), as defined in Guerrin (1995), that maps exactly the scale of notation used by experts to assess the descriptors of the dough, and with  $\{?\}$  the indecision symbol;

- For measurements whose domain of value is the set of real numbers, elements of  $Q$  are representative of the absolute order of magnitude based on a partition of the real line:  $]-\infty, x1]$ ,  $]x1, x2]$ ,  $]x2, x3]$ ,  $]x3, x4]$ ,  $]x4, x5]$ ,  $]x5, x6]$ ,  $]x6, +\infty[$ . For observations, a symbolic scale of a maximum of seven elements is used. The interpretation of an observation depends on the context, for a sensory descriptor the scale of assessment is: very insufficient, insufficient, slightly insufficient, normal, slightly excessive, excessive, very excessive.

- A qualitative equality ( $\approx$  meaning “possibly equal to”).  $\approx$  is reflexive, symmetrical, and intransitive in the general case:

$$(x \approx y)_{def}$$

$$\forall x \in Q, \exists y \in Q : x \approx y \Leftrightarrow \exists z : z \subseteq x \wedge z \subseteq y$$

- A qualitative addition ( $\oplus$ ), whose definition is given in tables 2.  $\oplus$  is commutative, associative, admits  $m$  as a neutral element and admits the symmetrical element ( $\forall x \in Q, \exists x' \in Q : x \oplus x' = x' \oplus x \approx m$ );

- A qualitative multiplication ( $\otimes$ ), whose definition is given in tables 2.  $\otimes$  is commutative, associative, admits  $h$  as a neutral element,  $m$  as an absorbing element, does not admit a symmetrical element and is qualitatively distributive compared to  $\otimes$ ;

- Two specific functions ( $T$  and  $\perp$ ), whose definitions are given in table 3.  $T$  and  $\perp$  have the following property:  $T(x) \oplus \perp(x) = x$ . Those two functions were introduced to represent non-linear evolution, such as saturation or initiation, up to or beyond a given threshold, respectively.

Operators and specific functions of the  $Q$ -algebra map the basic cognitive operations used by the experts to predict a dough condition. A complex reasoning is represented by a combination of these basics functions.

### 4.2. Modelling the breadmaking process

A breadmaking process is seen as a sequence of breadmaking operations. An operation is a transformation of the dough whose state is formally described by a set of state variable which represent the sensory descriptors (see section 3.1). Each operation accepts as input a set of state variable describing the dough resulting from the preceding operation, except the first one (pre-mixing) which transforms the ingredients into a dough. Each breadmaking operation is

controlled by its control variables that capture the settings and adjustments of the equipments or the baker's actions.

Tables 2. Definition of the qualitative addition ( $\oplus$ ) and multiplication ( $\otimes$ ) in the  $Q \cup \{?\}$  space (Ndiaye et al. 2009)

$\oplus$	vvl	vl	l	m	h	vh	vvh	?
vvl	vvl	vvl	vvl	vvl	$[vv1, vl]$	$[vv1, l]$	?	?
vl	vv1	vv1	vv1	vl	l	m	$[h, vvh]$	?
l	vv1	vv1	vl	l	m	h	$[vh, vvh]$	?
m	vv1	vl	l	m	h	vh	vvh	?
h	$[vv1, vl]$	l	m	h	vh	vvh	vvh	?
vh	$[vv1, l]$	m	h	vh	vvh	vvh	vvh	?
vvh	?	$[h, vvh]$	$[vh, vvh]$	vvh	vvh	vvh	vvh	?
?	?	?	?	?	?	?	?	?

$\otimes$	vv1	vl	l	m	h	vh	vvh	?
vv1	vvh	vvh	vvh	m	vv1	vv1	vv1	?
vl	vvh	vvh	vh	m	vl	vv1	vv1	?
l	vvh	vh	h	m	l	vl	vv1	?
m	m	m	m	m	m	m	m	m
h	vv1	vl	l	m	h	vh	vvh	?
vh	vv1	vv1	vl	m	vh	vvh	vvh	?
vvh	vv1	vv1	vv1	m	vvh	vvh	vvh	?

Table 3. Definition of specific functions  $T$  and  $\perp$  in  $Q \cup \{?\}$  space

$x$	vv1	vl	l	m	h	vh	vvh	?
$T(x)$	vv1	vl	l	m	m	m	m	?
$\perp(x)$	m	m	m	m	h	vh	vvh	?

## 5. KNOWLEDGE REPRESENTATION PROCEDURE, FROM EXPERTS TO THE KNOWLEDGE BASE

To elicit the way variables influence each other, we asked a group of three experts (two technologists and one rheologist) to fill in decision tables, that are basically relationship matrices between the input variables and the state variables (output variables) of a given operation of the breadmaking process. Experts collaborate to provide consensual relationship matrices. Relationship matrices are translated in tables of quantities following the projection from the experts' scale to  $Q$ .

The relationship matrices collected so far, capture the influence of one qualitative variable on another,  $x$  influences  $y$ , or two variables on another,  $x$  and  $y$  influence  $z$ . Most of the influences have a  $q$ -algebraic expression of the type  $y = f(x)$  or  $z = f(x, y)$ ,  $f$  being a qualitative function of the  $Q$ -algebra.

As an example of the knowledge representation procedure let us consider the prediction of the dough consistency at the end of the mixing operation, *Cons*. *Cons* is a state variable influenced by the dough self-heating  $\Delta T$ , caused by viscous dissipation, the consistency at the beginning of the mixing, *Cbm*, and the temperature at the end of the mixing, *Tem*. They can be considered as control variables since the first two are

tuned by an experienced baker whereas the third is actually a target temperature used by bakers to control the mixing process. Here is the corresponding influence graph (Fig. 1)

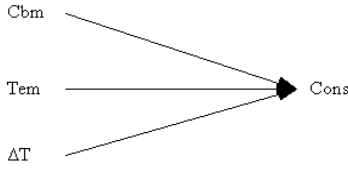


Figure 1. Influence graph of the prediction of the dough consistency at the end of mixing

#### Projection step

Measurements and observations used by experts to describe the mixing conditions are translated in quantities. This is done through a projection operation defined as follows:

$$Pr: \mathcal{R} \rightarrow \mathcal{Q} \quad \text{or} \quad V \rightarrow \mathcal{Q}$$

With  $V$  a vocabulary space representing the scale of assessment of a given observation. Tables 4 show the projection for the three control variables of the mixing operation in the quantities space  $\mathcal{Q}$ .

Tables 4. Projection tables for control variables

<b>Cbm</b>		<b>Tem</b>	
Measurement	Quantity	Measurement	Quantity
(UF)	(w)	(°c)	(x)
<b>Cbm≤350</b>	l	<b>Tem≤22</b>	l
<b>350&lt;Cbm≤450</b>	m	<b>22&lt;Tem≤25</b>	m
<b>Cbm&gt;450</b>	h	<b>Tem&gt;25</b>	h

<b>ΔT</b>	Quantity
Observation	(z)
<b>low</b>	l
<b>medium</b>	m
<b>high</b>	h

We note  $w=Pr(Cbm)$ ,  $x=Pr(Tem)$ ,  $y=Pr(\Delta T)$  and  $Cons_i$  the state variable  $Cons$  influenced by the variable  $i$ .

#### Relationship matrices and mapping as qualitative functions

Three relationship matrices (Tab. 5) define the individual influences of  $w$ ,  $x$ ,  $z$  on  $Cons$ . They are represented by the following qualitative functions:

$$Cons_w \approx w$$

$$Cons_x \approx l \otimes \perp(x)$$

$$Cons_z \approx l \otimes z$$

Tables 5. Individual influences of  $w$ ,  $x$ ,  $z$  on  $Cons$

$w$	$Cons_w$	$x$	$Cons_x$	$z$	$Cons_z$
<b>l</b>	l	<b>l</b>	m	<b>l</b>	h
<b>m</b>	m	<b>m</b>	m	<b>m</b>	m
<b>h</b>	h	<b>h</b>	l	<b>h</b>	l

We also need to know how to combine the three individual influences to determine the global qualitative function for the prediction of  $Cons$ . This requires relationship matrices describing the combined influences. The two relationships matrices allowing the identification of the global qualitative function are presented in Tables 6.

Tables 6. Relationship matrices for control variables

$Cons_{wx}$	$Cons_x$		$Cons$	$Cons_z$		
	<b>l</b>	<b>m</b>		<b>l</b>	<b>m</b>	<b>h</b>
<b>l</b>	vl	l	<b>vl</b>	vv	vl	l
<b>Cons<sub>w</sub> m</b>	l	m	<b>l</b>	vl	l	m
<b>h</b>	m	h	<b>m</b>	l	m	h
			<b>h</b>	m	h	vh

The patterns of the tables 6 match with the qualitative addition  $\oplus$  (Tab. 2), meaning that the way experts combine the influences of the control variables is additive. Here follows the representation as qualitative functions:

$$Cons_{wx} \approx Cons_w \oplus Cons_x$$

$$Cons_{wx} \approx w \oplus l \otimes \perp(x)$$

$$Cons \approx Cons_{wx} \oplus Cons_z$$

$$Cons \approx w \oplus l \otimes \perp(x) \oplus l \otimes z$$

The qualitative addition ( $\oplus$ ) represents a kind of reasoning that is recurrent in this work.

Final results of the qualitative calculus have to be interpreted in a vocabulary space to be handed on to domain experts and users. Formally, interpretation is the inverse operation of the projection. The table 7 presents the interpretation of the  $Cons$  values as observations of the dough consistency at the end of the mixing.



Table 7. Interpretation of the qualitative variable *Cons* as dough consistency at the end of mixing

Interpretation		Meaning for the users
<i>Cons</i>	Consistency end of mixing	Implication for the rest of the process
vvl	very insufficient	Dough very liquid, hard to manipulate and to make up . Poor quality bread expected
vl	insufficient	Soft dough, significant deviation from the reference. Requires short proofing stage and soft moulding
l	slightly insufficient	Slightly soft dough. Requires short proofing stage and soft moulding
m	normal	Normal value for a standard breadmaking process
h	slightly excessive	Slightly firm dough. Can still lead to good quality bread
vh	excessive	Firm dough, requires a long proofing stage, can undergo hard moulding

## 6. RESULTS, QUALITATIVE MODEL OF THE MIXING OPERATIONS

### 6.1. Qualitative model of the texturing phase and implementation

The  $Q$ -algebra had been developed so that the set of rules, experts use to assess a property of a wheat flour dough or bread, can be represented by a qualitative function. The  $Q$ -algebra was used to model the states of the dough at the end of the two successive operations of mixing: (1) pre-mixing during which the components are homogenized into a dough and (2) texturing which performs dough aeration and gluten network formation. The state of the dough at the end of the pre-mixing can be described by its consistency; it is influenced by the characteristics of the ingredients (%flour, water, protein and pentosan content...). The state of the dough at the end of texturing is defined by the following 8 descriptors: smoothing velocity (SV), smooth aspect (SA), Extensibility (Ext), stickiness (Stic), Stability (Stab), Consistency (Cons), Elasticity (Elas) and Creamy Colour (CC). This state is influenced by the consistency at the end of the pre-mixing operation ( $w$ ), by the target temperature at the end of mixing ( $x$ ), by the mixer setting and geometry, namely the difference in linear velocity between the arm and bowl ( $y$ ), and by the expected heat dissipated during texturing ( $z$ ). The selection of these variables and the technical terms is based on a glossary defining dough quality and bread baking, written in French by the same research group, and available on the Web (Roussel et al. 2010).

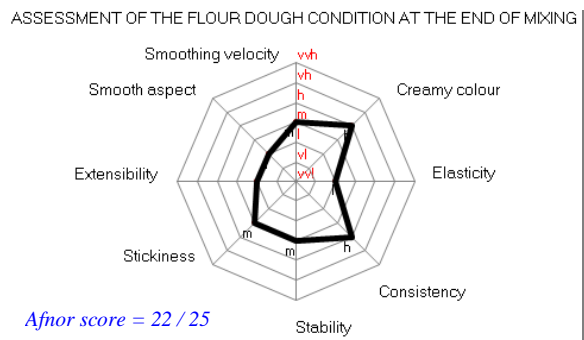
Figure 2 presents the set of qualitative functions forming the qualitative model of the texturing operation which allows to compute the state of a mixed dough from the initial consistency of the dough and from the operating conditions. There is a total of 8 functions, that is one function for each state variable of the model.

$$\begin{aligned}
 SV &\approx (1 \otimes w) \oplus \alpha(y, z) \\
 SA &\approx T((1 \otimes w) \oplus (1 \otimes \perp(x)) \oplus T(z) \oplus (1 \otimes \perp(z))) \\
 Ext &\approx (1 \otimes \perp(w)) \oplus (1 \otimes \perp(x)) \oplus T(z) \oplus (1 \otimes \perp(z)) \\
 Stic &\approx \perp((1 \otimes w) \oplus \perp(\perp(x) \oplus z)) \\
 Stab &\approx \perp((1 \otimes w) \oplus \perp(x) \oplus z) \\
 Cons &\approx w \oplus (1 \otimes \perp(x)) \oplus (1 \otimes z) \\
 Elas &\approx (1 \otimes \perp(w)) \oplus x \oplus T(z) \\
 CC &\approx (1 \otimes z)
 \end{aligned}$$

Figure 2. Qualitative functions of dough state descriptors after mixing (Ndiaye et al. 2009)

Such models are implemented in the KBS using the *Qualis*® expert-system shell and the outputs are displayed to the users as plain text using the domain vocabulary, or as radar charts (Fig.3).

(a)



(b)

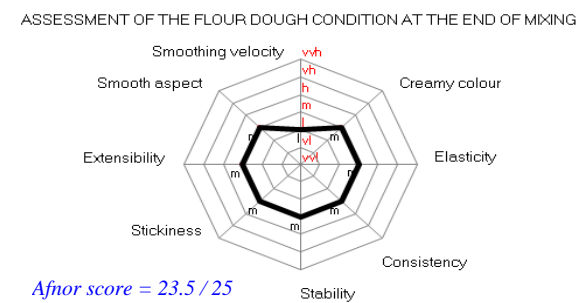


Figure 3. Examples of the KBS outputs: predictions of dough state at the end of texturing, starting from a standard consistency at the end of first mixing ( $350 \leq w \leq 450$  UB), a standard target temperature ( $22 \leq x \leq 25$  °C), an average difference of velocity  $y$  and a heat dissipation  $z$  (a) medium and (b) low. Scale ranges from vvh (very excessive) to vvl (very insufficient).

### 6.2. Validation

The validation of the knowledge base consists in comparing predictions given by experts with those of the KBS. When comparing distinct evaluations of the same food product, it is common, and experts do so, to take into account a sensibility level to distinguish acceptable divergences resulting from different sensibilities, from significant divergences (Allais et al. 2007). The threshold to do the distinction depends on the domain; in the case of breadmaking, the acceptable gap between two evaluations is of one level on the scoring scale of 7 levels (Tab. 1), e.g. very insufficient

and insufficient or normal and slightly excessive are two acceptable divergences.

The qualitative model of the texturing operation covers 81 solutions that ought to be valid in a normal context of production; in other words, unreasonable mixing settings were not considered by the experts during the knowledge elicitation phase. The set of solutions was checked over exhaustively with the experts. Only one significant disagreement between the KBS and experts has been identified and reported in Ndiaye et al. (2009) giving a rate of 98.8% of acceptable predictions.

The divergence comes from the prediction of the smoothing velocity, for high self-heating and high difference of linear velocity. The experts found difficult to refine the knowledge about the difference of linear velocity which reflects an incomplete or tacit knowledge about the influence of this variable on the dough condition. As a matter of fact, the linear velocity difference influences only the smoothing velocity and, rather oddly, has no effect on the other state variables (Fig. 2). Experts predict the influence of the kneader based mostly on the dough self-heating, which reflects the amount of energy transmitted to the dough. They initially came up with the difference of linear velocity to integrate the influence of the kinematics of kneader on the dough properties; however, in a normal production context the influence of this variable turned out to be difficult to uncouple from the energy.

## 7. PREDICTIONS AGAINST EXPERIMENTS

We performed experiments to investigate the influence of the mixing setting on the dough condition. As said section 6.2, the incomplete integration of the difference of linear velocity reflects a lack of knowledge. Thus, experiments were designed in a research context to investigate the effect of extreme mixing conditions on the dough by combining the duration with the rotation speed of the mixer's arm. To reveal the influence of the difference of linear velocity, independently from the one of energy, mixing was conducted with a high-speed spiral kneader instead of the more commonly used low-speed oblique-axis kneader. Experimental design is presented in figure 4.

The two technologists involved in the project performed sensory measurements of the dough condition for the nine trials. Measurements were limited to the six sensory descriptors of the normalised assessment grid (Fig.1): Smooth Aspect (SA), Stickiness (Stic), Consistency end of the mixing (Cons), Extensibility (Ext), Elasticity (Elas), Stability (Stab). According to the model, the difference of linear velocity should not be involved in the prediction of these descriptors (Fig. 2).

The experiment settings were converted into input values for the four control variables in the KBS. The composition of the dough and the pre-mixing operation were standard and the same for the nine trials so that the consistency of the dough at the beginning of the mixing was always normal: Cbm  $\approx$  m. Measurements of the

dough temperature at the beginning and at the end of the mixing enabled to determine the dough selfheating as input in the KBS.

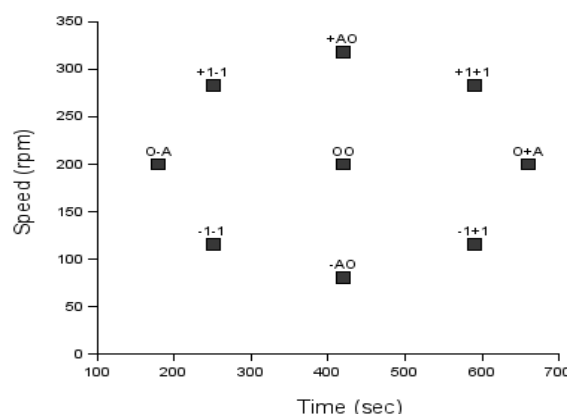


Figure 4. Repartition of the nine trials in the 2 dimension: space speed (rpm= rev per min) x texturing duration.

The results revealed three profiles of dough conditions corresponding to normal, low and high energy mixing (Fig. 5). The overall rate of acceptable predictions of the KBS is 88.9%, as illustrated Fig. 5a and b, predictions of the KBS for normal and low energy mixing is acceptable but significant disagreement comes from the case high speed and long duration (+1,+1) (Fig. 5c). Although four descriptors are correctly predicted, the smooth aspect is significantly better than the one predicted by the KBS and the dough elasticity is insufficient while the KBS predicts a slightly excessive elasticity.

This tends to show that out of a standard production context, the influence of the couple difference of linear speed and self-heating is incompletely captured by the model. It is likely that, for a long duration of mixing, the high speed mixing-arm impacts negatively the dough elasticity, while the smooth aspect remains correct. This is because spiral kneader applies much higher shear velocity to the dough than oblique-axis kneader do, resulting in a phenomenon of overmixing i.e. degradation of the dough protein network. Shear velocity is captured in the model by the difference of linear velocity, so better integration of this variable will extend the domain of validity of the KBS and incidentally the domain knowledge.

The physical relationship between the nature of the stress applied by the kneader and the dough properties is yet to be clarified (Stauffer 2007); the KBS was used as a support to point out the lack of knowledge; it also helped to perform well-focused experiments to characterise the relations between the dough behaviour and the mixing setting. The causes of the dough behaviour are now to be uncovered by the domain rheology.

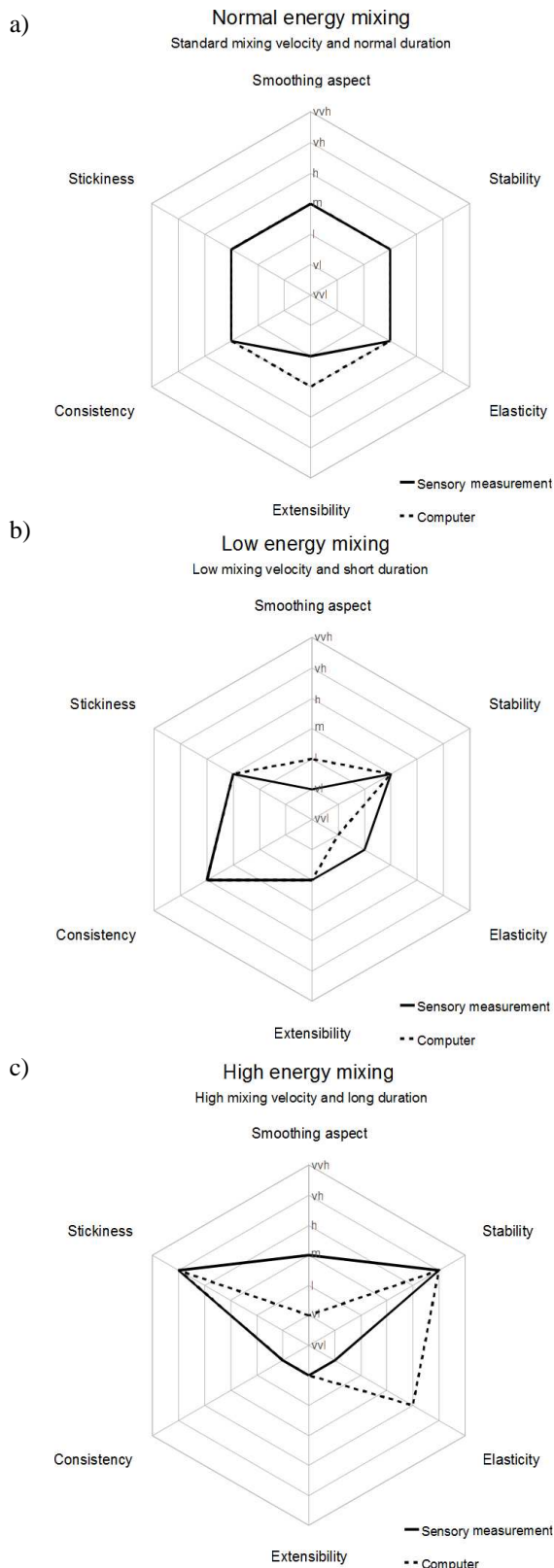


Figure 5. Confrontation between sensory measurements and predictions of the KBS for a) a standard mixing (trial 0,0), b) a low energy speed mixing (trial -1,-1) and c) high energy mixing (trial +1,+1).

## 8. DISCUSSION

We present a procedure to represent expert knowledge about breadmaking operations through qualitative functions. The value of each output variable is given by a unique function that integrates all the variables having an influence (e.g. Fig. 2); therefore, for each output variable it exists one and only one way to compute its value. This mode of representation, greatly facilitated by the use of the Q-Algebra, allows to avoid the problems of inconsistency of the knowledge base that can occur with other knowledge representations. To update the calculation method of one variable in the KB without jeopardizing the consistency of this last, it is necessary and sufficient to revise the algebraic expression of the concerned qualitative function.

The procedure, illustrated in section 5, reflects the majority of the situations we encountered so far. However two problems may occur: i/ a decision table given by an expert can only be represented by a complex qualitative function providing neither a concise expression of experts knowledge nor calculation benefits, in this case the decision table is used as such in the calculation process, ii/ the number of decision tables to be filled in is intractable for human experts because the number of combinations of variables is too large, in this case we incrementally prospect with the human experts the way they combine the variables. For instance, the prediction of the condition of the dough resulting from the pre-mixing operation requires the combination of 18 variables that gives about  $10^{11}$  distinct combinations. To model this operation, we first implemented a simple qualitative addition of all the calculated variables reflecting the individual influences of the 18 input variables and then refined the model after each validation turn from the experts comments and corrections (Kansou et al. 2008). This knowledge elicitation technique takes advantage of the functional representation of the knowledge presented in this paper.

## 9. CONCLUSION

A KBS for the prediction of the mixed dough condition has been built on the basis of the human experts' know-how for technologists and scientists in the French breadmaking domain. A knowledge representation procedure based on a qualitative algebra to represent the knowledge through qualitative functions has been presented. With the procedure, the calculation of each variable used in the KBS is performed using a unique qualitative function, eliminating, *de facto*, the risk of having inconsistencies in the KB.

Finally, the elicitation and representation of the expert knowledge lead to the identification of lack of knowledge. Confrontation of the KBS predictions with the real observations of the dough condition is also a way to refine the knowledge and to reduce the lack of knowledge in the domain.

In the near future, the qualitative models should also be used to support decision by determining the best settings of the mixing operation to make a dough with the desired properties.

## ACKNOWLEDGMENTS

This work has been developed in the frame of the ANR-Alia INCALIN and the authors are particularly thankful to the experts, Hubert Chiron (INRA-BIA) and Philippe Roussel (Univ. Paris VI)

## REFERENCES

- Van de Ven, A.H. Johnson, P.E., 2006. Knowledge for theory and practice. *Academy of Management Review*, 31 (4) 802-821.
- Girard, N., Navarrete, M., 2005. Quelles synergies entre connaissances scientifiques et empiriques ? L'exemple des cultures du safran et de la truffe. *Natures Sciences Sociétés*, 13, 33-44.
- Nonaka, I., Takeuchi, H., 1995. *The Knowledge-Creating Company*. New York, NY: Oxford University Press.
- Kansou, K., Della Valle, G., Ndiaye, A., 2008. Qualitative modelling to prospect expert's reasoning. In: Cesta, A., Fakotakis, N., eds, *Proceedings of the 4th Starting AI Researchers' Symposium (STAIRS'08)*, Frontiers in Artificial Intelligence and Application, 179, IOS Press, 94-105.
- Ndiaye, A., Della Valle, G., Roussel, P., 2009. Qualitative modelling of a multi-step process: The case of French breadmaking, *Expert Systems with Applications*, 36 (2, Part 1) 1020-1038.
- Stauffer, C.E., 2007. Principles of Dough Formation. Second Edition. In: Cauvain, S.P., Young, L.S., eds. *Technology of Breadmaking*, US: Springer, 299-332.
- Bloksma, A.H., 1990. Dough structure, dough rheology and baking quality, *Cereal Foods World* 35, 237-244.
- NF V03-716. *Farines de blé tendre (Triticum aestivum L.) - Essai de panification de type pain courant français*. AFNOR, Feb, 2002.
- Dieng, R., Corby, O., Lapalut, S., 1995. Acquisition and exploitation of gradual knowledge, *International Journal of Human-Computer Studies*, 42 (5), 465-499.
- Guerrin, F. (1995). Dualistic Algebra for Qualitative Analysis. In: *Proceedings of the 9<sup>th</sup> International Workshop on Qualitative Reasoning*, 64-73. May 1995, Amsterdam, Netherlands.
- Allais, I., Perrot, N., Curt, C., Trystram, G., 2007. Modelling the operator know-how to control sensory quality in traditional processes. *Journal of Food Engineering*, 83(2), 156-166.
- Roussel, P., Chiron, H., Della Valle, G., Ndiaye, A., 2010. *Glossaire terminologique appliqué aux pains français*. INRA CEPIA. Available from: <http://www4.inra.fr/cepia/Editions/glossaire-pains-francais> [June 2011].

# SIMULATION BASED MODELING OF WAREHOUSING OPERATIONS IN ENGINEERING EDUCATION BASED ON AN AXIOMATIC DESIGN

Rafael S Gutierrez<sup>(a)</sup>, Sergio Flores<sup>(b)</sup>, Fernando Tovia<sup>(c)</sup>, Olga Valerio<sup>(d)</sup>, Mariano Olmos<sup>(e)</sup>

<sup>(a)</sup> The University of Texas at El Paso

<sup>(b)</sup> Universidad Autonoma de Cd. Juarez

<sup>(c)</sup> Philadelphia University

<sup>(d)</sup> El Paso Community College

<sup>(e)</sup> El Paso Community College

<sup>(a)</sup> [rsgutier@utep.edu](mailto:rsgutier@utep.edu),

## ABSTRACT

Many engineering students have difficulties to resolve real life problems through a traditional instruction. Most of them do not apply the fundamental science-math knowledge to construct a functional understanding. A mathematical modeling learning approach named *Axiomatic Design* represents a didactical alternative to achieve not only the scientific skills but also the ability toward the design, creativity and innovation of engineering processes based on an adaptive expertise for learning using mathematics and physics principles. We present an axiomatic design application in the context of a block stacking situation and the corresponding learning outcomes.

**Keywords:** axiomatic design, mathematical modeling, space utilization, geometric reasoning

## 1. INTRODUCTION

Most of students have learning difficulties with science-math fundamental concepts in the engineering courses. A possible reason of this understanding problem is the traditional instruction. Through this educational approach, many students do not develop a functional understanding Flores et al. (2004) of the basic scientific knowledge they need to succeed in their engineering courses and a professional life. This kind of instruction is characterized by the use of: 1) textbook-like exercises, Flores.(2006) 2) a few of didactical strategies, 3) an

unidirectional teacher-student communication system, and 4) lectured-based class sessions. Although several efforts have been made to improve engineering understanding through a technology-base instructional modification, they have not been enough to develop a students' integral education. In the sense, we claim the necessity to include a science-math industry-base projects learning alternative to improve engineering students' academic preparation.

Ohland et al. (2004) report that among engineering programs, there is a consensus that mathematics (Calculus I and II) is the largest obstacle causing dropout in the freshman year.<sup>3</sup> To overcome the burden of mathematics and science in the freshman engineering curriculum, integration of science courses with engineering, and project- or problem-based teaching has proven effective in helping students to overcome these barriers (Bernold wt al. (2000) , Dichter (2001), and Dym et al. 2005). In addition, Prince and Felder (2006) state that learner-centered teaching methods (inductive teaching) supported by research findings have shown that students learn by fitting new information into existing knowledge structures that are unlikely to learn if they do not make the connections to what they already know and believe Gutierrez(2003).

In this article, we describe the General Framework of Warehousing and how we may teach warehousing operations using White et al. (1981), a typical sequence of events is to move unit loads from a packaging station, which and how we may teach warehousing operations, a typical sequence of events is to move unit loads from a packaging station, which has followed production operations, to temporary finished goods rack storage. Then, upon demand, the unit

loads are retrieved and moved from storage to shipping. This paper focuses on how the objectives of participating students and academic institutions are to provide the engineering students with a new pedagogical approach to understand and apply mathematics and science as a cognitive tool in the context of Engineering. This approach increases the recruitment of community college students and university undergraduates toward engineering with the required skills to succeed and graduate.

Section 1 describes the general framework of supply chain using axiomatic design, defines the general domains-of-use design problem. Section 2 explains how to maximize warehouse space when retrofitting is not an option.

In Section 3, we review the most relevant literature, definitions. Section 4 presents a general warehousing design model formulation. The development of block stocking model is describe in section 5. Finally section 5 provide recommendations for block stocking.

## 2. GENERAL WAREHOUSE DESIGN PROBLEM AND DESIGN WORLD DOMAINS

### 2.1 Axiomatic Design Framework

Figure 1 provides a conceptual view of the general design problem for warehousing operations at the warehouse level that is compatible with an integrated supply chain strategy. The management of warehouse resources planning with incoming customer orders are relied on the knowledge of warehouse experts. There are two design applications for warehousing operations: (1) retrofitting of existing facilities and (2) construction of new facilities. Retrofit is the more typical and challenging case.

**The main thrust of the paper is to describe a general framework of warehousing on helping students to develop skills and an adaptive expertise for learning using mathematics and physics principles.**

The basic postulate is that there are two fundamental axioms that are always present in good design such as product, process, or systems design. *The first axiom is called the independence axiom, which states that the independence of functional requirements (FRs) must always be maintained*, where FRs are defined as the minimum number of independent

requirements that characterize the design goals.

***The second axiom is called the information axiom, which states that, among those designs that satisfy the independence axiom, the design with the highest probability of functional success is the best design.***

The design world of the axiomatic approach is made up of four domains: customer domain, functional domain, physical domain, and process domain. The domain structure is illustrated in Figure 1. All design tasks are contained in these four domains. For example in figure 1, in the case of warehousing systems, customer attributes (CAs) may be the attributes desired by all customers; functional requirements (FRs) may be flexibility, efficiency and controllability; design parameters (DPs) may be the layouts and design of the supply chain elements themselves as composed of physical elements; and process variables (PVs) may be machine tools, people and material handling and so on. The capacitated location/allocation block on the right represents the proposed design solution of how we choose to satisfy the requirements specified in the left block.

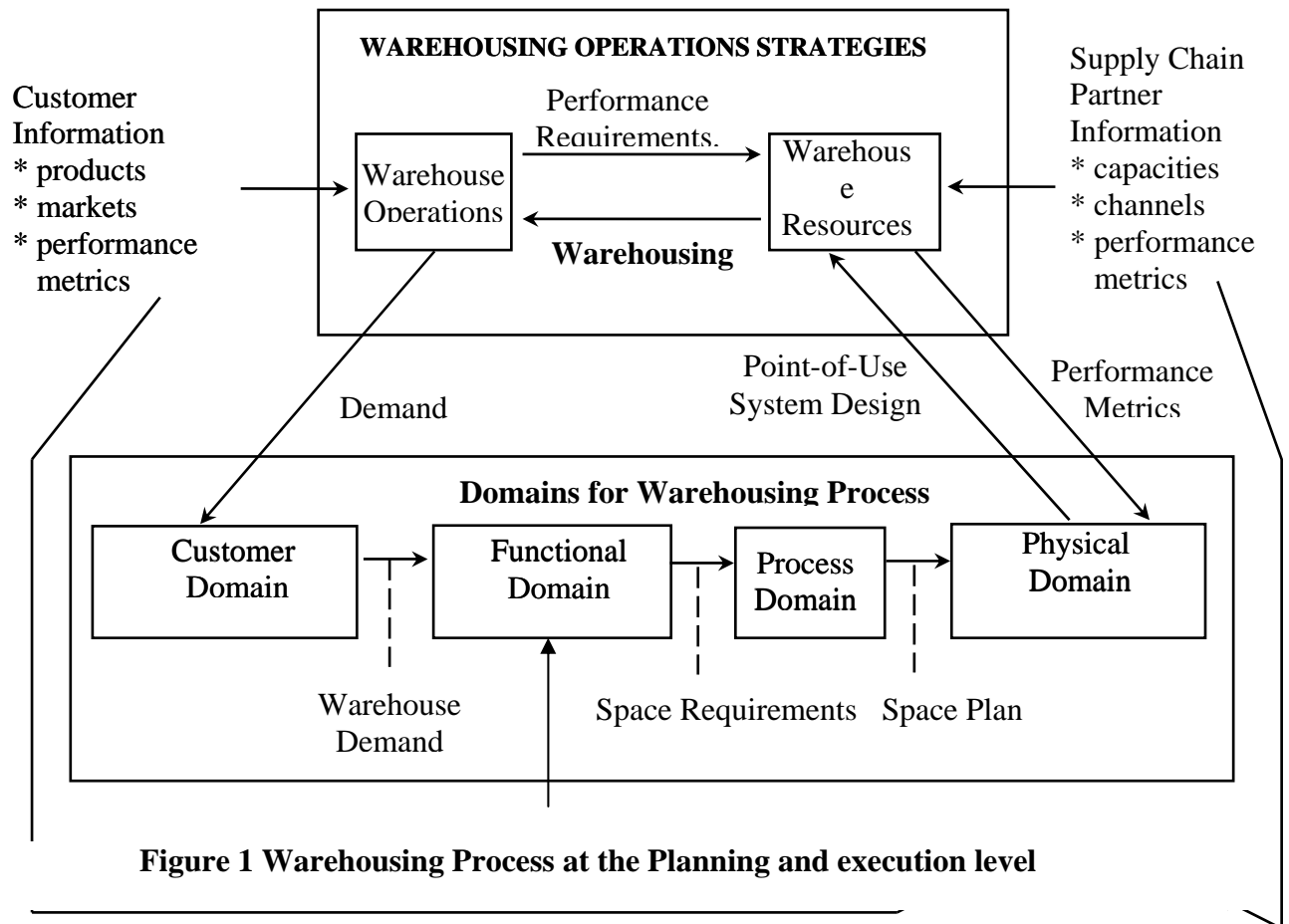
It is mandatory for matching functional requirements with parameter design. It is mandatory to explain the axiomatic design procedure for definition and design information required to matching functional requirements with parameter design and improve the satisfaction of the original need through the evaluation of the information content.

### 1.1 Physics –Mathematical Modeling Cycle

Physics – Mathematical Modeling Cycle provides useful information of our model describing how we use mathematics in physical systems as it is shown in figure 3. Math modeling cycles are applied between the different domains we described before in the figure 1 as mapping and data. Our process of math modeling cycles begin in the lower left corner by choosing a physical system we want to describe.

It is mandatory for matching functional requirements with parameter design and also It is mandatory to explain the axiomatic design procedure for definition and design information with parameter design and improve the outcome.





**Figure 1 Warehousing Process at the Planning and execution level**

### 1.1 Physics –Mathematical Modeling Cycle

Physics – Mathematical Modeling Cycle provides useful information of our model describing how we use mathematics in physical systems. Math modeling cycles begin in the lower left corner by choosing a physical system we want to describe. Within this box, we have to decide what characteristics of the system to pay attention to and what to ignore once. We have decided what we need to consider, we do step 1 called mapping. We map our physical structures.

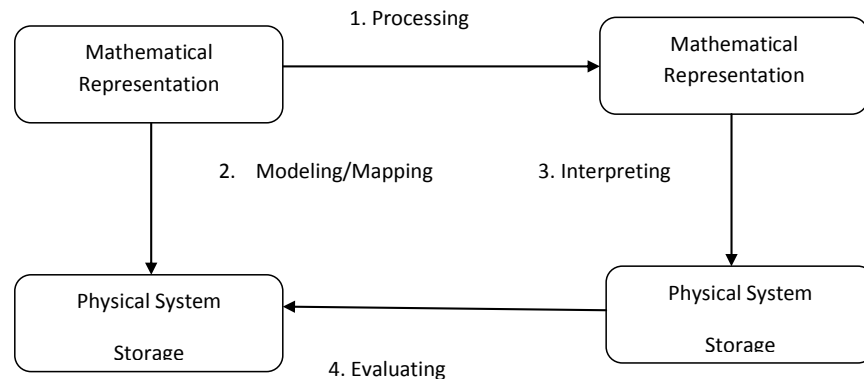
According to Redish (2005) – Mathematical Modeling Cycle provides useful information of our model describing how we use mathematics in physical systems as it is shown in figure 3. Math modeling cycles are applied between the different domains we described before in the figure 1 as mapping and data. Our process of math modeling cycles begin in the lower left corner by choosing a physical system we want to describe.

We have decided what we need to consider, we do step 1 called mapping. We map our physical structures into mathematical ones. To do this, we have to understand what mathematical structures are available and what aspects of them are relevant to the physical characteristics we are trying to model following this modeling process. After we have mathematized our system, we are ready for step 2 called process. At this level, we may solve an equation or deriving new ones. We still have to do step 3 called interpretation. We see what our resources tell us about our system in physical terms and then do step 4 called evaluation. We have to evaluate whether our results adequately describe our physical system or whether we have to modify our model. Our traditional approach does not help our students focus on some of these important steps. We tend to provide our students with the model ready made, and we rarely ask to our students to interpret their results and even less often ask them to evaluate whether we have to modify our model.

**1.1 Physics** After we have mathematized our system, we are ready for step 2 called process. At this level, we may solve an equation or deriving new ones. We still

have to do step 3 called interpretation. We see what our resources tell us about our system in physical terms and then do step 4 called evaluation. We have to evaluate whether our results adequately describe our physical system or whether we have to modify our model.

Our traditional approach does not help our students focus on some of these important steps. We tend to provide our students with the model ready made, and we rarely ask to our students to interpret their results and even less often ask them to evaluate whether we have to modify our model.



**Figure 2 Evaluation process of the information content**

### When Retrofitting is not an option

A warehouse will normally run out of space due to rapid growth, seasonal peaks, large discount buying, planned inventory builds for manufacturing shutdowns, facility consolidation, or even a slow sales period.

Generally there are three types of space deficiencies that occur in a warehouse. The first type results from simply having too much of the right inventory. The second is the result of having too much of the wrong merchandise, and the third comes from using the existing warehouse space poorly.

We will focus on the space deficiency due to the poorly use of the existing warehouse space. This condition is usually caused by steady growth, changing storage requirements (change in product mix) and ever increasing service requirements. Poorly utilized space is a common occurrence that happens in all warehouses occasionally and is non-exclusive of the inventory type or storage conditions in the warehouse.

Traditionally, warehouses are built to

handle projected volumes , a set number of products and limited unit loads. Then they are expected to adjust to customer demands as well as be more efficient over time. To accomplish these conflicting goals, warehouses generally accept long-term penalties to accomplish short-term goals like creating customized floor-ready customer's merchandise at the piece level, or creating mixed loads to simplify customer processing when goods traditionally ship in full case or full pallet quantities. All of these customization steps take valuable floor space and labor from primary warehouse functions. Other common instances of poor space utilization include low vertical space utilization, wide aisles (over nine feet), and multiple products in single bin locations and/or partial unit loads being stored in full unit load locations. These types of problems should be addressed with physical layout and workstation design changes.

### 3.2 Warehouse Redesign

Most of professionals working in industry consider that the warehouse redesign process is more art than science and more common sense than theory. The primary objectives of warehouse redesign are to:

- Use space efficiently
- Allow for the most efficient material handling
- Provide the most economical storage in relation to costs of equipment, use of space, damage to material, handling labor and operational safety
- Provide maximum flexibility in order to meet changing storage and handling requirements
- Make the warehouse a model of good housekeeping
- Provide maximum flexibility in order to meet changing storage and handling requirements
- Make the warehouse a model of good housekeeping

Eight steps are required to make this happen:

- 1 Measure the space you have to work with
- 2 Define the fixed obstacles (columns, walls, doors, clearances, etc.)
- 3 Understand the product stored and handled
  - Define storage condition zones
  - Throughput/replenishment requirements
  - Unit handling loads
- 4 Establish the material flow paths
- 5 Determine auxiliary facility requirements (offices, dock staging, hold and inspection, etc.)
- 6 Generate alternatives
- 7 Evaluate alternatives
- 8 Recommend and implement improvements

All alternatives must consider not only space, but also material handling, and impacts on labor.

### 4.0 Introduction to the Mathematical Modeling for Block Stacking Design Problem: the case of finite production rate

#### 4.1 Warehousing stocking Definitions and processes

**Unit load** is a term that is well known, to persons involved with material handling mainly with warehouse activities. Perhaps the most common example of a unit load is an arrangement of cartons, stacked in layers, on a pallet. The typical sequence of events is to move unit loads from a packaging station, which has all unit loads produced White J.A (1981).

Block stacking involves the storage of unit loads in stacks within storage rows. It is one of the most frequently used when large quantities of a few products are to be stored and the product is stackable to some reasonable height without load crushing. Frequently unit loads are block stacked 3-high in rows that are 10 or more loads deep. The practice of block stacking is very common for food, beverages, appliances, and paper products, among others.

**Among the most important design questions, there is a fundamental question of how deep should the storage rows be,** block stacking is typically used to achieve high space utilization at a low investment cost. Hence, it is often the case that storage rows are used with depths of 15, 20, 30 or more.

During the storage and retrieval cycle of a product lot, vacancies can occur in a storage row. To achieve FIFO lot rotation, these vacant storage positions cannot be used for storage of other products or lots until all loads have been withdrawn from the row. The space losses resulting from unusable storage positions are referred to as “honeycomb loss or vertical losses”; block stacking suffers from both vertical and horizontal honeycomb loss. Figure 4 depicts space losses resulting from honeycombing.

The **design of the block stacking storage system** is characterized by: the depth of the storage row ( $x$ ), the number of storage rows required for a given product lot ( $y$ ), and the height of the stack ( $z$ ), where the decision variables,  $x$ ,  $y$ , and  $z$ , must be integer valued. If the high of the stack is fixed, the key decision variable is the depth of the storage row.

For a single product, factors which may influence the optimum row depth include the lot size, load dimensions, aisle widths, row clearances, allowable stacking heights,

storage/retrieval times, and the storage/retrieval distribution.

## 4.2 Initial steps for collecting drawings and data gathering

The goal of the Texas College and Career Readiness Standards (CCRS) is to establish what students must know and be able to succeed in entry-level courses offered at institutions of higher education Texas of Higher Education Coordinated Board (2010)

Using Geometric Reasoning will initiate to construct and use drawings to represent the block stacking modeling problem. Figure 4 describes the block stacking process as the storage of unit loads in stacks. The use of unit loads represents the basic element of our process such as pallets, boxes, etc. We describe palletization as an example because it is the most common and universally accepted technique of unitization. A pallet is simply a platform upon which material is placed to form a movable collection of items. Figure 5 shows the rack dimensions that are considered for the modeling of the block stocking analysis.

Let assume that a production order for Q units will arrive in the inventory system as a lot size of Q units. We will assumed that every unit load produced go directly to the warehouse. The factory is set up to produce a lot. The production rate is P units per hour and the demand rate is D units per hour. For simplicity we should assume that the production rate and the demand rate leave a fixed rate of unit loads in the warehouse of one unit for analysis only. Therefore unit loads of one unit are received in the warehouse every hour until the warehouse capacity is reached until the level of Q unit loads of inventory. During the periods when the factory is not producing these unit loads, there is a rate of outflow of one unit load until the inventory reaches a level of zero unit loads.

The basic assumptions of our analysis are: the storage is selected on random basis, the first in first out (FIFO) lot rotation is used for moving the unit loads in the warehouse, no re-warehousing of stock, and known lot sizes remain the basic assumptions of this study. The following additional assumptions are made to facilitate analysis of the product design problem.

1. The storage retrieval distribution for the product is characterized with a uniform production and withdrawal rate, no bulk

withdrawals, zero safety stock, and no lot splitting.

2. For simplicity, identical load dimensions and clearances (L,W, and C) are assumed for all products.  
However,  $Z_i$  represents the number of stacking levels for product i, is determined by the characteristics of the individual product.
3. The space required for cross-aisles and staging will not be considered in this analysis
4. The number of lanes of each depth to be provided in the block stacking design is exactly equal to the maximum number required. That is, storage lanes will be available so that each product can be stored immediately after its arrival and at its optimum depth.

$$S = \frac{y}{2Q - 1} (w + c)(0.5A + xl)(2Q - xyz + xz - 1)$$

Our decision variable is x because Z is restricted to the conditions of the ceiling height of the warehouse and the characteristics of the product in stock. Y is defined as  $\left\lceil \frac{Q}{xz} \right\rceil$ . Therefore, the optimization problem is:

## 4.3 .Block Stacking Model

### 4.3.1 Notation

S : Average floor area required for block stacking

X: Storage lane depth (integer number of unit loads)

Y: Maximum number of storage lanes required

Z: number of storage tiers

Q: Lot size (integer number of unit loads)

L: length or depth of the unit load

W: Width of the unit load

C: clearance between lanes

A: aisle width

We shall assume that  $Y = \left\lceil \frac{Q}{xZ} \right\rceil$

Where  $\lceil \cdot \rceil$  the smallest integer than or equal to  $\lceil \cdot \rceil$

#### 4.3.2 The development of a block stacking model

The model considered is shown in figures 4 and 5. The characteristics of this distribution are:

1. Uniform storage and withdrawal rate
2. A storage and withdrawal size of one unit load.
3. Safety stock equals zero.
4. All storage lanes for a product lot are restricted to be of equal depth

The average floor space required for block stacking over the life of the product is given by the following equation,

$$S = \frac{y}{2Q-1} (w+c)(0.5A + xl)(2Q - xyz + xz - 1)$$

Our decision variable is  $x$  because  $Z$  is restricted to the conditions of the ceiling height of the warehouse and the characteristics of the product in stock.  $Y$  is defined as  $\left\lceil \frac{Q}{xZ} \right\rceil$ . Therefore, the optimization problem is:

*Minimize*

$$S = \frac{y}{2Q-1} (w+c)(0.5A + xl)(2Q - xyz + xz - 1)$$

$$s.t. xyz \geq Q$$

$$x, y, z \text{ are integers}$$

Using geometric reasoning, our students make connections between geometry, statistics and probability. We compute probabilities using as a

Table 1 is shown an example of how lane and space requirements vary over the entire life of the product, and it is shown the space utilization for the same example.

The space utilization formula for a given inventory level is based on the following expression.

$$Ut = \frac{100I_t WL}{Z(ww + C)(0.5A + xL)}$$

#### 5.0 Determination of space requirements as a main function in the analysis using Algebraic Reasoning

Review of single product model assumptions the major step in the analysis of block stacking designs for one product is the specification of aggregate floor space requirements. For a single product  $k$ , the optimum design was determinate by minimizing  $S_k$  the average floor space required during the life of the product lot. The specification of average space requirements followed from the assumption that a storage lane is charged to an individual product lot until all units of the lot have been withdrawn.

#### 5.1 The development of a block stacking model

The model considered is shown in figure 1. The characteristics of this distribution are:

- uniform storage and withdrawal rate
- a storage and withdrawal size of one unit load
- safety stock equals zero
- all storage lanes for a product lot are restricted to be of equal depth

reference the number of events (possible solutions) and we scatter plot and use it to make predictions such as probability of having 6 or

more pallets in the stockroom. Likewise students and instructor should measure, compute and use measures of spread to describe data. Assume linear approximation in this analysis and draw floor space utilization for this product and answer the following questions,

How the space utilization that rises when a lane is receiving items for stock, and decreases whenever a new storage lane is beginning to get used?

How the space utilization rises whenever a storage lane is freed and decreases as honeycombing occurs?

storage lane; the lane is available for storage of any other product or lot.

Using geometric reasoning, we apply right relationships for definition purposes of the definition of average space requirements for an individual product k implicitly assumes that a storage lane will be immediately required by or charged to our product as soon as all units of product k have been withdrawal

Does space utilization rises when a lane is receiving items from stock, and decreases whenever a new storage lane is beginning to use?

We have the opportunity of computing the empirical probability of an event and its complement.

**Table 1. Total number of feasible system states**

Inventory Level	Number of lanes required	Total of space required	Probability of inventory level	Contribution to average space	Percentage utilization
1	1	62.111	1/17	3.653	11.74
2	1	62.111	1/17	3.653	23.48
3	1	62.111	1/17	3.653	35.22
4	1	62.111	1/17	3.653	46.96
5	2	124.222	1/17	7.307	29.35
6	2	124.222	1/17	7.307	35.22
7	2	124.222	1/17	7.307	41.09
8	2	124.222	1/17	7.307	46.96
9	3	186.333	1/17	10.960	35.22
8	2	124.222	1/17	7.307	46.96
7	2	124.222	1/17	7.307	41.09
6	2	124.222	1/17	7.307	35.22
5	2	124.222	1/17	7.307	29.35
4	1	62.111	1/17	3.653	46.96
3	1	62.111	1/17	3.653	35.22
2	1	62.111	1/17	3.653	23.48
1	1	62.111	1/17	3.653	11.74
Average number of lanes		27/17	1.5882		
Average space		98.640			



We have the opportunity of computing the empirical probability of an event and its complement.

#### Solutions obtained with the model

As we observe the model obtained is nonlinear with integer decision variables. The table 1 is shown an example of the model obtained. The solutions are shown that the function  $S$  is non-convex with respect to  $X$  therefore, it is required to enumerate over values of  $X$  in order to identify the global minimum. In the example it was assumed  $X_{\max}$  equals to 15. The optimal value is  $X$  equals 8. It was assumed  $Z$  equals 3, lot size equals 147 and  $Y$  equals 7.

#### 5.2 the development of a block stacking model

The model considered is shown in figure 1. The characteristics of this distribution are:

- uniform storage and withdrawal rate
- a storage and withdrawal size of one unit load
- safety stock equals zero
- all storage lanes for a product lot are restricted to be of equal depth

In summary a block stacking representation has been developed and the independence axiom has been applied to it. This application of the independence axiom has led to the establishment of an uncoupled design for the example of block stacking. The definition of the FRs and DPs has led to the identification of two constraints relating to storage retrieval process. An application of the information axiom is now required to allow the alternative supply chain structures to be evaluated further.

#### 6.0 Conclusions

The axiomatic design is a relevant instructional tool. Through this learning methodology students can understand the fundamental mathematical modeling steps and the corresponding skills in the context of several engineering situations. It represents an opportunity to apply mathematical concepts in real-life situations. Moreover, it is a high cognitive contribution to develop design-

innovative process understanding for both, introductory and advanced engineering courses. In addition, many learning engineering situations can be adapted to a lab-based axiomatic design instructional micro-curriculum. The major contribution of this paper is the understanding of axiomatic design block stacking design problem as a math-context tool for the development of guidelines for industry to design inventory systems along the supply chain network. Finally, we will conduct an investigation based on the analysis and categorization of the students' understanding effectiveness of this proposal in introductory engineering courses.

#### ACKNOWLEDGMENTS

Our sincere appreciation to Dr. Jessica O. Matson for her advice and support

#### REFERENCES

- 1.- S. Flores, S. Kanim and H. Kautz, "Students use of vectors in introductory mechanics", *Am. J. Phys.* **72** (4), 460-468, (2004).
- 2.- S. Flores, "Students use of vectors in mechanics", PhD. dissertation, Department of Physics, New Mexico State University, 2006.
- 3.- Ohland, M. A., Yuhasz, A. G., and Sill, B. I., "Identifying and Removing Calculus Prerequisite as a Bottleneck in Clemson's General Engineering Curriculum," *Journal of Engineering Education*, Vol. 93, No. 3, 2004, pp. 253-257.
- 4.- Bernold, L. E., Bingham, W. I., McDonald P. H., and Attia, T. M., "Influence of Learning Type Oriented Teaching on Academic Success of Engineering Students", *Journal of Engineering Education*, Vol. 89, No. 1, 2000, pp. 191-199.
- 5.- Dichter, A. K., "Effective Teaching and Learning in Higher Education with Particular Reference to the Undergraduate Education of Professional Engineers," *International Journal of Engineering Education*, Vol. 17, No 1., 2001, pp. 24-29.
- 6.- Dym, C. I., Agogino, A. M., Eris, O., Frey, D. D. and Leifer L. J., "Engineering Design

Thinking, Teaching, and Learning,” *Journal of Engineering Education*, Vol. 94, No. 1, 2005, pp. 103-120.

7.-Gutierrez, Rafael S., “Experiences in the Use of Integrative Manufacturing Software in Closing Professional Gaps in El Paso/ Juarez Industry,” in *INEER Special Volume for 2003 to the titled: “Engineering, Education, and Research- 2002: A Chronicle of Worldwide Innovations,”* (Eds.: W. Aung, M. Hoffmann, R. King, W.J. Ng, and Luis M. Sanchez), Chapter 25, pps. 247-255, 2003. ISBN 0-9741252-0-2.

8. Gutierrez, Rafael S., “Mathematical Models for the Optimization of Storage System Design Problem. The Case of Finite Production Rate, Working Paper, 2009

9. White J.A, Demars, N.A. and Jessica O. Matson, Optimizing Storage System Selection, 4<sup>th</sup> International Conference on Automation in Warehousing, Tokyo Japan September 30-October 2, 1981

10.Nam P. Suh, *The Principles of Design* (New York: Oxford University Press, 1990).

11. Prince, M., Felder, R., “Inductive Teaching and Learning Methods, Definitions, Comparisons, and Research Bases,” *The Journal of Engineering Education*, Vol. 95, No. 2, 2006, 123-138.

12. Redish, E.F., “ Problem solving and the use of math in physics courses”, *Proceedings of the International Conference on Physics Education in 2005, Focusing on Change, Delhi, August 21-26, 2005*

13. Texas Higher Education Coordinating Board, “Texas College and Career Readiness Standards,” Texas- HECB

# A MULTI-ITEM MULTI-RACK APPROACH FOR DESIGNING LIFO STORAGE SYSTEMS: A CASE STUDY FROM THE FOOD INDUSTRY

Elisa Gebennini, Andrea Grassi, Bianca Rimini, Rita Gamberini

Department of Science and Engineering Methods  
University of Modena and Reggio Emilia  
Via Amendola 2 – Pad. Morselli, 42100, Reggio Emilia – Italy

[elisa.gebennini@unimore.it](mailto:elisa.gebennini@unimore.it), [andrea.grassi@unimore.it](mailto:andrea.grassi@unimore.it), [bianca.rimini@unimore.it](mailto:bianca.rimini@unimore.it), [rita.gamberini@unimore.it](mailto:rita.gamberini@unimore.it),

## ABSTRACT

In a variety of industrial sectors rack storage is adopted for holding stock-keeping units (SKUs) between production (or purchasing from external supplier) and delivery. It is well known that, among the different types of rack storage, racks accessed in a Last-In-First-Out manner are the most economically convenient solutions. Nevertheless, especially when product shelf-life is critical as occurs in the food industry, the design of LIFO storage systems is not trivial. Thus, this paper presents an approach able to take into consideration two different measures for assessing the performance of a storage system, with the aim of assigning each item type in inventory to the optimal type of lane racks. Lastly, a significant case study from the food industry is discussed.

Keywords: LIFO storage, efficiency, optimal design, food industry

## 1. INTRODUCTION

The design of rack storage systems into a multi-item environment presents several challenges and opportunities. A first distinction that needs to be made is between racks accessed according to a First-In-First-Out (FIFO) policy and a Last-In-First-Out (LIFO) policy. The reader may refer to Bartholdi and Hackman (2006) for a review of the main advantages and disadvantages of the different types of rack storage.

In particular, LIFO racks are the most space-saving (i.e. economically convenient) solution but require periodic replenishing/emptying cycles in order to allow all the stock-keeping units (SKUs), which are not independently accessible, to be retrieved within a reasonable period of time. As a consequence, especially in industries such as food products, where items in inventory typically have critical shelf lives, the adoption of LIFO solutions requires a thorough analysis taking into account more than a single performance measure.

One of the most commonly used measures for assessing the performance of storage systems is *space utilization* (see, e.g., surveys such as Van Den Berg, 1999, and Gua et al., 2010). Briefly, space utilization refers to the amount of aisle space per pallet location.

The higher the capacity of a lane is, the higher the value of space utilization is (the same aisle space is spread over a larger number of pallet locations).

Recently, Ferrara et al. (2011) have stressed the importance of integrating such a traditional performance measure with a new indicator, called *storage efficiency*. Storage efficiency (see Ferrara et al., 2011) is a performance measure related to the number of pallet locations that are actually occupied by SKUs. Note that, once the first SKU of a certain item type has been placed into a LIFO lane, that lane is devoted to that item type until it becomes empty again. In other words, the other pallet locations, even if empty, are “constrained” (i.e. they can accept that specific item type only). As a consequence, storage efficiency for LIFO racks is typically less than 1. Moreover, as the number of “constrained” pallet locations arises, the adoption of such a solution becomes more critical. Nevertheless, works such as Ferrara et al. (2011) and Rimini et al. (2011) show that it is possible to maintain satisfactory values of the storage efficiency also in LIFO storage systems, if they are properly designed and operated.

Hence, the main contribution of this work paper is to integrate the two performance measures, i.e. space utilization and storage efficiency, into an innovative approach for designing optimal configurations of LIFO rack systems.

The remainder of the paper is organized as follows. Section 2 describes the problem under analysis. Section 3 presents the solution approach for assigning each item type in inventory to the optimal lane type. In Section 4 the solution approach is applied to a real case study from the food industry. Finally, Section 5 draws some conclusions.

## 2. PROBLEM DESCRIPTION

### 2.1. Objective

The focus of this paper is on storage systems where a set  $I$  of different item types (i.e. types of products presenting different input/output flows and physical features) is held as inventory in LIFO lane racks. Specifically, a set  $J$  of different lane types is assumed to be available, where lane types differ from each other in

their capacity (expressed in number of available pallet locations).

In this paper a case study from the food industry is presented. Since the focus is on perishable products, the capacity of the LIFO lane should not increase too much in order to avoid that any SKU spends an excessive time in inventory. Specifically, the height of all the lanes is supposed to be one pallet (pallets cannot be stacked one on top of each other). Nevertheless, for an efficient use of space, lane racks may have more than a single level. Let us denote with  $H$  the number of levels of independent LIFO lanes of the same type. As an example, Figure 1 shows an agglomeration of  $H=2$  levels of lanes with a capacity of 3 pallet locations each.

The objective of the study is to develop a solution approach for assigning each item type to the lane type that best suits the inventory requirements by taking into account both space utilization and storage efficiency.

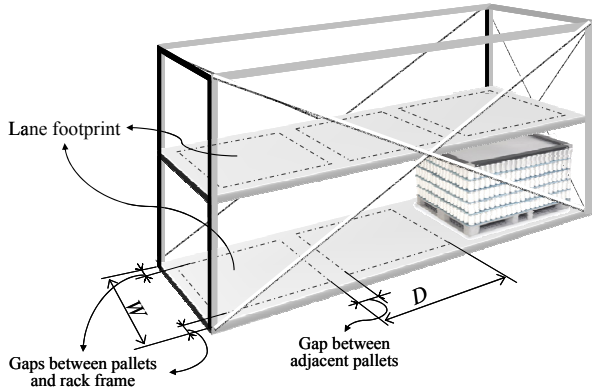


Figure 1: 2-level aggregation of 3-pallet deep lanes

## 2.2. Problem Data

Once the set  $I$  of item types and the set  $J$  of lane types has been identified, problem data must be collected or simulated when not available.

Typically, any Enterprise Resource Planning (ERP) system provides the firm with information about the input/output flows of each item type. Thus, even if actual data on inventory levels are not available, they can be easily simulated. We assume to consider a period of time composed of  $T$  time units (t.u.). Let us denote with  $q_{i0}$  the inventory level of item type  $i$  at the beginning of the first time unit  $t=1$ . Then, the inventory level of item type  $i$  at the end of t.u.  $t$  is as follows:

$$q_{it} = q_{i,t-1} + p_{it} - o_{it}, \quad \forall i, t. \quad (1)$$

where  $p_{it}$  is the input flow of  $i$  during  $t$  (coming from the production area or from external suppliers) and  $o_{it}$  is the output flow of  $i$  during  $t$  (going toward the shipping area).

As regards lane racks, let us denote with  $C_j$  the capacity of lane type  $j$ , where  $C_{j'} \neq C_{j''}$ ,  $\forall j', j'' \in J$ .

Note that, given a certain item type  $i$  and lane type  $j$  (i.e. of capacity  $C_j$ ), the average number of lanes of that type necessary for storing  $i$  is:

$$N_{ij} = \left\lceil \frac{\sum_t q_{it}}{C_j T} \right\rceil, \quad \forall i, j, \quad (2)$$

We assume that pallet dimensions are fixed, e.g. the dimensions of a standard pallet. In Figure 1, two 3-pallet deep lanes (i.e.  $C_j=3$ ) are shown, stacked on top of each other in a 2-level agglomeration. For each lane we denote as  $W$  the total pallet location width (pallet width + gaps) and as  $D$  the total pallet location depth (pallet depth + gaps). Thus, the width of any lane is  $W$  and the total depth of a lane type  $j$ , i.e. with a capacity  $C_j$ , is  $D C_j$ . As a result, the footprint  $F_j$  of a lane type  $j$  is as follows:

$$F_j = W D C_j, \quad (3)$$

The lane footprints are indicated by light-grey areas in Figure 1.

Since lanes are assumed to be stacked in agglomerations of  $H$  levels each (e.g.  $H=2$  in Figure 1), lanes of the same agglomeration are accessed from the same aisle. Hence, in order to compute the total amount of horizontal space required in the system, one-half the aisle width (denoted as  $A/2$ ) must be considered for each agglomeration of  $H$  lanes. Note that the aisle width  $A$  depends on the type of forklift trucks or Automated Guided Vehicles (AGVs) used in the storage system for pallet handling.

Given the input data set discussed in this section, the solution approach for allocating each item type to the optimal lane type is discussed in the sequel.

## 3. SOLUTION APPROACH

### 3.1. Storage Efficiency and Inefficiency Weight

According to the notation and conventions defined in the previous section, the storage efficiency related to lanes of capacity  $C_j$  housing  $q_{it}$  SKUs of item type  $i$  during time unit  $t$  (see Ferrara et al., 2011) is:

$$e_{ijt} = \frac{q_{it}}{C_j \left\lceil \frac{q_{it}}{C_j} \right\rceil}, \quad \forall i, j, t. \quad (4)$$

Thus, the average storage efficiency for item type  $i$  and lane type  $j$  during the whole observation period  $T$  can be expressed as follows:

$$e_{ij} = \frac{\sum_t e_{ijt}}{T}, \quad \forall i, j. \quad (5)$$

As explained in Ferrara et al. (2011), a threshold level  $e_j^T$  can be identified for the storage efficiency of LIFO storage racks of capacity  $C_j$ . The threshold level  $e_j^T$  is the minimum storage efficiency required to store

any item type into lanes of type  $j$ . It can be computed simply as the ratio between the space utilization of selective racks and the space utilization achievable with LIFO racks of capacity  $C_j$ . The comparison with selective racks is motivated by the trade-off between space utilization and storage efficiency. If  $C_j=1$ , i.e. in case of selective racks, the storage efficiency is always 1 for any item type, while the amount of aisle space required per pallet location arises. Thus, if both the performance measures are jointly considered, the efficiency threshold allows us to estimate the convenience of adopting lane racks instead of selective racks. Thus, given the storage efficiency computed according to Eq. (5), if  $e_{ij} \geq e_j^T$ , it is more convenient to stock item type  $i$  into lanes of capacity  $C_j > 1$ . Otherwise, the adoption of selective racks is suggested.

For the sake of convenience, we now define the inefficiency weight  $w_{ij}$  as the complement of the storage efficiency. So,

$$w_{ij} = 1 - e_{ij}, \quad \forall i, j. \quad (6)$$

This parameter is used in the mathematical formulation of the linear-programming model discussed in the next section.

### 3.2. Optimization Model

In this section a linear allocation model is presented in order to assign each item type to the optimal lane type. The objective is to minimize the total lane and floor space (including gaps between lanes and aisle space) required on average to store the SKUs of all the item types under analysis. Moreover, solutions presenting low values of the inefficiency weight, i.e. high values of the storage efficiency, are preferred.

Prior to the model formulation, the following additional assumptions must be made:

- The storage system is assumed to be uncapacitated (a large number of lanes of any type is available);
- Each item type can be assigned to a single lane type;
- Given a certain item type  $i$ , there exist at least one lane type  $j$  so that  $e_{ij} \geq e_j^T$ ;
- All the agglomerations of lanes have the same number of levels.

The integer linear-programming model is as follows:

$$\min \sum_{i \in I} \sum_{j \in J} F_j N_{ij} w_{ij} x_{ij} + W \frac{A}{2} \sum_{j \in J} z_j \quad (7)$$

subject to

$$\sum_{j \in J} x_{ij} = 1, \quad \forall i, \quad (7.1)$$

$$e_j^T x_{ij} \leq (1 - w_{ij}), \quad \forall i, j, \quad (7.2)$$

$$\frac{1}{H} \sum_{i \in I} N_{ij} x_{ij} \leq z_j, \quad \forall j, \quad (7.3)$$

$$z_j \text{ integer}, \quad (7.4)$$

$$x_{ij} \in \{0, 1\}, \quad (7.5)$$

where the input data are as follows:

$i \in I$	item type;
$j \in J$	lane type;
$F_j$	footprint of lane type $j$ according to Eq. (3);
$e_j^T$	threshold level for the storage efficiency of lane type $j$ (refer to Section 2);
$N_{ij}$	number of lanes type $j$ necessary on average for storing item type $i$ according to Eq. (2);
$w_{ij}$	inefficiency weight for the assignment item type $i$ into lane type $j$ (refer to Section 2);
$A$	aisle width;
$W$	lane width;
$H$	number of levels of a lane agglomeration;

The decision variable is  $x_{ij}$ . It is a binary variable that in the optimal solution is set to 1 if item type  $i$  is assigned to lane type  $j$ , 0 otherwise. Variable  $z_j$  allows the objective function to be linear: according to constraints (7.3) and (7.4),  $z_j$  assumes integer values only and represents the minimum number of agglomerations of  $H$  lanes of type  $j$ .

The objective function is composed of two contributions:

- The first term represents the total lane footprint for storing all the item types under analysis, weighted by the inefficiency weights  $w_{ij}$ . Hence, since the objective function defines the optimization problem as a minimization task, solutions with low values of the inefficiency weight are preferred;
- The second term represents the aisle space by considering a value equal to  $\left(W \frac{A}{2}\right)$  for each agglomeration of  $H$  lanes (recall the above definition of  $z_j$  and the correspondent constraints).

Constraint (7.1) states that each item type must be assigned to a single lane type. Constraint (7.2) guarantees that each item type  $i$  can be assigned to lane type  $j$  only if the corresponding storage efficiency, equal to the complement of  $w_{ij}$  according to Eq. (6), is higher than the threshold level as defined in Section 2. Constraints (7.3) and (7.4) regard the values assumed by  $z_j$ , as described above. Constraints (7.5) is the classical integrity constraint.

In the optimal solution, each item type is assigned to the optimal lane type. Thus, the number of lanes of each type that should be installed in the storage area is:

$$N_j^* = \sum_i N_{ij} x_{ij}^*, \quad \forall j, \quad (8)$$

where  $x_{ij}^*$  represent the values assumed by the decision variable in the optimal solution.

Similarly, the total storage efficiency for the whole storage system in the optimal solution can be computed as follows:

$$e^* = \frac{\sum_j \sum_i e_{ij} N_{ij} x_{ij}^*}{\sum_j \sum_i N_{ij} x_{ij}^*}. \quad (9)$$

#### 4. CASE STUDY

This case study has been done in collaboration with an Italian company producing glass and plastic containers for food products such as pasta sauces, pickled vegetables and marmalades.

##### 4.1. Input Data

Item types differ from each other in either the type of container or the type of product or both. Thus, the same sauce in a 314 ml container or in a 720 ml container corresponds to two different item type. Specifically, the set  $I$  under analysis includes 1 597 item types.

The ERP system provides the analyst with daily information about the input/output flows of all item types. Thus, a past period of 228 days has been considered and the inventory level for each item type at the end of each day within the observation period has been evaluated according to Eq. (1). As an example, Table 1 shows the inventory levels of 4 different item types over 5 days.

Table 1: Inventory levels for 3 item types over 5 days

Days	Item Types $i$			
	1	2	3	4
Day 1	18	18	18	13
Day 2	23	15	15	4
Day 3	32	32	16	12
Day 4	5	5	5	2
Day 5	3	3	0	0

The set  $J$  of lane types comprises 13 typologies of lane racks that can potentially be installed in the storage area. Each lane type  $j$  has a certain capacity  $C_j$ . Let us consider standard pallets, i.e. width of 1.2 m and depth of 0.8 m. Since the dimensions of a pallet location should include the gaps between SKUs and the rack frame (see Figure 1), the width of a pallet location  $W$  can be set to about 1.5 m and the depth  $D$  to 0.85 m. Thus, the footprint of any lane type  $j$  can be computed according to Eq. (2). Table 2 lists the available lane types along with their capacities (expressed in terms of pallet locations) and footprints (expressed in square meters).

Table 2: Set of lane types

Lane Types $j$	$C_j$ [# of pallet locations]	$F_j$ [m <sup>2</sup> ]
1	3	3 825
2	4	5 100
3	5	6 375
4	6	7 650
5	7	8 925
6	8	10 200
7	9	11 475
8	10	12 750
9	11	14 025
10	12	15 300
11	13	16 575
12	14	17 850
13	15	19 125

Lanes of the same types are stacked in agglomerates. For each agglomerate the number of levels (i.e. number of lanes) is  $H=6$ . In this case study an AGVs system is adopted for pallet handling. So, the aisle width  $A$  can be set to 3.6 m.

##### 4.2. Model Implementation and Results

Given the data set discussed above, the solution approach described in Section 3 can be applied to the case study.

Firstly, the threshold levels of the storage efficiency have been computed for each lane type according to Ferrara et al. (2011). The threshold levels are reported in the second column of Table 3.

Table 4: Storage Efficiency and Threshold levels

Lane Types $j$	Thres -hold $e_j^T$	Storage Efficiency $e_{ij}$			
		Item Types $i$			
		1	2	3	4
1	<b>99%</b>	99%	99%	91%	94%
2	<b>89%</b>	76%	92%	89%	97%
3	<b>83%</b>	87%	84%	76%	98%
4	<b>78%</b>	75%	94%	86%	85%
5	<b>75%</b>	64%	93%	81%	91%
6	<b>73%</b>	56%	81%	71%	83%
7	<b>72%</b>	93%	77%	63%	74%
8	<b>70%</b>	84%	69%	57%	98%
9	<b>69%</b>	80%	79%	87%	89%
10	<b>68%</b>	74%	90%	81%	82%
11	<b>67%</b>	68%	83%	87%	75%
12	<b>67%</b>	63%	77%	81%	70%
13	<b>66%</b>	59%	72%	76%	71%

Then, the storage efficiency  $e_{ij}$  of each item type  $i$  into lane type  $j$  can be obtained according to Eq. (5). For the sake of clarity, Table 4 shows the values of the storage efficiency for 4 item types if assigned to the lane types under analysis. As described in Section 3.1, only if  $e_{ij} \geq e_j^T$  it could be convenient to assign item type  $i$  to lane type  $j$ . As an example, if item type  $i=1$  is assigned to lane type  $j=3$  (i.e. 5-pallet deep lanes



according to Table) the corresponding storage efficiency is 87%. Since it is higher than the threshold level (83%), such an assignment is allowed. On the other hand, if the same item is assigned to lane type  $j=2$  (i.e. 4-pallet deep lanes) the storage efficiency is lower than the threshold level. As a consequence, if lane type  $j=2$  were the only one available in the system, it would be more convenient to store  $i$  into selective racks.

Since the liner-programming model described in Section 3.2 makes use of the parameter  $w_{ij}$  as defined by Eq. (6), in Table 5 the values of the storage efficiency are converted into inefficiency weights. If a certain assignment is not allowed (i.e.  $e_{ij} < e_j^T$ ) a dash (-) is entered in the corresponding cell. Recall that in our solution approach any assignment that does not satisfy  $e_{ij} \geq e_j^T$  is prevented by constraint (7.2).

Table 5: Inefficiency Weights

Lane Types $j$	Inefficiency Weights, $w_{ij}$			
	Item Types $i$			
	1	2	3	4
1	0.01	0.01	-	-
2	-	0.08	0.11	0.03
3	0.13	0.16	-	0.02
4	-	0.06	0.14	0.15
5	-	0.07	0.19	0.09
6	-	0.19	-	0.17
7	0.07	0.23	-	0.26
8	0.16	-	-	0.02
9	0.20	0.21	0.13	0.11
10	0.26	0.10	0.19	0.18
11	0.32	0.17	0.13	0.25
12	-	0.23	0.19	0.30
13	-	0.28	0.24	0.29

The linear-programming model was solved by using ILOG CPLEX 10.1 on a Pentium IV-3.2 GHz PC.

Once the optimal solution has been found, the number of necessary lanes of each type can be determined according to Eq. (8) as reported in Table 6.

The total amount of space required on average to store the SKUs and support the lane racks is 82 485 m<sup>2</sup> (78 894 m<sup>2</sup> of lane space + 3 591 m<sup>2</sup> of aisle space). Finally, according to Eq. (9) the total storage efficiency is 90,5%.

Since both the space requirements and the total storage efficiency are satisfactory and consistent with the company needs, this design solution can be adopted in practice.

Table 6: Number of lanes in the optimal solution

Lane Types $j$	# of necessary lanes
1	78
2	1 260
3	1 254
4	960
5	948
6	660
7	630
8	438
9	354
10	318
11	558
12	186
13	336

## 5. CONCLUSIONS

The paper addresses the design of LIFO storage systems. Particularly, a new solution approach able to assign each item type to the optimal lane type has been developed. The application to a significant case study from the food industry has been described and the ability of the method to produce a satisfactory solution has been proved.

## ACKNOWLEDGMENTS

The authors would like to thank to Eng. Andrea Ferrara for his support and assistance in data analysis.

## REFERENCES

- Bartholdi, J. & Hackman, S., 2006. *Warehouse & Distribution Science, Release 0.76*. Available from: <http://www.warehouse-science.com>.
- Van Den Berg, J.P., 1999. A literature survey on planning and control of warehousing systems. *IIE Transactions*, 31, 751-762.
- Gua, J., Goetschalckx, M., McGinnis, L.F., 2010. Research on warehouse design and performance evaluation: A comprehensive review. *European Journal of Operational Research*, 203, 539-549.
- Ferrara, A., Gebennini, E., Grassi, A., Rimini, B., 2011. Efficiency estimation of LIFO storage systems managed with automated guided vehicles (AGVs). *Proceedings of 21st International Conference on Production Research*, Germany, July 31-August 4, 2011, Stuttgart.
- Rimini, B., Grassi, A., Gamberini R., Gebennini, E., Lolli, F., Ferrara, A., 2011. Perishable item allocation in LIFO storage systems served by Automated Guided Vehicles. *Proceedings of First international workshop on Food Supply Chain*, Italy, June 26-July 1, 2011, Bertinoro (FC).

# A Process Mining Approach Based on Trace Alignment Information for Very Large Sequential Processes with Duplicated Activities

Pamela Viale

Laboratoire des Sciences de  
l'Information et des Systèmes  
Marseille, France

STMicroelectronics  
Rousset, France

pamela.viale@lisis.org

Claudia Frydman

Laboratoire des Sciences de  
l'Information et des Systèmes  
Marseille, France

CIFASIS, Centro Internacional  
Franco Argentino de Ciencias  
de la información y de  
Sistemas  
Rosario, Argentina

claudia.frydman@lisis.org

Jacques Pinaton

STMicroelectronics  
Rousset, France

jacques.pinaton@st.com

## ABSTRACT

This paper presents an approach to construct models for very large sequential processes with duplicated activities. Very large meaning processes containing hundreds of activities. The algorithm uses traces of the process we are interested in modeling and construct an alignment between them. The alignment obtained is then used to guide the model construction. We present also the application of our approach to STMicroelectronics' manufacturing processes. The constant modifications of semiconductor specifications forces engineers to continuously update processes. For this reason, STMicroelectronics is interested in re-constructing models of its manufacturing processes to being able to control their definitions and quickly react in case a problem is detected.

## Keywords

Sequential Processes, Sequence Alignment, Process Mining, Block-structured Models

## 1. INTRODUCTION

Mining process models from process traces is an alternative to constructing process models from scratch. It also offers a mean to analyze and optimize already existent models. This paper presents an approach for constructing (or reconstructing) process models using process trace data. The idea of this approach is to construct process models using an alignment between different traces of a same process.

The method we propose in this paper could be applied for mining different kind of sequential processes. However, we

are specially interested in mining very large sequential processes (i.e. processes containing hundreds of tasks), such as STMicroelectronics' manufacturing processes. This enterprise is interested in applying process mining methods over its trace data for reconstructing its process models. Semiconductor specifications are modified very often (STMicroelectronics has an average of 50 modifications to its manufacturing process per week). Therefore, process definitions are continuously updated. Products cycle time (i.e. period of production) is mostly longer than sixty days. This means that there are objects being manufactured that can follow different process definitions for a same final product. So, reconstructing process models using process mining techniques will help STMicroelectronics' experts to easily detect added/deleted/modified options to processes and will enable to react quicker if an abnormal situation is detected. The final objective is to reduce reaction time to problems detected on process definitions.

The two main characteristics of the processes we are interested in mining are: 1) they contain hundreds of activities; and 2) duplicated activities are allowed. For this reason, we consider that is helpful to differentiate when a repetition of an activity is part of normal execution of the process and when it is due to problems encountered. In fact, a repetition of a an activity can be due to a situation that arrives frequently during the execution of a process. When executing a process some controls are done. If the desired results are not achieved some actions (special actions) has to be done to correct problems. These actions usually consist on performing special activities and/or the repeating already executed activities.

This is the reason why we propose to align the different process traces to identify equivalent activities over the different executions. It is straight forwards to think that when a subsequence of activities is present in all process traces this subsequence is part of normal execution. Subsequences of tasks executions not aligned can be consider to be special subsequences. If we consider STMicroelectronics production traces, these special subsequences could be: 1) tasks

executed to correct problems; or 2) tasks representing modifications suffered by the process in the period between the dates of the process traces considered.

We have analyzed several multiple sequence alignment methods considering optimality and time and memory needs. After this analysis, we have decided to work with the T-Coffee (Tree-based Consistency Objective Function For alignment Evaluation) algorithm presented in [10]. The original implementation of T-Coffee accepted only biological sequences as input. We have re-implemented this method for process traces sequences.

We present in this article all the necessary steps for creating process models from an alignment. The alignment is used to identify subsequences of tasks that are contained in all traces. As a consequence, all options of the process are also identified. We could analyze if our method is complete and minimal. Completeness describes that a model preserve all the dependencies between activities that are present in a process trace. Minimality assures that the model should neither introduce paths of execution nor spurious dependencies between activities which are not present in the traces considered. In the article we will explain why our method is complete but not minimal.

We have already tested our ideas using STMicroelectronics' manufacturing process traces obtaining encouraging results.

In the following Section (Section 2), we will present the process mining state of art and we will explain why we did not use neither of the algorithms studied. Section 3, introduces the vocabulary that will be used in the rest of the article. In Section 4 we define trace alignment and briefly introduce the alignment methods. In the same Section we present the method chosen, T-Coffee algorithm, and we expose the reasons for this choice. Afterward, in Section 5 we present our process mining method and propose a small example for its well understanding. In Section 6 we show the application of the method for mining STMicroelectronics' production processes. Finally, in Section 7 we make a conclusion and we talk about future work.

## 2. PROCESS MINING STATE OF ART

In this section we will present a small analysis of some process mining methods that have been studied during our research.

The first process mining technique was proposed in the context of software engineering process in [3]. In that article, three methods are described for process discovery: one using neural networks, one consisting in a purely algorithmic approach and one Markovian approach. The neural network approach is not well explained in that article. The purely algorithmic approach constructs *finite state machines* where states are fused if their futures (in terms of possible future behavior in the next  $k$  steps) are identical. The markovian approach uses a mixture of algorithmic and statistical methods and is able to deal with noise. The authors consider the last two approaches to be the most promising. Notice that all of them deal with sequential processes. This could be useful for us because our work is limited for the moment to sequential processes. The model representation used by

these approaches is finite state machines. This representation accepts duplicated activities. However, we decided to avoid using this kind of representation due to the size of the models we are interesting in modeling. There is a consensus to consider that finite state machines are difficult to understand and to validate. The same authors extended their approach to deal with concurrent processes in [4] and [2]. This extension cannot deal with duplicated activities.

Another interesting process mining algorithm is presented in [1]. Authors divide process mining problem into two different sub-problems: graph mining and condition mining. The graph mining problem consist to discover the structure of the process and the condition mining problem consists on finding the conditions associated to the different execution paths. They only treated the first problem in their article, the graph mining. The model representation they used for their processes are *process graphs*. One characteristic of this kind of model is that an activity can appear only once in a model. We find here again the same problem that we found before, the method cannot be used for modeling processes containing duplicated activities.

Other approach has been published in [16]. The method presented in that article can deal with noise and concurrency. The model representation they use are a subset of Petri nets, *WF-nets*. We know that Petri-nets provide a graphical formal language designed for modeling concurrency, what it is appropriate for the kind of processes they want to discover. Their method consists on a series of steps: (i) construction of dependency/frequency table, (ii) induction of a graph from the previous table, (iii) reconstruction of a WF-net from the dependency/frequency table and the graph obtained in the previous steps. Working from the dependency/frequency table forbids the method to deal with duplicated tasks. They transform traces into a dependency/frequency table and they consider that repetitions of an activity in a trace are due to loops. Once again, the proposed approach is not useful for mining processes with duplicated tasks.

[11] introduces a different approach from the others just described. The main difference is that authors of this last article have considered activities' lifespans. All methods presented before consider that activities are executed in an atomic way. Here the approach is different, they consider that execution of task may take some time, so the use two events for identifying the starting and ending times of each activity. In this way they are able to easily discover concurrency between activities. The disadvantage of this method from our optic is the same as with the previous ones. Their model representation is based on process graphs. So, we cannot represent processes containing duplicated activities.

Two articles [13] and [14] present an algorithm for mining exact models for concurrent process. This method consider, as the last presented, that activities are non-atomic. So, two events are considered for each activity, events stating the starting and ending moments of execution of each activity. This method constructs complete and minimal models, complete in the sense that all the traces used for its construction can be obtained from the model and minimal meaning that no spurious trace can be generated from the final model. The

model representation used is block-oriented. The building blocks are terms and operators. This kind of model allows to represent processes containing duplicated tasks, because basic terms are pointers to tasks or sub-processes. However, the approach proposed in both articles [13] and [14] does not consider the mining of processes containing duplicated tasks. We will go back to the block-oriented representation in Section 5.1.

We can see from this analysis, that neither of the approaches studied provided a solution to our problem of mining sequential processes with duplicated activities. However, we found an interesting work in [8]. This article proposes to use trace alignment for process diagnostics. We agree with the authors that “*trace alignment uncovers common execution paths patterns and deviations in the log*”<sup>1</sup>. Based on this, we propose to use trace alignment not just for process diagnostics but also for mining processes.

### 3. NOTATIONS

We will introduce here the vocabulary that will be used in subsequent sections. For the sake of simplicity, we will use notations introduced in [8]:

- Let  $\Sigma$  denote the set of activities.  $|\Sigma|$  is the number of activities.
- $\Sigma^+$  is the set of all non-empty finite sequences of activities from  $\Sigma$ .  $T \in \Sigma^+$  is a trace over  $\Sigma$ .  $|T|$  denotes the length of a trace  $T$ .
- The set of all  $n$ -length sequences over the alphabet  $\Sigma$  is denoted by  $\Sigma^n$ . A trace of length  $n$  is denoted as  $T^n$  i.e.,  $T^n \in \Sigma^n$ , and  $|T^n| = n$ .
- The ordered sequence of activities in  $T^n$  is denoted as  $T(1)T(2) \dots T(n)$  where  $T(k)$  represents the  $k^{th}$  activity in the trace.
- $T^{n-1}$  denotes the  $n-1$  length prefix of  $T^n$ . In other words  $T^n = T^{n-1}T(n)$ .
- An event log,  $\mathcal{L}$ , corresponds to a multi-set (or bag) of traces from  $\Sigma^+$ .

### 4. TRACE ALIGNMENT

In this section we define formally what a trace alignment is. This definition was first introduced in [8].

**DEFINITION 1.** *Trace alignment over a set of traces  $\mathbb{T} = \{T_1, T_2, \dots, T_n\}$  is defined as a mapping of the set of traces in  $\mathbb{T}$  to another set of traces  $\overline{\mathbb{T}} = \{\overline{T}_1, \overline{T}_2, \dots, \overline{T}_n\}$  where each  $\overline{T}_i \in (\Sigma \cup \{-\})^+$  for  $1 \leq i \leq n$  and*

- $|\overline{T}_1| = |\overline{T}_2| = \dots = |\overline{T}_n| = m$ ,
- $\overline{T}_i$  by removing all “-” gap symbols is equal to  $T_i$ ,
- $\nexists k, 1 \leq k \leq m$  such that  $\forall_{1 \leq i \leq n}, \overline{T}_i(k) = -$

<sup>1</sup>A log is the set of traces considered

$m$  in the definition above is the length of the alignment. An alignment over a set of traces can be represented by a rectangular matrix  $\mathcal{A} = \{a_{ij}\} (1 \leq i \leq n, 1 \leq j \leq m)$  over  $\Sigma' = \Sigma \cup \{-\}$  where  $-$  denotes a gap. The third condition in the definition above implies that no column in  $\mathcal{A}$  contains only gaps ( $-$ ). It is imperative to note that there can be many possible alignments for a given set of traces and that the length of the alignment,  $m$ , satisfies the relation  $l_{max} \leq m \leq l_{sum}$  where  $l_{max}$  is the maximum length of the traces in  $\mathbb{T}$  and  $l_{sum}$  is the sum of lengths of all traces in  $\mathbb{T}$ .

#### 4.1 Trace Alignment Algorithms

It is important to realize that there are many different possible alignments between a set of traces. The choice that make us consider that one is better than another one is based on a function *Score* that assigns a numerical value to an alignment. There are many propositions in the literature for this function *Score*, but we will not speak about this here. Refer to [9], [10] or [15] if interested. Function *Score* is calculated based on a matrix that we will present here. The scores for aligning the different activities in  $\Sigma$  are obtained from a matrix  $\mathcal{S} = \{s_{x,y}\} (x \in \Sigma, y \in \Sigma)$ , called similarity matrix. The value  $s_{x,y}$  represents the similarity between activity  $x \in \Sigma$  and activity  $y \in \Sigma$ . There is also a particular value that is considered, the gap penalization value, that is a value associated to the alignment of an activity from  $\Sigma$  with a gap ( $-$ ). So, the objective of alignment algorithms is to find alignments that maximize the *Score* function. However, finding a similarity matrix is not an easy problem.

In biology, there exists some well-known similarity matrices that are used for aligning, for example, DNA sequences. However, this problem has also been encountered in many other areas, i.e. social science field, where there are no well known similarity matrices pre-calculated. The problem in social science is that there is no equivalent to such a strong theoretical and empirical framework like evolutionary theory. Thus the relationship between the symbols constituting sequences of social events remain at least partly uncertain. Some ideas have been proposed to find a way to construct the similarity matrix  $\mathcal{S}$  in the social science context. We consider that it could be interesting in the future to investigate this for constructing special similarity matrices for any kind of sequences, in particular, for process traces sequences, following ideas published in [5].

Pairwise trace alignment methods can be used to find the best alignment between two traces. Alignment methods can be classified into global and local alignment algorithms. The difference between global and local alignments is that global alignments forces the alignment to span the entire length of all query traces whereas local alignment algorithms only identify regions of similarity within all query traces. There is a dynamic programming algorithm proposed in [9] that computes the optimal global alignment between two traces. [15], by contrast, proposes a local alignment method. Sometimes, there are more than one alignment associated to the optimal score. In these situations, an arbitrary choice is made.

Multiple trace alignment is an extension of pairwise trace alignment to align more than two traces at a time. Multiple alignment methods try to align all of the traces in a given

query set. Multiple traces alignments are computationally difficult to produce and most formulations of the problem lead to NP-complete combinatorial optimization problems.

The technique of dynamic programming is theoretically applicable to any number of sequences; however, because it is computationally expensive in both time and memory, it is rarely used for more than three or four sequences in its most basic form. Although this technique is computationally expensive, its guarantee of a global optimum solution is useful in cases where only a few sequences need to be aligned accurately.

Progressive, hierarchical, or tree methods generate a multiple alignments by first aligning the most similar traces and then adding successively less related traces or groups to the alignment until the entire query set has been incorporated into the solution. Progressive alignment results are dependent on the choice of most related sequences and thus can be sensitive to inaccuracies in the initial pairwise alignments.

## 4.2 T-Coffee alignment algorithm

T-Coffee is an algorithm that can be used for multiple trace alignment. This method is broadly based on the popular progressive approach to multiple alignment but avoids the most serious pitfalls caused by the greedy nature of this algorithm. T-Coffee pre-process a data set of all pairwise alignments between the traces. This provides a library of alignment information that can be used to guide the progressive alignment. Intermediate alignments are then based not only on the sequences to be aligned next but also on how all the sequences align with each other.

The T-Coffee algorithm is composed of several steps that we enumerate below:

1. Generating a primary library of alignments;
2. Derivation of the primary library weights;
3. Combination of the libraries;
4. Extending the library;
5. Progressive alignment strategy.

We are not going to describe in detail each step of the approach. Refer to [10] if interested. The first four steps of the approach are concerned in the construction of libraries. T-Coffee makes use of the information in the library to carry out the last step, the progressive alignment, in a manner that allows to consider the alignments between all the pairs while it carry out each step of the progressive multiple alignment. This provides a progressive alignment, with all its advantages of speed and simplicity, but with a far lesser tendency to make errors. We decided to use this algorithm for aligning traces of very large sequential processes because of its advantages over other progressive alignment methods and its complexity.

### 4.2.1 Complexity Analysis

#### Original Implementation

Considering  $N$  is the number of sequences to align and  $L$  is the average length of the considered sequences, the complexity of the whole procedure in its original implementation is given by:

$$O(N^2L^2) + O(N^3L) + O(N^3) + O(NL^2) \quad (1)$$

where  $O(N^2L^2)$  is associated with the computation of the pairwise library,  $O(N^3L)$  the extension,  $O(N^3) + O(NL^2)$  the computation of the progressive alignment.

The CPU time consumption of T-Coffee original implementation was analyzed empirically by the authors. Measurements indicate the apparent complexity of the program is quadratic, both relative to the average sequence lengths and to the number of sequences. This result can be explained by the fact that in the cases analyzed by the authors,  $L \gg N$ . Therefore, the time required for the library and the alignment computation is much larger than the time required for the library extension:  $O(N^2L^2) + O(NL^2) \gg O(N^3L)$ .

The efficiency of this algorithm was the reason why we decided to use it. The original implementation of this algorithm was oriented toward biological sequences and was capable of working with  $\max|\Sigma| = 23$ . In our process traces we sometimes find that  $|\Sigma| = 1000$  so a re-implementation has been necessary. We have analyzed the source code of this original implementation and we re-implement it for our traces.

#### Our Implementation

Now we will consider  $N$  as the number of traces to align,  $L$  as the average length of the considered traces and  $|\Sigma|$  as the number of different activities. We can analyze the different steps that compose the T-Coffee method (Steps 1, 2, 3, 4, 5) and calculate the complexity of our implementation. Looking into the implementation done, the complexity of the first three steps (1, 2, 3) is the following:

$$O(N) + O(N^2L) + O(N^2|\Sigma|^2) + O(N^2L^2) + O(L) \quad (2)$$

Then, the complexity of the library extension (step 4) is the following:

$$O(N^3L^2) \quad (3)$$

in the worst case. However, this will only occur when all the included pairwise alignments are totally inconsistent. In practice, for the data sets used in our examples the complexity is closer to  $O(N^3L)$ .

Considering the implementation of the progressive alignment step (step 5), we obtain the following results:

$$O(NL) + O(LN^2) + O(L^2N^3) + O(N^3) + O(N) \quad (4)$$

As for the previous step, the term  $O(L^2N^3)$  is in the worst case. For our traces, this term is closer to  $O(LN^3)$  so the complexity of this step results :

$$O(NL) + O(LN^2) + O(LN^3) + O(N^3) + O(N) \quad (5)$$

instead.

The CPU time consumption of our T-Coffee implementation was analyzed empirically (following the same idea of the original paper [10]). Our empirical results indicate similar behavior as the one observed by the authors of the original paper [10]. The apparent complexity of the algorithm is quadratic. This result can be explained by the fact that in the cases considered, using STMicroelectronics' traces (see Section 6 for an explanation of these traces),  $L \gg N$  ( $L = 550$ ,  $N \in \{1, \dots, 50\}$ ). This means that  $O(N^2L^2) + O(NL^2) \gg O(N^3L) + O(N^3)$ .

The complexity of our implementation is not as efficient as the original one, concerning the number of comparisons required. However, the apparent quadratic time complexity is also observed in our version of the algorithm.

## 5. OUR PROCESS MINING METHOD

### 5.1 Model Representation

The model representation chosen, i.e., the meta model of the extracted models are based on, is a block-oriented model originally proposed in [13] and [14].

This meta model defines that each model consists of an arbitrary number of nested building blocks. These building blocks are differentiated into operators and terms. Operators combine building blocks and define the control-flow of a workflow. We use the following operators: *Sequence*, *Parallel*, *Alternative*, and *Loop*.

The *Sequence* operator defines that all embedded blocks are performed in a particular sequential order. Let  $\mathcal{S}$  denote the *Sequence* operator and let  $a$  and  $b$  denote two blocks.  $\mathcal{S}(a, b)$  is then a model defining that block  $a$  is always performed before block  $b$ .

The *Parallel* operator defines that all embedded blocks can be performed concurrently, i.e. in parallel or in any order. Let  $\mathcal{P}$  denote the *Parallel* operator and let  $a$  and  $b$  denote two blocks.  $\mathcal{P}(a, b)$  is then a model defining that block  $a$  and  $b$  can be performed concurrently.

The *Alternative* operator defines a choice of exactly one block out of all its embedded blocks. This operator is supplemented by a set of rules determining the choice. Let  $\mathcal{A}$  denote the *Alternative* operator and let  $a$  and  $b$  denote two blocks.  $\mathcal{A}(a, b)$  is then a model defining that either block  $a$  or block  $b$  is performed. Additionally, we define a *Loop* operator  $\mathcal{L}$ . A *Loop* operator contains only one block that is executed repeatedly while its loop condition holds.

Basic terms are pointers to tasks and sub-workflows, which are embedded in a control-flow.  $\epsilon$  represents an 'empty process', used for describing empty paths in alternatives.

We call the model that conforms to this meta-model a block-structured model. We can construct a model top-down by setting one operator as the starting point of the model and nesting other operators until we get the desired control-flow structure. At the bottom of this structure, we embed basic terms into operators which terminates the nesting process. Beside the representation of block-structured models as terms, we represent them in the form of diagrams. Such a diagram depicts the model's blocks in the form of nested rectangles. The control-flow is explicated by arrows and operator symbols.

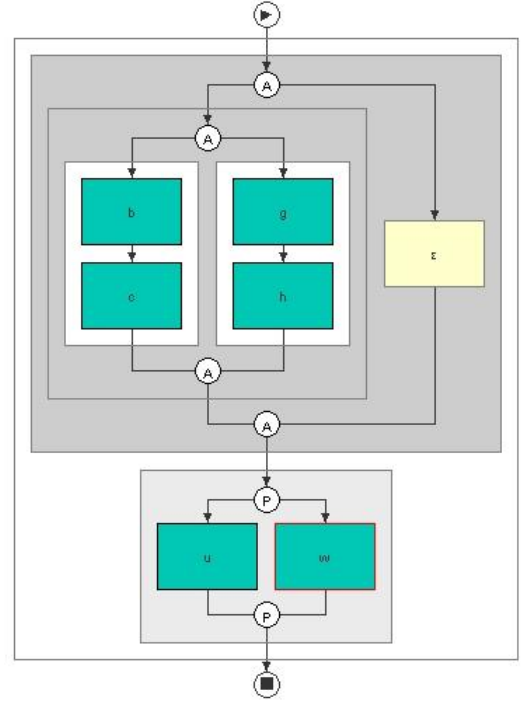


Figure 1: A simple process model

For example, Figure 1<sup>2</sup> [12] shows the process model  $\mathcal{S}(\mathcal{A}(\mathcal{S}(b, c), \mathcal{S}(g, h)), \epsilon), \mathcal{P}(u, w))$  in the form of a diagram. The root of this model is a *Sequence* operator block that contains all other blocks in a hierarchical manner, so that the task pointers  $b, c, g, h, \epsilon$  (empty process),  $u$  and  $w$  are embedded in a nested control-flow.

For this meta-model there is an algebra that consists of a set of axioms covering distributivity, associativity, commutativity and idempotency. These axioms are the basis of term rewriting systems that can be used in order to transform models. A set of basic axioms is shown in Table 1. More details about this meta-model can be found in [13] and [14].

Block-structured models have some advantages: they are well formed and always sound. We have decided to construct

<sup>2</sup>Figure created using ProcessMiner tool, downloaded from <http://www.processmining.de/>



**Table 1: Basic Algebra**

A <sub>1</sub> :	$\mathcal{A}_1(x_1, \dots, x_n) = \dots = \mathcal{A}_n(x_n, \dots, x_1)$
A <sub>2</sub> :	$\mathcal{A}(\dots, x_1, \dots, x_n, \dots) = \mathcal{A}(\dots, \mathcal{A}(x_1, \dots, x_n), \dots)$
A <sub>3</sub> :	$\mathcal{A}(x, x, \dots) = x$
A <sub>4</sub> :	$\mathcal{S}(\mathcal{A}(x_1, \dots, x_n), y_1, \dots, y_m) =$ $\mathcal{A}(\mathcal{S}_1(x_1, y_1, \dots, y_m), \dots, \mathcal{S}_n(x_n, y_1, \dots, y_m))$
A <sub>5</sub> :	$\mathcal{S}(\dots, x_1, \dots, x_n, \dots) = \mathcal{S}(\dots, \mathcal{S}(x_1, \dots, x_n), \dots)$
A <sub>6</sub> :	$\mathcal{P}_1(x_1, \dots, x_n) = \mathcal{P}_n(x_n, \dots, x_1)$
A <sub>7</sub> :	$\mathcal{P}(\mathcal{A}(x_1, \dots, x_n), y_1, \dots, y_m) =$ $\mathcal{A}(\mathcal{P}_1(x_1, y_1, \dots, y_m), \dots, \mathcal{P}_n(x_n, y_1, \dots, y_m))$
A <sub>7</sub> :	$\mathcal{P}(\dots, x_1, \dots, x_n, \dots) = \mathcal{P}(\dots, \mathcal{P}(x_1, \dots, x_n), \dots)$
A <sub>8</sub> :	$x = \mathcal{S}(x, \epsilon)$
A <sub>9</sub> :	$x = \mathcal{S}(\epsilon, x)$

models based on this meta-model because they can be easily converted to many models applicable in different workflow systems. We plan to automate, afterward, the comparison between the workflows constructed from this process mining method to workflow models defined by experts.

## 5.2 Mining Process Models

In this section we will introduce the process mining algorithm we propose to construct models from process traces.

**ALGORITHM 1.** *Given a set of traces  $\mathbb{T} = \{T_1, T_2, \dots, T_n\}$  of process  $P$  where  $\forall 1 \leq i \leq n$ ,  $T_i \in \Sigma^+$ , a provided similarity matrix  $\mathcal{S} = \{s_{x,y}\} (x \in \Sigma, y \in \Sigma)$  and a gap penalty  $d$ . We propose to construct a process model from this set of traces  $\mathbb{T}$  in 3 steps:*

- *Step 1: Construct a trace alignment  $\mathcal{A} = \{a_{ij}\} (1 \leq i \leq n, 1 \leq j \leq m)$  for  $\mathbb{T}$  using T-Coffee alignment algorithm considering  $\mathcal{S}$  and  $d$ ;*
- *Step 2: Considering the alignment  $\mathcal{A} = \{a_{ij}\} (1 \leq i \leq n, 1 \leq j \leq m)$  constructed in the previous step, identify all  $j \in \{1, \dots, m\}$  where  $a_{ij} = a_{kj} \in \Sigma, \forall i, \forall k \in \{1, \dots, n\}$ . We call this set SAA (Same Activity Aligned). Set SAA will contain the indexes of columns where only identical activities are aligned (– gap symbol is not permitted).*
- *Step 3: Considering the set SAA obtained in the previous step we can construct the process model based on the meta model introduced in subsection 5.1 using the following instructions:*
  - *Starting from index  $j$  equal to 1 up to  $m$  (this means we have to consider all columns in the alignment  $\mathcal{A}$ ), identity all pairs of indexes  $(start, end)$  marking the beginning and the ending of indexes belonging to SAA and to  $\overline{SAA}$ <sup>3</sup>, respectively.*
  - *For all pairs  $(start, end)$  where  $start, end \in SAA$  use function *Construct\_Sequence* to construct the block model representing the only sequence admitted by the traces.*

<sup>3</sup> $\overline{SAA}$  means the complement of SAA

Function *Construct\_Sequence*( $\mathcal{A}, start, end$ ) construct a Sequence block model containing the elements of  $\Sigma$  found in  $\mathcal{A}$  between indexes  $start$  and  $end$ .

- *For all pairs  $(start, end)$  where  $start, end \in \overline{SAA}$  use function *Construct\_Alternative\_Sequences* to construct the block models representing each of the different paths introduced by the traces. Function *Construct\_Alternative\_Sequences*( $\mathcal{A}, start, end$ ) construct an Alternative block model, containing a block for each alternative path proposed in alignment  $\mathcal{A}$  between indexes  $start$  and  $end$ . If a path contain more than one activity, the corresponding Sequence block model is constructed.*
- *Finally, respecting the order of all the pairs  $(start, end)$  found, the final model is a Sequence block model containing all the block models constructed in the previous steps, respecting the order of indexes.*

Let us show how it is working by an example.

### 5.2.1 Example

Consider the set of traces  $\mathbb{T} = \{T_1, T_2, T_3, T_4\}$  of process  $P$ . Imagine that:

$T_1 = bcdffghijk$

$T_2 = bcdghijk$

$T_3 = bcdffghi$

$T_4 = bcghik$

$\Sigma = \{b, c, d, f, g, h, i\}$  in this example.

We apply our process mining method over this log.

*Step 1:* We apply T-Coffee alignment algorithm over  $\mathbb{T}$ . Imagine we obtain the following alignment matrix:

$$A = \begin{pmatrix} b & c & d & f & f & g & h & i & j & k \\ b & c & d & - & - & g & h & i & j & k \\ b & c & d & f & f & g & h & i & - & - \\ b & c & - & - & - & g & h & i & - & k \end{pmatrix}$$

*Step 2:* Once we have constructed alignment  $\mathcal{A}$  we proceed to find the indexes of the columns where in each row appears the same identifier of  $\Sigma$  (and no gaps). We construct set SAA. Looking at  $\mathcal{A}$  it is easy to see it is the case of the first two columns and also columns 6, 7 and 8. The resulting set is  $SAA = \{1, 2, 6, 7, 8\}$ .

*Step 3:* We must identity all pairs of indexes  $(start, end)$  marking the beginning and the ending of indexes belonging to SAA and to  $\overline{SAA}$ . We obtain:  $(1, 2)$  where  $1, 2 \in SAA$ ,  $(3, 5)$  where  $3, 5 \in \overline{SAA}$ ,  $(6, 8)$  where  $6, 8 \in SAA$ ,  $(9, 10)$  where  $9, 10 \in \overline{SAA}$ .

We use the appropriate function in each case:

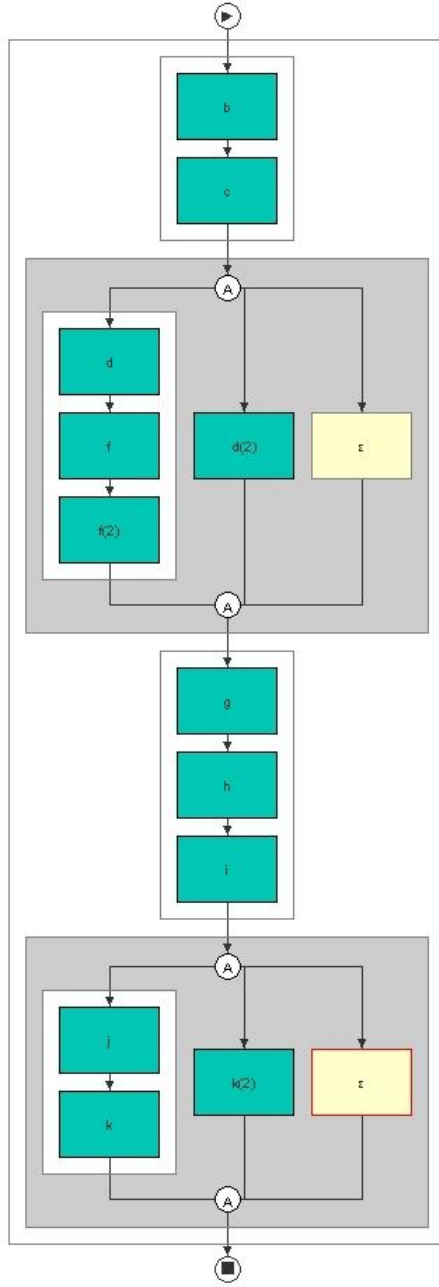


Figure 2: Resulting process model of our example

- *Construct\_Sequence*( $\mathcal{A}, 1, 2$ ). This function will construct the following model:  $\mathcal{S}(b, c)$ .
- *Construct\_Alternative\_Sequences*( $\mathcal{A}, 3, 5$ ). This function construct a model that its root is an Alternative operator and each of the alternative blocks are the different paths proposed in the alignment between indexes 3 and 5. There are 3 paths proposed:  $\mathcal{S}(d, f, f)$ ,  $d$  and  $\epsilon$ . So, the result proposed by function *Construct\_Alternative\_Sequences*( $\mathcal{A}, 3, 5$ ) would be:  $\mathcal{A}(\mathcal{S}(d, f, f), d, \epsilon)$ .
- *Construct\_Sequence*( $\mathcal{A}, 6, 8$ ). The resulting model is

$$\mathcal{S}(g, h, i).$$

- *Construct\_Alternative\_Sequences*( $\mathcal{A}, 9, 10$ ) constructs the model:  $\mathcal{A}(\mathcal{S}(j, k), \epsilon, k)$ .

The final model is a Sequence block model containing the concatenation of the previous blocks just obtained:

$$\mathcal{S}(\mathcal{S}(b, c), \mathcal{A}(\mathcal{S}(d, f, f), d, \epsilon), \mathcal{S}(g, h, i), \mathcal{A}(\mathcal{S}(j, k), \epsilon, k)).$$

A graphical representation of this model is shown in Figure 2.

### 5.2.2 Results

Analyzing the procedure for constructing models proposed we can state that the models we obtain are always complete. The reason is that there is no step that omits a task. All tasks are represented in the resulting model and order between different tasks is also respected. However, the models we obtain are not minimal. We can prove this just looking at the previous example. The resulting model accepts the following sequence of activities:  $b \rightarrow c \rightarrow d \rightarrow f \rightarrow f \rightarrow g \rightarrow h \rightarrow i \rightarrow k$  that is not present in the log used for the model construction.

We must also remark that models we construct could sometimes be transformed into simpler ones. Look again at the result obtained in the example. Consider the constructed block:

$$\mathcal{A}(\mathcal{S}(j, k), \epsilon, k).$$

There we can see that in two paths, activity  $k$  is referenced. We could simplify this model using the basic algebra proposed in Table 1. We could transform that block in the following way:

$$\begin{aligned} \mathcal{A}(\mathcal{S}(j, k), \epsilon, k) &\equiv_{(A9)} \\ \mathcal{A}(\mathcal{S}(j, k), \epsilon, \mathcal{S}(\epsilon, k)) &\equiv_{(A1)} \\ \mathcal{A}(\mathcal{S}(j, k), \mathcal{S}(\epsilon, k), \epsilon) &\equiv_{(A2)} \\ \mathcal{A}(\mathcal{A}(\mathcal{S}(j, k), \mathcal{S}(\epsilon, k)), \epsilon) &\equiv_{(A4)} \\ \mathcal{A}(\mathcal{S}(\mathcal{A}(j, \epsilon), k), \epsilon) & \end{aligned}$$

We are thinking already a way to automatize the construction of simpler models. However, the first interest that was finding a model for easily differentiate normal executions paths from optional executions paths for very large sequential processes has been achieved using our process mining method.

## 6. APPLICATION

We are going to show now the results obtained using our process mining method in a particular case: STMicroelectronics' sequential manufacturing processes. But, before that, we are going to introduce the vocabulary associated to STMicroelectronics' production processes (vocabulary used by its experts).

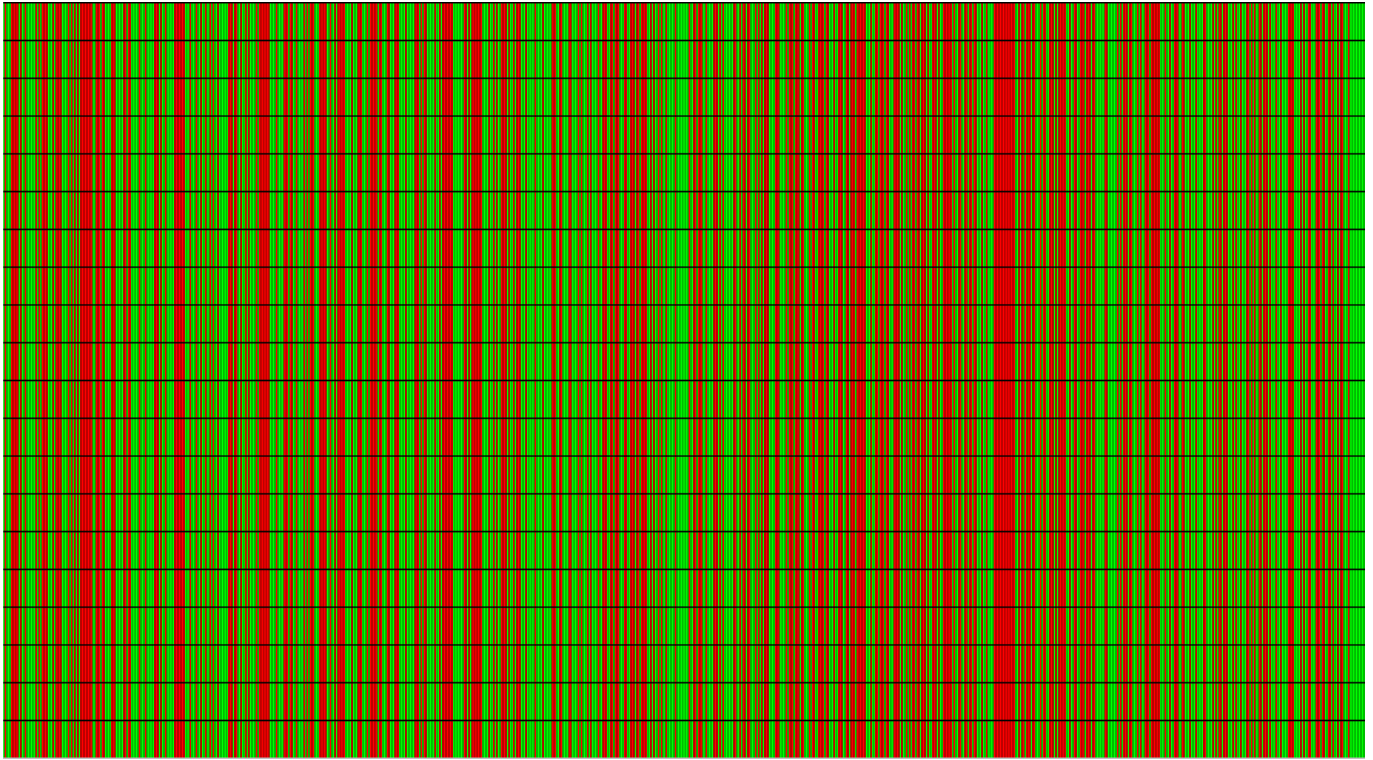


Figure 3: Colored representation of alignment for 20 process traces from STMicroelectronics

### 6.1 STMicroelectronics Experts' Vocabulary

A manufacturing process is called '*Production Route*' in STMicroelectronics' vocabulary. Each of these production routes is composed of a sequence of '*Operations*'. Each operation has a numerical identifier and inside a route, operations' identifiers are ordered respecting an increasing order. Each operation is, as well, a sequence of '*Steps*' and '*Events*'. In fact, a step is always associated to an event. A step is just a numerical identifier used to order the events in an operation; increasing order is also respected here. An event is a desired treatment over the manufactured object. Associated to each event there are '*Entities*'. An entity is a machine that is qualified, i.e. that has been tested and validated for the desired treatment. Depending on the event to perform and the entity chosen, there is a '*Recipe*' associated. A recipe is a list of machine instructions that are given to the entity chosen to perform the desired treatment. There may be different recipes associated to a same event. The reason is that there exists different machines in STMicroelectronics that are used for performing same treatments. So, the list of machine instructions are not the same if the machines are not equal.

A manufactured object is called a *lot* in STMicroelectronics' vocabulary. In fact, a lot is a box containing, at most, 25 silicon plates (*wafers*). During production, the lot goes from one machine to another to be transformed according to the process definition.

### 6.2 STMicroelectronics Data Volume

STMicroelectronics plant situated in the south of France, in Rousset, counts actually with more than 1000 active produc-

tion routes. Each of these production processes is composed of, in average, more than 500 different operations. Each operation is composed of steps, steps number vary from 1 to 4, 5 steps in the longest operations. This means that a lot must go through (in average) 600 different transformations steps before obtaining the final product. Some other interesting numbers can be cited to show the volume of data related to this enterprise processes definitions. There exist approximately 10000 different recipes (list of machines instructions) and there are more than 300 equipments in this plant.

These are the reasons why the alignment method chosen for performing the alignment has been studied in detail, there is a huge data volume to treat. Optimality has been considered. However, time and memory needs were the priority.

### 6.3 STMicroelectronics Traces

STMicroelectronics counts with a Supervision System of Production that is called '*WorkStream*'. Each time a lot is treated, the corresponding information is saved in WorkStream's databases. Information saved in these databases are lot identifier, production route identifier to which the lot is associated, operation, step, event, recipe being executed, starting and ending times of the treatments, machine used, etc. All this information is saved for all lots in production. We use a subset of all this information for testing our approach.

Each trace in this example will be associated to a lot. For each lot, the trace contain an ordered sequence of the activities that the lot went through from the beginning to the

end of its production. Considering the description of STMicroelectronics' processes given in Subsection 6.1, the names of the activities will be considered to be the concatenation between the name of the event and the name of the recipe chosen for execution. We do not consider information such as operations and steps numbers in our traces because they are used by STMicroelectronics' engineers to order the activities. Introducing this information will implicitly introduce information about the process model and we do specially want to avoid that.

## 6.4 Results

We have tested our approach over STMicroelectronics' real production traces. We extracted a set of 20 traces  $\mathbb{T} = \{T_1, \dots, T_{20}\}$  of a same process. Traces length varied from 520 to 620 recorded activities. A total of 551 different activities have been found in  $\mathbb{T}$ , this means  $|\Sigma| = 551$ . We performed an alignment between these traces using our implementation of T-Coffee algorithm, obtaining the alignment matrix  $\mathcal{A} = (\overline{T}_1, \dots, \overline{T}_{20})^T$  as result. We have used a similarity matrix  $\mathcal{S}$  such that  $s_{x,y} = 1$  when  $x = y$  and  $s_{x,y} = -1$  when  $x \neq y$ , where  $x$  and  $y \in \Sigma$ . Aligning any activity identifier with a gap symbol (-) has been penalized using a  $d$  value set to  $-1$ .

In Figure 3 we can see a colored representation of the alignment matrix  $\mathcal{A}$  obtained as result. The dimensions of this matrix are 20 rows  $\times$  884 columns. The construction of  $\mathcal{A}$  corresponds to the *Step 1* of the algorithm proposed in Subsection 5.2. Each trace is represented in one row in the matrix. Due to the huge dimensions of this alignment, we have colored columns for a better understanding of the results. In fact, we do not show exactly the content of this matrix in Figure 3, we just show a colored version of it for the readers' better understanding. Green columns show columns where all tasks identifiers (for all 20 lots) match (set  $SAA$ ). Identifying these green columns corresponds to the *Step 2* of the same algorithm. We colored in red the rest. Analyzing the alignment contents we can appreciate that even though differences have been found between traces, a common process is identified from the 462 columns colored in green (for the sake of brevity this content is not shown in the article). This means that 462 similar tasks have been identified between the 520 – 620 that compose the traces.

Doing an analysis of alignment  $\mathcal{A}$ , we can confirm that differences are due to (in most of the cases) to repetitions of tasks for correcting errors in production, repetition of measures and modifications done to the process definitions between the different dates of the traces that compose the log.

The construction of the model from this colored matrix is straight forwards, *Step 3* in our mining algorithm. However, we are not going to show the model obtained here. In fact, the construction procedure has been already explained for the example of Subsection 5.2.1.

## 7. CONCLUSION AND FUTURE WORK

We are satisfied by the results obtained in the application of our approach over STMicroelectronics traces. The results obtained encourage us to continue to investigate the use of alignments to differentiate normal process executions and special activities executions inside process traces. Our fu-

ture work will surely consist in the transformation of the models obtained by our method into simpler models.

This work is part a project that searches to construct generic process models. These generic process models will be obtained merging models obtained from two different sources: activity (process traces) and experts process specification. For being able to merge models from these two different sources, all models will need to be translated to a common formalism. The formalism that we will use are Timed Sequential Machines (TSM) [7] and [6]. This will enable us to simulate if required in the future.

## 8. REFERENCES

- [1] R. Agrawal, D. Gunopulos, and F. Leymann. Mining process models from workflow logs. In *Sixth International Conference on Extending Database Technology*, pages 469–483, 1998.
- [2] J. E. Cook, Z. Du, C. Liu, and A. Wolf. Discovering models of behavior for concurrent workflows. *Computers in Industry*, 53:297–319, 2004.
- [3] J. E. Cook and A. L. Wolf. Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology*, 7:215–249, 1998.
- [4] J. E. Cook and A. L. Wolf. Event-based detection of concurrency. In *in: Proceeding of the Sixth ACM SIGSOFT International Symposium on the Foundation of Software Engineering (FSE-6)*, volume 53, pages 35–45, 1998.
- [5] J. A. Gauthier, E. D. Widmer, P. Bucher, and C. Notredame. How much does it cost? optimization of costs in sequence analysis of social science data. *Sociological Methods and Research*, 2008.
- [6] N. Giambiasi. An introduction to timed sequential machines.
- [7] N. Giambiasi. From sequential machines to devs formalism. *International Simulation Multiconference (ISM'09)*, 2009.
- [8] R. Jagadeesh Chandra Bose and W. van der Aalst. Trace alignment in process mining: Opportunities for process diagnostics. 6336:227–242, 2010.
- [9] S. B. Needleman and C. D. Wunsch. A general method applicable to the search of similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48:443–453, 1970.
- [10] C. Notredame, D. G. Higgins, and J. Heringa. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, 302:205–217, 2000.
- [11] S. Pinter and M. Golani. Discovering workflow models from activities' lifespans. In *Special issue: Process/workflow mining*, volume 53, pages 283–296, 2004.
- [12] G. Schimm. Process miner – a tool for mining process schemes from event-based data. In *Proceedings of the European Conference on Logics in Artificial Intelligence*, volume 2424, pages 525–528, 2002.
- [13] G. Schimm. Mining most specific workflow models from event-based data. In *Business Process Management'03*, pages 25–40, 2003.
- [14] G. Schimm. Mining exact models of concurrent

- workflows. *Comput. Ind.*, 53:265–281, April 2004.
- [15] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147:195–197, 1981.
- [16] A. Weijters and W. van der Aalst. Process mining discovering workflow models from event-based data. In *Proceedings of the ECAI Workshop on Knowledge Discovery and Spatial Data*, pages 283–290, 2001.

# MODELING PSYCHOLOGICAL MESSAGES AND THEIR PROPAGATION

Colette FAUCHER

LSIS, Laboratoire des Sciences de l'Information et des Systèmes,  
Université Paul Cézanne,  
Avenue Escadrille Normandie-Niemen, 13397  
Marseille, Cedex 20, FRANCE

Colette.Faucher@lsis.org

## ABSTRACT

*In the context of the stabilization phase of modern conflicts, actions of influence are essential and take precedence over combat actions. Among those actions, psychological operations (PSYOPS) are the most important for influencing info-targets by spreading a message that they must read, listen to and understand for them to adopt the desired behavior, which is made by the modification of their attitudes by acting on their perceptions. In this article, the author thoroughly analyzes the different parameters that describe psychological operations (info-targets, means of conveyance, expected behaviors and feelings and so on) as well as the way messages are propagated through social networks. This study is carried out in the context of the system CAPRICORN, whose goal is to assess psychological effects of combat actions, CIMIC actions and psychological actions of the coalition on the allied forces, the actors of the threat and the population, in the framework of the ongoing conflict in Afghanistan.*

**Keywords:** Stabilization phase, PSYOPS operations, message propagation, social networks

## 1. INTRODUCTION

Nowadays modern conflict management is divided into three phases, the intervention, the stabilization and the normalization phases, (Faucher 2011), each of them being characterized by a main action (coercion for the intervention phase, security maintaining for the stabilization phase and aid for the normalization phase). During the stabilization phase, combat actions are relatively limited compared to actions of influence, e.g. actions of communication (Operational Communication), (N°297/DEF/CICDE 2007), psychological actions (PSYOPS), (N°069/DEF/CICDE 2008), and civil-military actions (CIMIC), (N°262/DEF/EMP.1 2005).

The CAPRICORN system is a simulation system that is used in the stabilization phase of the ongoing conflict in Afghanistan, (Bruzzzone 2009). It takes as inputs different types of actions of the coalition force: PSYOPS actions, CIMIC actions and combat actions and it aims at assessing the psychological effects of those actions on the friendly forces, the actors of the threat and the population.

This article deals with CAPRICORN researches on PSYOPS management models.

## 2. WHAT ARE PSYCHOLOGICAL ACTIONS ?

Psychological actions aim at elaborating and spreading a message that must be read, listened to and understood by the *info-targets* in order to get the desired effect, that is, influencing the info-targets to get from them the desired behavior by the modification of their attitudes, by acting on their perceptions, (N°069/DEF/CICDE 2008).

A message must generate in the info-targets, reasoned thoughts, spontaneous feelings and emotions and/or reflex behaviors, depending on the means used to spread the messages.

When an analyst designs a psychological message, he must produce an efficient message that will reach the goal the analyst has in mind, e.g. the latter must get the expected effect on the info-targets.

The paper will then highlight the characteristics of a message that facilitate its understanding, the trust one can have in it and its spreading.

According to Clark, (Clark 2009), efficient messages have the following characteristics: they tap into people's anxieties, they are surprising, but coherent, they reflect the *zeitgeist*, they are simple and concrete and are difficult to disprove. According to Rosnow, (Rosnow 1991), credibility and ambiguity are also important factors.

DiFazo and Bordic, (DiFazo & Bordic 2007), also mention some mechanisms that urge receivers to spread a message: their uncertainty and anxiety, the importance of the message to their eyes, the fact that they believe in the message, their motivation.

Actually, it is possible to distinguish between two types of messages: *propaganda* and *rumor*, (Defencejournal 2000).

- Propaganda

According to the International Encyclopedia of Social Sciences, (Wiki propaganda 2011), «Propaganda is the relatively deliberate manipulation, by means of symbols (words, gestures, flags, images, monuments, music and so on) of other people's thoughts or actions, with respect to beliefs, values and behaviours which these people (reactors) regard as controversial». E.D. Martin says, «Propaganda offers ready-made opinions for the unthinking herd», (quoted in (Choukas 1965)).

- Rumor

In defining a rumor, one can say it is an information on news without verification, content-oriented, (Defencejournal 2000). "A Psychology of Rumor" was



published by Robert Knapp in 1944, (Knapp 1944), in which he reports on his analysis of over one thousand rumors during World War II that were printed in the Boston Herald's "Rumor Clinic" Column. He defines rumor as a proposition for belief of topical reference disseminated without official verification. *So formidably defined, rumor is but a special case of informal social communications, including myth, legend, and current humor. From myth and legend it is distinguished by its emphasis on the topical. Where humor is designed to provoke laughter, rumor begs for belief.*

Rumors tend to be distorted overtime. A rumor travels, it tends to grow shorter, more concise, more easily grasped and told. In successive versions, fewer words are used and fewer details are mentioned. The rumor sharpens, from a larger context to the selective perception, retention and reporting of a limited number of details are developed.

Different *means of conveyance* can be used for spreading out messages:

- audio, visual or audiovisual means,
- «means of opportunity».

When a message is spread out, it's important to be able to measure its impact on the info-targets by using *efficiency criteria*.

In the next paragraph, the author is going to detail the different parameters of psychological actions that have been mentioned in this paragraph.

### 3. PSYCHOLOGICAL ACTIONS PARAMETERS

#### 3.1. Info-targets

The info-targets are the people to whom psychological messages are intended. They can be either opponents to the desired final state (opposing sources), indecisive or neutral people whose support the actor of the psychological activity is trying to gain, or local allies whose engagement he is trying to reinforce. Info-targets can be specific people or one or several groups of people.

The selection of the info-targets, according to the desired final effect and the defined goals is based on several criteria: demographic, geographical, ethnic, religious, behavioral, emotional sensitivity and so on, for all the people and, as far as insurgents are concerned, they can also be discriminated according to their origin, their affiliation or their motivation or a combination of these criteria, (Bruzzzone 2011).

Info-targets can also be classified according to their mode of accessibility:

- *Intentional info-targets*: they are the targets towards whom the message is directed, either they are direct, that is, they receive the message from the means that are used to spread the message, either they are indirect and they get the message from direct info-targets or

other indirect info-targets who propagate the message.

- *Non-intentional info-targets*: the message may reach targets that the analyst who designed the psychological message had not thought of, that's why they are called non-intentional.

### 3.2. Means of conveyance for messages in the framework of a psychological action

#### 3.2.1. Audio, visual and audio-visual means

They consist of printed media (leaflets, posters, magazines, newspapers ads), radio programs, TV programs, local telecommunication networks (messages are spread via texts or mobile phone calls), loudspeakers, internet, goodies (pens, satchels and so on) and face-to-face direct verbal communication.

It is necessary to must stress that some means of conveyance better fit either propaganda or rumors. For propaganda, printed media, radio programs, television programs, loudspeakers and goodies are preferentially used, while phone calls and texts (using local telecommunication networks) as well as the internet and direct verbal communication are better tools to disseminate rumors.

#### 3.2.2. «Means of opportunity»

They convey more instinctive, more reflex, less intellectualized messages that are rather dependent of the info-targets' local customs. An outstretched hand, the symbolic or ritual impact of an action provoke opportunity behavioral stimuli.

They consist of non-verbal communication and other means of action that are called «parades», for example shows of power from the Forces by means of a parade of the military vehicles or airplanes.

- Non-verbal communication

Non-verbal communication, (non verbal communication 2011), that is, communication that is not expressed by means of words, includes: external appearance, rites, spacial relationship behavior, body movements (trunk, limbs, head), facial expressions, looks and visual contacts, para-verbal communication (voice tonality, silences and so on).

In the context of Afghanistan, non-verbal communication is essentially made through appearance and behavior.

##### \* Appearance

For instance, an armed soldier will be regarded as aggressive by the population of a village. Another soldier who will wear sunglasses will be judged as intrusive. Certain type of clothes will also denote certain impressions that a soldier will like to provoke on the population.

##### \* Behavior

Soldiers coming to the local market will be perceived as a symbolic gesture of sympathy and familiarity and will be appreciated as such. Likewise, if a male soldier

enters the part of a house for the women only, he will trigger a feeling of fear in the women and anger in the men.

It is necessary to distinguish between global actions where the behavior and the appearance of people are part of a non-verbal communication, for instance, soldiers on patrol in a village or around a mosque or a meal with a notable, and more punctual gestures or behaviors that are part of those global actions, for example, a handshake between a soldier and the notable or the way that soldier will give the salt to another person during the meal.

In the context of Afghanistan, the ways of behaving in many contexts depend on the culture and the customs of the Afghan people and sometimes largely differ from those of Western people, so an inappropriate gesture or way of behaving from a soldier can cause negative feelings in an Afghan person who will observe it.

For each global action of non-verbal communication, the author will list some local customs to be respected.

Among those ones, it is possible to distinguish between those:

- that will be regarded as neutral, if they are not respected, but positive if they are (for instance, hugging your interlocutor when greeting him).
- that will be regarded as negative, if they are not respected and neutral otherwise (for instance, shaking your interlocutor's hand with your right hand instead of the left one).

When a very strong negative feeling is triggered, it can be «eliminary», in the sense that even if the global non-verbal action of communication is supposed to generate positive feelings, those positive feelings will be erased by the eliminary mistake.

- The other means of action («parades»)

In a logic where the search of effects like appeasement, support, confusion, fear, means of opportunity defined by the context can be used. Those means function through the satisfaction of a need or the loss of satisfaction of this need by the info-target. It is referred to the needs expressed in the Maslow pyramid (physiological needs, need of safety, need of belonging, need of self-esteem, need of actualization), (Maslow 2011).

For instance, the Army, the Marine, a land force can obtain psychological effects by means of shows of strength, shows of presence, intimidation and so on. It must be stressed that in the context of the war in general and in Afghanistan in particular, the logic of the Maslow pyramid may not be respected. For example, a kamikaze soldier will satisfy his need of self-esteem and actualization without obviously satisfying his need of safety.

### 3.3. Efficiency criteria

When a psychological action is carried out, analysts must assess the efficiency criteria, which consist of two parameters:

- The *impact of the message* on the info-targets, that is, whether the message has effectively reached the info-targets or not,
- The *effect(s) of the message*. Has the message generated the desired effects, behaviors, feelings and so on ?

The **impact** of a message can be evaluated through several indicators:

- the *level of insecurity* (number of aggressive actions like IED (Improvised Explosive Devices), direct or indirect shot (rocket, mortar), assassinations directed towards specific people and so on).
- the *contact of the force with the population* (crowds to economic events, number of attenders to *shuras*, number of people attending the inauguration of a CIMIC project, number of people who initiate CIMIC projects, number of people from the population or key leaders who visit the FOB and so on).
- The *amount of information that is communicated* (number of phone calls to the Forces, number of relevant pieces of information given and so on).
- *Control of the Taliban over the population* (racket, curfew, number of night letters, number and intensity of spread rumors (pig fever for example), number of destroyed or damaged CIMIC projects and so on).

The **effects** of a message can be classified according to two typologies.

The first one concerns the type of effects that are looked for:

- *Effects on the perception*: they can be felt through questions that are asked or addressed at *shuras*, goods that are sold at a market and so on.
- *Effects on attitudes* (for instance, propensity to act that can be measured at a collective scale by the questions the locals may ask or on the contrary by the lack of interest they show concerning certain topics).
- *Effects on behaviors* (for example, people coming to the market).

The second typology for classifying the effects to be obtained depends on the info-targets they aim, as said in §3.1:

- *Effects for comforting allied people*: for instance, the friendly *malek* of a village will be

considered as a good governor and will be reelected, the population of a given valley will trust a given positive influencer and so on.

- *Effects for making new allied from neutral people*: the population of such or such village will actively participate in a CIMIC project despite the propaganda of the enemies, Afghan authorities will be perceived positively and trusted for the development of certain projects and so on.

or for putting apart neutral people from opponents (for example, the population understands actions that are bad for the local authorities and try to avoid them).

- *Effects for neutralizing the opponents*, that is, dividing them, dissuading them... (AAF no longer put IED on such road, don't carry out operations of more than 30 people on such an area and so on).

#### 4. DISSEMINATION OF A MESSAGE

When it comes to studying the propagation of psychological messages, it is necessary to distinguish between messages conveying *propaganda* and messages conveying *rumors*.

- Messages conveying propaganda only have an effect on direct info-targets. The messages delivered to them are not propagated to other people.
- Messages conveying rumors, on the contrary, are largely propagated through word of mouth and the high-tech equivalents of word of mouth: phone calls, texts and the internet.

An approach for the dissemination of a message conveying rumors to info-targets is to divide the process in two phases:

- the *spreading of the information* conveyed by the message to direct info-targets as they are defined by the analysts,
- the *propagation of the message* by direct or indirect info-targets to indirect info-targets.

##### 4.1. The first phase of the dissemination process

For this step, it is used the means of conveyance above mentioned in §3.2. In the context of CAPRICORN, the author is working on the region of Afghanistan where the French Forces are based, the province of Kapisa. In this province, not all the means of conveyance are available, nor used at this step of the process. Actually the Forces use printed supports (but not that much because only 30% of the population is literate in this area), radio programs, loudspeakers, goodies, mobile communication, face-to-face verbal communication, non-verbal communication and parades.

It's important to notice that certain means of conveyance require specific conditions for the info-targets to get the message: for example, for a printed message, the info-targets must read, for a radio

program, the info-target must have a radio. It is designated the potential info-targets by the expression «practically receptive info-targets».

It is used the adverb «practically», because this type of receptivity relies on logistic constraints. The receptivity of an info-target can be understood otherwise, if it is considered the way a message affects an info-target depending on their sensitivity (religious, ethnic) or their political opinion, their family status and so on.

Intentional info-targets are defined by the analysts. However, the means used for spreading messages define the physical scope of the message, that is, if a message is disseminated by means of the radio, for instance, the message will reach people of a determined area. Within the area where the message is reachable, it can indeed reach other people than the direct and intentional info-targets. Those people are the non-intentional info-targets above mentioned in §3.1.

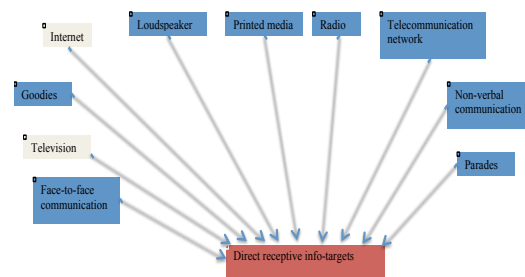


Figure 1: The first phase of the dissemination process (the means of conveyance in white are those that are not used in our system in this phase)

##### 4.2. The second phase of the dissemination process: the propagation

It's a direct or an indirect info-target who propagates a message to an indirect info-target. A message is propagated by word of mouth, mobile phone, if both the sender and the receiver of the message have a mobile, e-mail, if both have an internet access.

The sender of a message propagates a message to people belonging to one or several networks he is part of.

It will be reviewed the type of networks that exist in Afghanistan in §5, some of them being specific to this country.

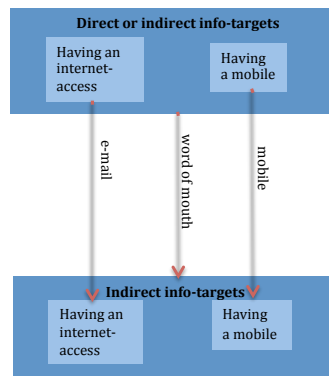


Figure 2: The second phase of the dissemination process: the propagation

## 5. TYPES OF SOCIAL NETWORKS

### 5.1. What is a social network ?

A social network is a network that links people. It is represented by means of vertex and edges. A vertex stands for an individual or a group of individuals. An edge represents a link that connects two vertex and can have different semantics. It can express a family link, a clannish link, an ethnic link, a political link, an economic link and so on.

### 5.2. The social networks in Afghanistan

In Afghanistan, there are two types of social networks, some are stable and others unstable.

#### 5.2.1. Stable networks

- Ethno-linguistic groups

Afghanistan is an ethnic mosaic, (Program for culture 2011), (Tribalism 2011), where there are over 40 major ethnicities who speak over 50 separate languages or dialects. Its citizens naturally identify with those who speak their language and share their culture. Their loyalty is first to their local leaders and their tribe. Actually, ethnicity only plays one part in understanding Afghanistan and its people. It is organized according to many other factors the author will review in the following paragraphs. While the larger ethnic identities of Pashtuns, Uzbeks and Tadjiks and so on, do exist, and while they are important at a general level in understanding Afghan society, politics, economics and security, these categories are only general descriptions of how Afghans view themselves and one another. In some areas of the country, the Western notion of ethnicity has become so politicized that it has become rude to inquire immediately of an Afghan's ethnic identity (i.e. Tadjik, Pashtun, Uzbek).

In the province of Kapisa, the main ethnic groups are: Tadjik, Pashtun, Nuristani.

The Tadjiks tend to live in settled communities as opposed to a nomadic lifestyle.

The Pashtun are divided into tribes: the Pashai, the Safi, the Ghilzai, the Nuristani into the Kata and Ashtu Nuristani.

#### 5.2.2. Unstable networks

- *Qawms*

Although it would seem likely or even obvious that ethnic loyalties should be a strong determining factor for mobilization and social organization on the ground of Afghanistan and, for example, Pashtun tribal loyalties are largely described in the literature and the media, ethno-linguistic groups are not that primordial in Afghanistan.

Olivier Roy, (Roy 2002), argues against assuming that all members of an ethnic group defined by its spoken language actually share a coherent identity with a "will to express themselves politically." Many others agree that loyalties are strongest within local communities, not at a national or ethnic level. Roy concedes that ethnic identities are important, but argues that "primordial" local identities take precedence. These local identities are usually called *qawms*.

A *qawm* is the term used to describe any segment of society bound by solidarity ties, whether it be an extended family, clan, occupational group or village. *Qawm* is based on kinship and patron-client relationships; before being an ethnic or tribal group, it is a solidarity group, which protects its members from the encroachments of the state and other *qawms*, but it is also the scene of internal competition between contenders for local supremacy, (Roy 1989).

Richard Tapper cites the flexibility of the *qawm*:

According to context and situation, *qawm* may involve a varying number of individuals, close kinsmen, a village, an ethnic group, a religious sect or a linguistic group. It is therefore a highly ambiguous and flexible concept allowing for strategic manipulations of identity, (Tapper 1988).

Afghans usually will identify themselves by their *qawm*. Roy states that when they identify themselves by the language they speak they do so "without any ethnic connotation", (Roy 1995), (Roy 2002).

The author is conducting researches in order to use the model of *qawm* described in (Geller 2007) to be used in CAPRICORN. This *qawm* model consists of ten actor types: politicians, religious leaders, commanders, businessmen, warriors, civilians, farmers, drug farmers, organized criminals and drug dealers. Geller assumes that some people can have several roles at the same time. There is a reciprocal dependence between the different actors. For instance, if a politician is in need of military protection, he approaches a commander. In return, a commander receives political appreciation by mere cooperation with a politician. Likewise, if a businessman wants to be awarded an official construction contract by the government, he relies on a politician's political connections. In return, the politician receives a monetary provision, for example bribes, and so on.

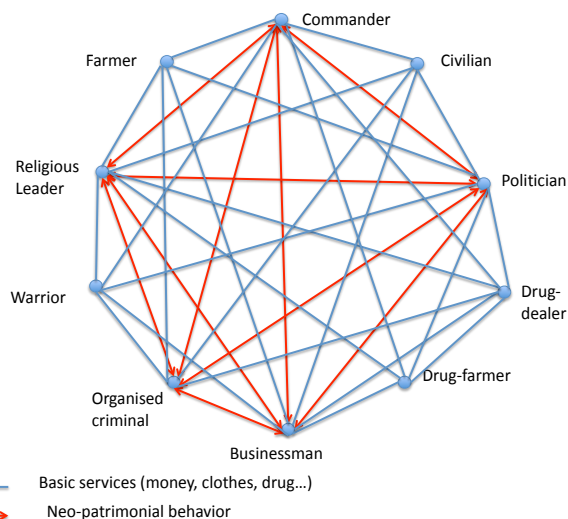


Figure 3: Structure of a *qawm* according to (Geller 2007)

- *Shuras*

*Shura* constitutes one of the *four cardinal principles* in the Islamic perspective on socio-political organization, (Sulaiman 2010). The other three are *justice, equality, and human dignity*.

What is the *shura* principle in Islam? It is essentially *parallel* to the democratic principle in Western political thought, having analogous aspects and about the same tendency or direction. It is predicated on *three basic precepts*. *First*, that all persons in any given society are equal in human and civil rights. *Second*, that public issues are best decided by majority view. And *third*, that the three other principles of *justice, equality and human dignity*, which constitute Islam's moral core, and from which all Islamic conceptions of human and civil rights derive, are best realized, in personal as well as public life, under *shura* governance.

Concretely, a *shura* is a meeting where the tribal elders come together to discuss and listen to what their senior members or representatives of the government are saying. In these days, the *shura* is usually attended by members of the ISAF-NATO forces and civilians from USAID. The members of a *shura* discuss common interest questions. The membership to a *shura* is not fixed, but varies from one meeting to another.

4 traditional *shuras* exist in Tagab valley: *Nazir mohamed, Safi, Population and Security*. In Kapisa, it is possible to find the *shuras Development and Pashai*.

- *Jirgas*

In Afghanistan, there is an official *shura* called *Loya Jirga* (Pashtu). Its role is to make the big decisions concerning the Afghan people, (Wardak 2010).

Apart from the *Loya Jirga*, there are two other official *shuras*. The House of the People or *Wolesi Jirga* is the lower house of the bicameral National Assembly of Afghanistan, alongside the House of Elders. The House of the People has the primary responsibility for making and ratifying laws and approving the actions of the president.

*Mesherano Jirga* or the House of Elders, is the upper house of the bicameral National Assembly of Afghanistan. This house forms more of an advisory role rather than a maker of law. Still, the house has some veto power.

- *Tanzims*

The *tanzim* of Afghanistan have a storied past, (Role of *tanzims* 2011). While one of the earliest was *Hezb-i Islami*, founded by Gulbuddin Hekmatyar in 1973 as an anti-communist resistance movement, most of the *tanzims* emerged as the result of the 1979 Soviet invasion and subsequent influx of foreign aid in response to the invasion. As such, they can be understood as post-revolutionary ethnic movements whose goal was the promotion of narrow sectarian interests, and not political parties with set ideologies and membership sets. It should be emphasized that Afghans today tend to mistrust political parties of all varieties as relics or reminders of the communist era and later the civil war. In other words, the *tanzims* of Afghanistan do not operate like Western or even African or other Asian parties, but are generally viewed by most Afghans as guardians of narrow ethnic interests and are often associated with extra-government militias.

The Main Seven *Tanzims* were founded by groups of mujahideen factions known as the «Peshawar Seven» that were responsible for the majority of guerilla activity and political activism in Afghanistan in the 1980s and 1990s. These seven groups were:

- \* *Hezb-i Islami (Gulbuddin)*: Led by Gulbuddin Hekmatyar, and explicitly Pashtun-nationalist.

- \* *Hezb-i Islami (Khalis)*: maulavi Younas Khali split this faction from Hekmatyar's party in 1979. This group also birthed prominent warlords like Abdul Haq, Jalaludin Haqanni, and Mullah Malang.

- \* *Jamiat-i Islami (Islamic Society)*: Led by former President Berhanuddin Rabbani, it is primarily Tadjik. Most of the party's leadership, including Ahmed Shah Massoud, comes from the Naqshbandi School of Sufism.

- \* *Itehad Islami (Islamic Unity)*: Led by Abdul Rasul Sayaf, this faction was funded primarily by Saudi wahhabists.

- \* *Mahaz-i Milli Islam (National Islamic Front of Afghanistan)*: Led by Pir Sayed Ahmad Gilani of the Qadiri school of Sufism. Once centered in Nangarhar, it was known for being royalist and associated with western-based interests.

- \* *Jabha-i Nijat Milli (Afghan National Liberation Front)*: Led by Sibgrtullah Mojadidi, who was the first post-communist President of Afghanistan, this group often acted as a mediator between the other six factions.

- \* *Harakat-i Inqilab-i Islami (Islamic Revolutionary Forces)*: A party of mostly intellectuals originally led by Mohammad Nabi Mohammadi.

It is important to note that, while the Peshawar Seven were, obviously, based in Peshawar, they did not focus on or emphasize activity exclusively in RC-East.



In fact, they can be divided into three general groups, according to the effect they had on RC-East:

- *Minimal/Unknown effect* (Jabha-i Nijat Milli, Itihad Islami, Harakat-i Inqilab-i Islami)
- *Some effect, little visibility* (Mahaz-i Milli Islam, Hezb-i Islami (Khalis))
- *Major impact, high visibility* (Hezb-i Islami (Gulbuddin), Jamiat-i Islami).

- Temporary networks

Apart from constituted networks the boundaries of which can be unstable, but the content homogeneous according to certain criteria, there are networks that the author calls *temporary networks*, in the sense that they gather together people with no obvious common characteristics, but who are connected to each other because of their common attendance to an event. The author is alluding to events like markets, for example. Those networks of connected people during an event must not be neglected in the study of the dissemination of messages, because they do communicate during events and their role is important, because they actually allow information to be spread between constituted groups (stable or unstable), otherwise information would «stay» within the types of constituted networks the author just listed in the previous paragraphs.

These networks are not permanent. They vary according to the moments. The links they rely on are not systematically activated.

As far as their graphical representation is concerned, they have vertex that represent events. From those nodes go out, in a radial way, links towards participants to the considered event.

Such a network can be activated:

- *permanently during a determined period of time* (for example, a network representing a CIMIC project),
- *periodically* (for instance, a weekly market).

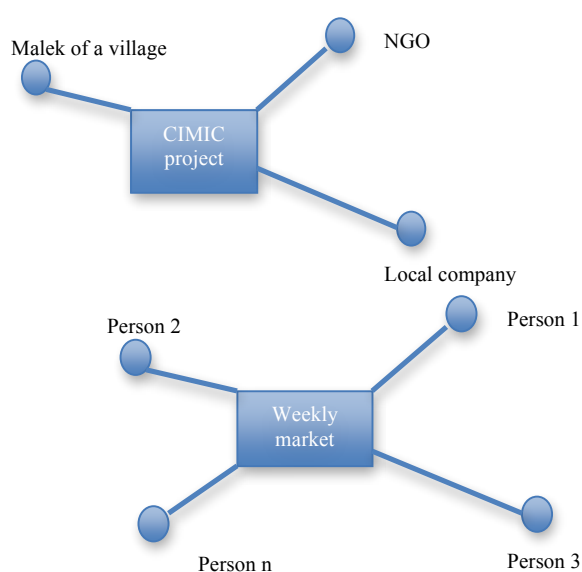


Figure 4: Examples of temporary networks

(network gathering the partners of a CIMIC economic project, network gathering the participants of a weekly market)

### 5.3. Characteristics of social networks that are relevant in the context of Afghan types of networks

#### 5.3.1. Propagation of messages within a group and between groups

Within a constituted group (network), certain people, the *opinion leaders*, have a central position and the dissemination of information within the group largely relies on them, (Lazersfeld 1948), (Weimann 1982). Leaders have strong links with the members of the group. The links that are internal to a group are called *strong links*. The links that connect people belonging to different groups are called *weak links*.

The dissemination of a message within a group depends on the cultural characteristics of the group. The dissemination between groups depends on the cultural proximity of the groups.

Moreover, information that is related to the preservation of the group's interests are more likely to circulate within the group than between the groups, (Lin 1986).

#### 5.3.2. Cultural factors influencing the sharing of information within a group

- People tend to share information with other people to whom they are strongly tied and that information reach more quickly the receivers than the information sent through weak links. This is due to the intensity of the relationships between strongly linked people and the frequency of their contacts.
- The density of the relational network of a country is strongly linked to the cohesion of the groups of that country, the strength of the family and friendly links, the depth and the duration of strong links between people belonging to the same group. This density is more important according to the mentality of the country – a community spirit versus individualism favors such a density, (Svenmarck & al. 2010).

#### 5.3.3. Characteristics discriminating country cultures according to Hofstede

Hofstede has stated five dimensions allowing to describe the culture of a country in a general way, (Hofstede 1980), (Hofstede 2001):

- Power Distance

It indicates the extent to which a society accepts the fact that power in institutions and organizations is distributed unequally. It's reflected in the values of the less powerful members of society as well as in those of the most powerful ones. In particular, power distance controls the way resources and information are managed and shared. In high power distance cultures,



only few people control resources and information, whereas low power distance cultures have a more equal distribution of resources and information.

- **Uncertainty Avoidance**

It indicates to which extent a society feels threatened by uncertain and ambiguous situations and tries to avoid these situations by establishing more formal rules, not tolerating deviant ideas and behaviors, and believing in absolute truths and the attainment of expertise. Nevertheless, societies in which uncertainty avoidance is strong are also characterized by a high level of anxiety and aggressiveness.

- **Individualism vs. Collectivism:**

This dimension encompasses individualism and its opposite, collectivism. Individualism implies a loosely knit social framework in which people are supposed to take care of themselves and of their immediate families only, while collectivism is characterized by a tight social framework in which people distinguish between in-groups and out-groups; they expect their in-group (relatives, clan, organizations) to look after them and in exchange for that they feel they owe absolute loyalty to them.

- **Femininity vs. Masculinity**

Measurements in terms of this dimension expresses to which extent the dominant values in society are «masculine» - that is, assertiveness, the acquisition of money and things, and not caring for others, the quality of life or people. In other words, women tend to take care and to feed, while men need to achieve. These values were labeled «masculine» because within nearly all societies, men scored high in terms of the values' positive sense than of their negative sense – even though the society as a whole might veer toward the «feminine» pole. Interestingly, the more an entire society scores to the masculine side, the wider the gap between its «men's» and «women's» values.

- **Long Term Orientation**

Long term orientation stands for the fostering of virtues oriented towards future rewards, in particular perseverance and thrift. Its opposite pole, short term orientation, stands for the fostering of virtues related to the past and present, in particular, respect for tradition, preservation of «face» and fulfilling social obligations.

#### **5.3.4. Influence of Hofstede's dimensions on the dissemination of messages within a group**

The dimension that most influences the dissemination of a message is uncertainty avoidance.

The author would expect that uncertainty avoidance would urge people to seek more information through networks, but the reality is sometimes different. In countries where people look for certitude, it happens that, in some cases, the arrival of new information seems more threatening than the absence of information, but it's not the most frequent case.

According to (Khalile & Rohani 2009), information is more spread out in countries having:

- a low power distance,
- a higher uncertainty avoidance,

- a higher long term orientation,
- a higher institutional collectivism and a lower in-group collectivism.

Collectivism's role is then rather ambiguous.

#### **5.3.5. Between-culture factor**

In the same way as Hofstede's dimensions allow to evaluate the propensity of people having the same culture for disseminating messages within their group, when a message is sent from a person to another person belonging to another culture, it is possible to evaluate to which extent the cultures are closed to each other which influences the way the message will tend to be communicated from the first person to the another one, (van Vliet 2011). The comparison between cultures depends on the following parameters:

- Presence of previous conflicts in the cultures' history,
- Similarity of the values of the following characteristics in the cultures:
  - \* habits,
  - \* appearance,
  - \* language,
  - \* economic and social status,
  - \* family size,
  - \* political participation,
  - \* collectivism.

#### **5.4. Propagation of information within each type of networks in Afghanistan**

##### **5.4.1. Within constituted stable or unstable networks**

If the message reaches the leader, the latter propagates the information to all the members of his network.

If the message reaches a member of the group, he communicates the information to the leader, who propagates it to all the members of his group.

##### **5.4.2. Within a *qawm***

If the message reaches a member of the *qawm*, he propagates the information to all the people who are connected to him via a link modeling a neo-patrimonial behavior or a link representing the sharing of resources.

##### **5.4.3. Within a temporary network**

When a member receives a piece of information, it is accessed the event node to which this member is connected and it is propagated the information in a radial way to the other members of the network.

### **6. DYNAMICS OF MESSAGE PROPAGATION**

#### **6.1. General scheme**

Contrary to most systems that use models of messages dissemination that are inspired from the epidemiological model of propagation, (Keeling & Rohani 2009), (Salter 2011), the author has defined in his system an original model as proposed in the present paragraph.

The author will explain progressively the following

message propagation scheme.

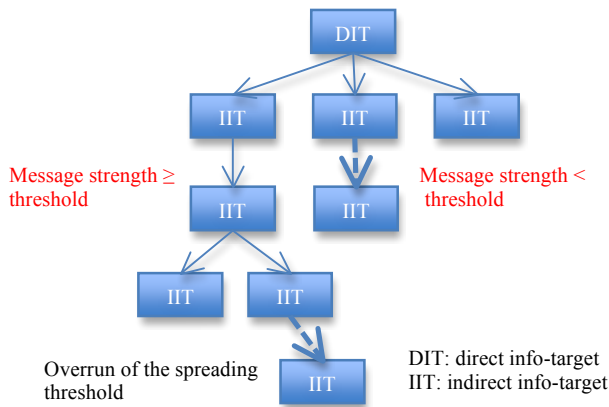


Figure 5: Message propagation model

## 6.2. Strength of a received message

The strength of a message received by the direct info-targets is defined by the analyst (a military) as the *importance of the message* for each group which the info-targets can belong to.

The importance is equal to:

- 1 for trivial information,
- 2 for a moderately important information,
- 3 for crucial information.

Initially, the strength of a message is equal to its importance/3 (the value is normalized between 0 and 1). For indirect info-targets, the strength of the received message is calculated functions of:

- the *strength of the message before the transmission*, i.e. the strength of the message when it was received by the present sender of this message,
- the info-target's *receptivity*,
- the sender's *credibility*,
- the *strength of the transmission*.

The info-target's receptivity is given by the analyst. An info-target's receptivity expresses the interest of the receiver concerning the content of the message. For example, if the message concerns the construction of a *kareze* in a rural part of the country, urban people will not feel very interested in the information and their receptivity will not be very high.

The sender's credibility depends on the concerned info-target. For instance, if the sender is a leader of a *shura* the receiver is part of, his credibility will be high to the eyes of the receiver.

The possible values of the receptivity and credibility factors are low, medium and high and they are respectively quantified 1, 2 and 3.

The strength of the transmission depends, among other things, on the strength of the connection between the sender and the receiver (if they are linked via several networks, the strongest link is chosen). It's evaluated 0 or 1.

The strength of a message can be computed as follows, using the abbreviations:

SMBT: strength of the message before transmission,

RR: receiver's receptivity,

SC: sender's credibility

SMT: strength of the message transmission,

SM: strength of the message when the receiver gets it.

$$\boxed{SM = SMBT * RR/3 * SC/3 * SMT}$$

This value is defined between 0 and 1.

## 6.3. Strength of a message transmission (SMT)

### 6.3.1. If the sender and the receiver belong to the same ethnic group

The strength of a message transmission depends on the Intra-Culture Factor (ICF) of the ethnic group, the strength of the link (LS) connecting the sender and the receiver, the product of the stability factors of, respectively, the sender's and the receiver's zone and the easiness of communication between those zones.

- If the means of conveyance is word of mouth:

$$\boxed{SMT = ICF * LS/4 * \text{product of the stability factors of the zones} / 9 * \text{easiness of communication} / 3}$$

- If the means of conveyance is telephone or e-mail:

$$\boxed{SMT = ICF * LS/4 * \text{product of the stability factors of the zones} / 9}$$

### 6.3.2. If the sender and the receiver belong to different ethnic groups

The strength of a message transmission depends on the Between Culture Factor (BCF) of the ethnic groups, the strength of the link (LS) connecting the sender and the receiver, the product of the stability factors of, respectively, the sender's and the receiver's zone and the easiness of communication between those zones.

The means of conveyance is always word of mouth, as only two very closed people (belonging to the same ethnic group) communicate via telephone or e-mail.

$$\boxed{SMT = BCF * LS/4 * \text{product of the stability factors of the zones} / 9 * \text{easiness of communication} / 3}$$

## 6.4. Computation of the moment M a message is received given the moment T of its dissemination to the direct info-targets

### 6.4.1. If the sender and the receiver belong to the same ethnic group

- If the means of conveyance is word of mouth:

$$\boxed{M = T + \text{average number of days for a person to join the other by usual means of transportation}}$$

- If the means of conveyance is mobile phone or e-mail:

$$\boxed{M = T + 1}$$

### 6.4.2. If the respective ethnic groups of the sender and the receiver of the message differ

The means of conveyance is word of mouth.

- If the sender and the receiver do not belong to a common temporary network:

**$M = T + \text{average number of days for a person to join the other by usual means of transportation}$**

- If the sender and the receiver belong to a common temporary network:

**$M = T + \text{the number of days between the diffusion of the message and the next opportunity of a new meeting gathering them}$**

#### 6.5. Propensity to diffuse a message within an ethnic group: Intra-Culture Factor (ICF)

It depends on the value of the following characteristics in the ethnic group's culture: Collectivism, Power Distance, «Femininity», Uncertainty Avoidance, Long Term Orientation. Its value is defined between 0 and 1.

A characteristic reinforces the value of the ICF, if its value goes along with a more important dissemination of the information.

Each contribution is weighted according to the importance of the associated characteristics. Let a, b, c, d, e, be the respective weights associated to the different characteristics. According to the remarks of §5.3.4, it is possible to define:

Collectivism (C): Given the ambiguity of the collectivism factor, its contribution is always evaluated 0.5.

Power distance (PD): If it's low, its contribution is 1, 0 otherwise.

Femininity (F): If it's high, its contribution is 1, 0 otherwise.

Uncertainty avoidance (UA): If it's high, its contribution is evaluated 1, 0 otherwise.

Long term orientation (LTO): If it's high, its contribution is evaluated 1, 0 otherwise.

**$ICF = aC + bPD + cF + dUA + eLTO / a+b+c+d+e$**

#### 6.6. Propensity to disseminate a message between ethnic groups: Between-Culture Factor (BCF)

It is defined between 0 and 1. It depends on:

- the existence of previous conflicts in the history of the ethnic groups,
- the similarity of the values of the following characteristics in the respective cultures of the ethnic groups:

- Habits,
- Appearance,
- Language,
- Economic and social status,
- Family size,
- Political participation,
- Collectivism.

Two values are possible for evaluating the existence of previous conflicts (PC) (0, if no conflicts, 1 otherwise). For the other characteristics, it is possible to evaluate the contribution (respectively H, A, L, ESS,

FS, PP, Col) to 1, if they have identical values in both ethnic groups and 0, otherwise.

**$BCF = a(1-PC)+bH+cA+dL+eESS+fFS+gPP+hCol / a+b+c+d+e+f+g+h$**

#### 6.7. Value of the link strength between two people of a network (LS)

This value depends on the semantics of the link connecting the two people.

- If the network is family or ethnic group type:  **$LS = 4$**
- If the network is a religious constituted network:  **$LS = 3$**
- If the network is a constituted network of another type:  **$LS = 2$**
- If the network is a temporary network:  **$LS = 1$**
- If the network is a *qawm*: **Strength of the neo-patrimonial link= 3**

**Strength of the link «resources sharing» = 1**

#### 6.8. Value of the stability factor related to a person of a network (SF)

That is the measure of the stability factor of the zone where the group the person belongs to is located.

Its value is:

- **$SF = 3$** , if the zone is in the stability state 2 ("Self-sustaining peace"), according to the classification found in (Driedzic 2008),
- **$SF = 2$** , if the zone is in the stability state 1 ("Assisted stability"),
- **$SF = 1$** , if the zone is in the stability state 0 ("Imposed stability").

#### 6.9. Value of the factor «easiness of communication» between two people (EC)

This factor evaluates the easiness to communicate between two places (for instance, no mountains between the two places) where the people who communicate are located. The easiness of communication between regions depends on geographical factors, as well as political and economic factors.

- **$EC = 3$** , if the two regions can easily communicate,
- **$EC = 2$** , if the two regions can communicate rather easily,
- **$EC = 1$** , if the communication is difficult.

The area covered by our system is divided into regions. The matrix of «easiness of communication» is calculated between the regions taken two by two. The latter must be updated regularly.

#### 6.10. Stopping of a message propagation: dissemination threshold

The message is no longer propagated to a potential receiver when the dissemination threshold is reached, that is, either:

- the strength of the message for this potential receiver becomes lower than a certain threshold to be determined or,
- the time passed since the diffusion of the initial message (to the direct info-targets) exceeds a certain threshold or,
- the total number of informed people exceeds a certain threshold (the information has lost its value of «news»).

#### 7. FORMAT DEFINING A PSYOPS MESSAGE

When the user of CAPRICORN provides a psychological operation as an input for the system, he must specify the following parameters (the parameters must not be systematically instantiated, depending on the values of some of them):

- **Name:** it is specified the name given to the psychological operation.
- **Means of conveyance:** leaflet, magazine, newspaper, radio program, television program, text, phone call, loudspeaker, internet, goodies, direct verbal communication, appearance, behavior, parade.
- **Scenario:** For complex behaviors including several sub-behaviors that are related to customs to be respected, the analyst has to fill up a form where he will indicate for each sub-behavior if it conforms the customs or not. If the sub-behavior can be eliminatory, the consequence will be very negative on the effects of the initial complex behavior.
- **Content of the message:** It must be specified only if the message is printed.
- **Type of the content of the message:** propaganda or rumor.
- **Info-targets:** list of individuals or groups of individuals.
- **Info-targets' receptivity:** list <group, level of receptivity>  
For each group, it is specified its level of receptivity.
- **Importance of the message:** list <group, importance of the message>  
For each group, it is specified the importance of the message defined initially by the analysts.  
This value is the strength of the message initially, when it is sent by the Allied Forces.
- **Expected effects:** tuples <group, feeling, high-average-low, target-group>, <group, behavior, more frequent-less frequent>, <group, need type, need, increased-decreased>  
For each group, is specified:

\* each feeling of the group towards a target-group, as well as the intensity of the feeling (high, average, low).

\* each behavior adopted by the group and it is specified if it has become more frequent or less frequent.

\* each need felt by the group, the type of need and the need itself, and it is specified if it has increased or decreased.

The need refers to the Maslow pyramid.

- **Efficiency criteria:**  
<list of impacts, list of effects>  
Both the impacts and the effects are specified.
- **Date:** precise date when the operation is to take place.
- **Geographical area:** village or list of villages, hamlet, MGRS zone.

#### 8. CONCLUSION

In this article, the author has defined the parameters describing psychological messages and the way those parameters affect the dissemination of the messages through the social networks to which the direct and indirect info-targets belong. The simulation of the propagation of this type of messages is important because it allows the analysts to realize the impacts and the effects of the sent messages over the friendly forces, the actors of the threat and the population, because the system is able to figure out all the people reached by the messages as well as the feelings, needs or behaviors those messages have triggered in them, whereas the analysts had not thought of some of those people when designing the messages.

#### REFERENCES

- Bruzzzone A.G., Frydman C., Cantice G., Massei M., Poggi S., Turi M., 2009. "Development of Advanced Models for CIMIC for Supporting Operational Planners", Proc. of I/ITSEC2009, Orlando, November 30-December 4.
- Bruzzzone, A., 2011. CAPRICORN, Deliverable D-210-2, *State of the Art for Human Behavior Models in a CIMIC and PSYOPS*.
- Choukas, M., 1965. *Propaganda Comes Of Age*, Washington, D.C.: Public Affairs Press.
- Clark, T., 2009. The 81/2 laws of rumor spread, *Psychology Today online*, November 1, 200.
- Defencejournal, 2000. Available from: <http://www.defencejournal.com/2000/jan/p-warfare.htm>, [Accessed 29 July 2011].
- DiFonzo, N., Bordia, P., 2007. *Rumor, gossip and urban legends*, Diogenes, 54, pp.19-35.
- Doctrine interarmées, 2006. *Les opérations d'information*, DIA-3.10, Doctrine interarmées, Centre interarmées de Concepts, de Doctrines et d'Expérimentation, N°570/DEF/EMA/EMP.1/NP du 29 mai.
- Doctrine interarmées, 2008. *Les opérations militaires d'influence*, DIA-3.10.1, Doctrine interarmées,

- Centre interarmées de Concepts, de Doctrines et d'Expérimentation, N°069/DEF/CICDE/NP du 05 mars.
- Doctrine interarmées, 2005. *La coopération civilo-militaire* (CIMIC), CIA-9, Doctrine interarmées, Centre interarmées de Concepts, de Doctrines et d'Expérimentation, N°262/DEF/EMP.1/NP du 03 mars.
- Doctrine interarmées, 2007. *La communication opérationnelle*, DIA-3.10.2, Doctrine interarmées, Centre interarmées de Concepts, de Doctrines et d'Expérimentation, N°297/DEF/CICDE/NP du 26 juillet.
- Driedzic, M., Sotirin, B., Agoglia, J., 2008. *Measuring progress in conflict environments, A metrics framework for assessing conflict transformation and stabilization*, Available from: [www.usip.org/files/resources/MPICE\\_final\\_compl ete%20book%20\(2\).pdf](http://www.usip.org/files/resources/MPICE_final_compl ete%20book%20(2).pdf), [Accessed 29 July 2011].
- Faucher C., 2011. CAPRICORN, Deliverable D-130-1, *EU MS Requirements Joint Requirements for Simulation and CGF in Non-Article 5/San Petersburg Crisis Management Operations Planning Tasks (Peace keeping, Peace enforcement, Peace making)*.
- Geller, A., Moss, S., 2007. Growing qawms. A case-based declarative model of Afghan power structures, *The Fourth Conference of the European Social Simulation Association, ESSA 2007*, Sept. 10-14, Toulouse, France.
- Hofstede, G., 1980. *Culture's consequences*, Beverly Hills, CA : Sage.
- Hofstede, G., 2001. *Culture's consequences. Comparing values, behaviors, institutions and organizations across nations*. Thousands Oaks, CA: Sage.
- Keeling, M.J., Rohani, P., 2008. *Modeling infectious diseases in humans and animals*. Princeton: Princeton University Press.
- Khalile, O.E.M., Seleim, A., 2009. Natural culture practices and societal information dissemination capacity, *Proceedings of the 2009 conference on Information Science, Technology and Applications*, Kuwait : Kuwait, pp.104-113.
- Knapp, R.H. , 1944. A psychology of Rumor, *Public Opinion Quaterly*, 8, pp.22-37.
- Lazarsfeld, P.S., Barelson, B., Gaudet, H., 1948. *The people's choice*, New York, NY: Columbia University Press.
- Lin, N., 1986. Conceptualizing social support, in N. Lin, A. Dean & W.M. Ensel (Eds) *Social support, life events and depression*, Orlando, FL: Academic Press.
- Maslow, 2011. *Pyramide de Maslow*, Available from: [www.communicationorale.com/maslow.htm](http://www.communicationorale.com/maslow.htm), [Accessed 29 July 2011].
- Non verbal communication, 2011. *Exploring non verbal communication*, Available from: <http://nonverbal.ucsc.edu/>, [Accessed 29 July 2011].
- Program for culture, 2011. *Program for culture & conflict studies*, Available from: [www.nps.edu/programs/ccs](http://www.nps.edu/programs/ccs), [Accessed 29 July 2011].
- Rosnow, R.L., 1991. Inside rumor : A personal journey, *American Psychologist*, 46, pp.484-496.
- Roy, O., 2002. Afghanistan: Internal Politics and Socio-Economic Dynamics and Groupings, *WRITENET Paper*, No. 14.
- Roy, O., 1995. *Afghanistan: From Holy War to Civil War*. Princeton, NJ: Darwin Press.
- Roy, O., 1989. Afghanistan: Back to Tribalism or on to Lebanon ?, *Third World Quarterly*, Vol. 10, No. 4. October, pp. 70-82.
- RRC-AFI-08-0023, 2010. *Role of tanzims (Political-military parties) in RC-EAST*, unclassified //FOUO, Tracking Number RRC-AFI-08-0023.
- Salter, W.J., Mc Cormack, R., 2011. Applying Epidemiological Modeling to Idea Spread, *Advances in Cross-Cultural Decision Making*, Dylan Schmorow and Denise Nicholson Eds., CRC Press, Taylor and Francis Group.
- Sulaiman, S.J., 2010. *The Shura principle in Islam*, available from: <http://www.alhewar.com/SadekShura.htm>, [Accessed 29 July 2011].
- Svenmarck, P., Huibregtse, J.N., van Vliet, A.J., van Hemert, D.A., Lundin, M., Sjoberg, E., van Amerongen, J.M., 2010. Message Dissemination in Social Networks for Support of Information Operations Planning, NATO Research and Technology Organization, Papers presented at the RTO Human Factors and Medicine Panel (HFM), *Symposium held in Amsterdam, Netherlands on 18-20 october 2010, RTO-MP-HFM-202, Human Modelling for Military Applications*.
- Tapper, R., 1988. Ethnicity, Order, and Meaning in the Anthropology of Iran and Afghanistan, in J.-P. Digard (Ed.) *Le Fait Ethnique en Iran et en Afghanistan*. Paris: Editions du CNRS.
- Training and doctrine command, 2011. Tribalism in Afghanistan, the cultural geography of Afghanistan, *Training and Doctrine Command, Theater specific, Individual Requirement Training Course*.
- van Vliet, T., Hinbregtse, E., van Hemert, D., 2011. Generic message propagation simulator, *Advances in Cross-Cultural Decision Making*, Dylan Schmorow and Denise Nicholson Eds., CRC Press, Taylor and Francis Group, 2011.
- Weimann, G., 1982. On the importance of marginality: one more step into the two-step flow of communication, *American Sociological Review*, 47, pp.17-30.
- Wikipedia 2011, the Free Encyclopedia, Propaganda, Available from: [www.wikipedia.org/wiki/Propaganda#Techniques](http://www.wikipedia.org/wiki/Propaganda#Techniques), [Accessed 29 July 2011].

# NAV – THE ADVANCED VISUALIZATION STATION: A MOBILE COMPUTING CENTER FOR ENGINEERING PROJECT SUPPORT

Gabriel A. Fernandes, MSc.<sup>(a)</sup>, Gerson G. Cunha, DSc.<sup>(b)</sup>, Tiago Mota, MSc.<sup>(c)</sup>, Celia Lopes, DSc.<sup>(d)</sup>

<sup>(a)</sup> <sup>(b)</sup> <sup>(c)</sup> <sup>(d)</sup> Universidade Federal do Rio de Janeiro – LAMCE/COPPE

<sup>(a)</sup> [gabriel.ufri@gmail.com](mailto:gabriel.ufri@gmail.com), <sup>(b)</sup> [gerson@lamce.coppe.ufri.br](mailto:gerson@lamce.coppe.ufri.br), <sup>(c)</sup> [tiagosmota@lamce.coppe.ufri.br](mailto:tiagosmota@lamce.coppe.ufri.br),  
<sup>(d)</sup> [celia@lamce.coppe.ufri.br](mailto:celia@lamce.coppe.ufri.br)

## ABSTRACT

This paper reviews the development progress of the Advanced Visualization Center (NAV). A mobile laboratory focused on outdoor Augmented Reality (AR) Research and collaborative project development. The NAV is composed of a series of integrated solutions which combined aim to improve project management performance, by improving, on the field, information access and remote project supervision. The approach uses several mobile work stations linked by a wireless network, which can provide positioned based information, such as, comment ready CAD sheets, 3D models, outdoor Augmented Reality visualization and constant GPS navigation/tracking. The NAV overall system also provides panoramic telepresence for remote project supervision, inspection and decision making. The work presents the NAV overall concept, and then describes the main solutions and their current development status and last finishes by discussing the tool integration and possible future continuous research.

Keywords: Augmented Reality, Collaborative Decision Making, Construction, Project Management

## 1. INTRODUCTION

This paper reviews the development progress of the Advanced Visualization Center (NAV). A mobile laboratory focused on outdoor Augmented Reality (AR) Research and collaborative project development. The NAV is composed of a series of integrated solutions which combined aim to improve project management performance, by improving, on the field, information access and remote project supervision. The approach uses several mobile work stations linked by a wireless network, which can provide positioned based information, such as, comment ready CAD sheets, 3D models, outdoor Augmented Reality visualization and constant GPS navigation/tracking. The NAV overall system also provides panoramic telepresence for remote project supervision, inspection and decision making. The work presents the NAV overall concept, and then describes the main solutions and their current development status and last finishes by discussing the tool integration and possible future continuous research.

The evolution of AR is a continuous search for the complete tangible interface which completely integrates visual perception and information in a seamless manner. Several researches have been published discussing the use of AR in Civil Engineering, among those is the extensive work of (Wang and Dunston 2006) which lists the possible applications in projects and the ergonomic difficulties from mobile visualization gear, there is also the work of (Shin and Dunston 2008) which have categorized in detail the activities in the construction environment and identified possible areas which could benefit from AR use. Another important consideration is the potential of remote observation allied to collaborative decision making and possible ramifications when integrated with solutions such as the virtualization gate by (Petit, Lesage et al. 2009). Not only must we map the activities that could benefit from AR but also predict what new services and resources could be created to improve current project work routines in the entire production chain. Examples of collaborative work using AR can be found in (Kim and Maher 2008), (Kim and Jun 2008), (Chen, Chen et al. 2008).

Although very present in marketing campaigns, toys and games the practical use of AR solutions in engineering projects has been discussed and elaborated over the past few years. The difficulty lies in the development of solutions which could be widely adopted by the market. Main issues still lie in heavy/large precision tracking devices, head mounted displays are still clumsy/fragile and are not prepared for outdoor heavy duty activities. We could also add that they do not adhere to safety regulation in dangerous environments, such as a construction site. The NAV main objective is to fill the gap between experimental and practical application in mobile visualization for engineering. Over the past years, several solutions at NAV have been put in constant field testing. These experiments have led to several improvements in interface design, hardware specifications and wireless network configuration. This mobile laboratory was originally developed for long term research in remote areas with small or no technological infrastructure, as such, it has solar panels, rechargeable high capacity



batteries, accommodations, office work environment, kitchen, bathroom, local network system, air conditioner, surveillance, high performance computers, RTK GPS kits and a full computer controlled infrastructure. Its spaces are composed of draws which can be retracted and provide it with a common container shape.

Currently the NAV mobile computing support center combines a series of visualization tools to integrate functions and activities in engineering projects, its main role is to develop and implement practical Augmented Reality applications and collaborative work solutions. Currently the prototype is localized at the Technological Park at the Federal University of Rio de Janeiro Campus and is under the supervision of the Applied Virtual Reality Group (GRVa) which is a part of LAMCE. Figure 1 shows NAV and LAMCE expansion. It has been positioned near the LAMCE expansion construction site which has been for the past years the main test ground for new developments in AR tools. The LAMCE expansion presents many construction challenges due to its unusual interior and exterior design, and so pushes the development of solutions to new boundaries. The current challenges include the transition between the current AR sensor tracking system (based on Intersense IS-1200) to the parallel tracking proposed by Klein (Klein and Murray 2007; Klein and Murray 2009).



Figure 1: NAV mobile lab and LAMCE Construction.

## 2. AR AND THE NAV SOLUTION

Practical AR depends on a combination of hardware, software and content. The following text will present a brief overview of general AR implementation demands: Physical Device, Interaction, Tracking and Application. Through a “device” point of view the equipment must weight mobility, high processing power, network performance and screen visibility under diverse weather and light conditions. Interaction can be achieved through gesture recognition, multi-touch controls, voice recognition, normal input buttons/controls and sensor input (device orientation and movement). As for tracking, there are two main approaches: Sensor and Camera tracking. These can be used separately or combined in a fused priority/opportunity based solution. Sensor tracking is a straightforward method and can be implemented by the use of gps, gyroscopes and accelerometers, as for camera tracking, a solution can range from fiducial (Hornecker and Psik 2005) placement to feature recognition (Zhuo and Xinyu 2010) and/or parallel tracking (Castle, Klein et al. 2011). When a camera based tracking system is being

used, there will also be the need for computer vision calculation to determine the users point-of-view, otherwise sensor tracking is reasonably straightforward and requires only alignment between real/virtual content. The Application is where the user can access and input general visual information and compare expected and obtained results. Interface requirements can vary from one solution to another, such as touchscreen input, 3d model navigation options and CAD style layer control. The following items present three of the undergoing NAV AR solutions.

### 2.1. Panoramic AR Remote Supervision

The remote supervision tool allows a decision maker to inspect a project from any location. Usually this can be easily achieved through a simple camera surveillance system. The Panoramic AR remote supervision steps further into this idea by adding new layers of 2D/3D information over the video stream. This tool allows a user to view undergoing activities and overlay this with the final expected work, or synchronize this with other project management tools and compare real status with a conclusion schedule. The current prototype is based on two modules: remote system and user terminal, Figure 2 shows a schematic view of the system. The remote module is composed of four main elements: high resolution GPS, mobile/handheld computer, surveillance camera with one or two rotation axis and long duration battery, Figure 3 shows some of the proposals for the module. The module is small and is supported an adapted medium load tripod. Determining the camera point of view is crucial for the system to function, in this case, position is obtained through the GPS and “look-at” direction is specified by an orientation sensor coupled to the camera. The remote module can be placed in strategic fixed locations or be entirely mobile and constantly repositioned. Since it is not composed of any majorly expensive components it can be manipulated by inexperienced personnel. In a situation the GPS is not reaching good resolution the user can always manually adjust virtual camera position to compensate small tolerance errors. The terminal is composed by a 3D model viewer with a video stream background, and can be controlled by a Bluetooth remote (camera movement and layer adjustment) which makes easy to use in with large screens and projections. Figure 4 shows a photo of an access terminal during field tests.

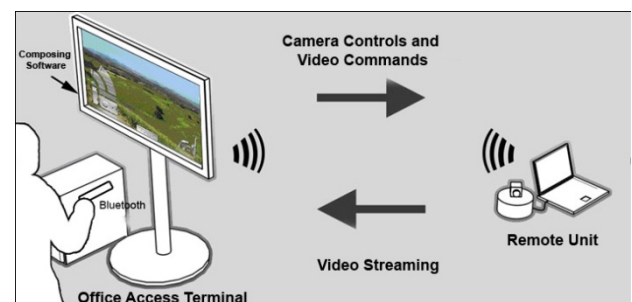


Figure 2: Schematic view of the Remote AR System.

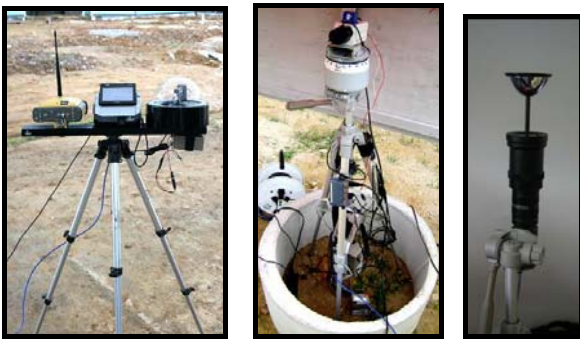


Figure 3: Mobile module prototypes.



Figure 4: Interface terminal for mobile camera system.

## 2.2. Outdoor AR Multipurpose

Outdoor AR has had much progress and is viable support tool ((Behzadan and Kamat 2005), (Boun Vinh, Kakuta et al. 2010), (Joonsuk and Jun 2010), (Karlekar, Zhou et al. 2010)). Precision, stability and reliability are currently dependent on wide variety of sensor technologies which comes in all shapes, sizes and prices. The NAV outdoor AR development has had the opportunity to experiment with a wide range of hardware combinations, outdoor image capture devices, interface designs, lcd screens and other performance related elements. The objective has always been to provide practical and reliable on site information system for engineers. The system specifications included low weight mobile hardware, touch screen interface, good visibility under strong sunlight, long battery life and the ability to use unconverted project 3D models.

The initial trials covered a wide variety of outdoor tracking systems using a selected group of image detection systems, such as, game controllers with sensors, electrical controls, fiducial marking, infrared lights and others which led to an illumination dependent and low resolution tracking system. Although, at this point, the outdoor mobile solution seemed hard to achieve using visual tracking, these activities led to the development of the Remote Panoramic AR system, based on digital servo controls already presented. The continuous experiments led to a different kind of sensor technology commercially available, the vision-inertial

self-tracking system (IS-1200)4. This sensor proved to be extremely reliable outdoors, but limited to areas which have visual fiducials. The new software developed on this technology had higher performance benchmarks on mobile computers, mostly because the sensor did most of the tracking within itself and passed the result to the software.

Tracking is an important factor, but there was also much discussion on how to give the CAD like look to models being seen on low performance computers. One approach was the render streaming library being developed by the GRVa software team, which still hadn't overcome bandwidth and resolution demands. Visualization experiments were done comparing formats and rendering engines, which led to consider the shortest path between the CAD system and the mobile AR, in this case it was perceived that OGRE3D provided a good similarity with original models and needed few to no adjustments for export. This engine also had achieved good performance scores when used on mobile computers. The coordinate systems between sensor and 3D space integrated smoothly with the need for conversions.

Video Capture was also an issue as the outdoor area around the LAMCE expansion is usually under a strong sunlight. This extreme light element showed that most of the web cameras used for indoor AR could not compensate or reduce exposure to provide a good image. Tests with several cameras and filters brought other issues, related to the size and aesthetic and ergonomic issues related to cameras attached to mobile computers. The current solution uses a high quality web cam, with an altered casing, allowing easy attachment and removal from the computer. The current camera model in use is a Microsoft HD-5000.

The specification of the mobile computer fitted the description of a typical TabletPC, but the main problem any computer outdoors is screen visibility which needs a brightness of at least 400 nits (cd/m<sup>2</sup>). Although some vendors can change the display hardware to match specific needs, the screen usually needs to be at maximum brightness to meet visibility demands. The sensor and camera already use up a lot of energy (and produce a lot of heat), so keeping the screen brightness at maximum only adds up to a short battery life, even for a 9 cell battery. The energy issue is still a matter to be resolved. The current Tablet PC on field use is the DELL Latitude XT. Figure 5 shows the combined hardware being used in a field experiment; the final solution was heavier than originally anticipated.

The last aspect of the solution is the application itself, which combines camera capture a 3D rendering engine and interface design. The many resources being used were integrated using a multimedia software toolkit known as Adobe Director, and extensions were developed in C++ to include sensor support, camera video capture and Ogre3d engine. This integration was chosen so it would give freedom for the design team to work and experiment with interface possibilities, animations and visual ergonomic factors. The current

interface is fairly simple and has only layer visibility control. Figure 6 shows a screenshot of the application with a visible 3d model in outdoor environment. This outdoor tool has served as base for the NAV AR project management system toolset which is explained in the following topics.



Figure 5: Outdoor AR system for viewing large scale projects.

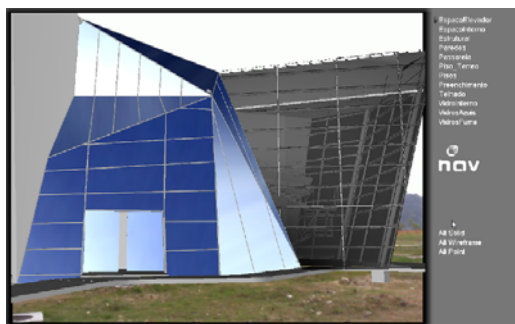


Figure 6: Outdoor AR screenshot, Expected versus Achieved Results.

### 2.3. AR Content Management (ARCOM)

Content management has always been a major setback to practical AR implementations. Content can also be considered a kind of visual annotation as observed by (Wither, DiVerdi et al. 2009). Currently there are many stable tracking options, as describe before, such as parallel, feature, fiducial and sensor tracking. Final build of AR applications have intuitive/ergonomic designs but, in most cases, have complex input, arrangement and removal of content, which forces users into manually editing parameter files which, in the best possible scenario, is in XML format. AR solutions, mostly exist as extensions for general multimedia development tools which are not developed to provide complete content management support.

The ARCOM system has been designed to fulfill a construction project demands for content management. The system is built considering a fiducial tracking system. The fiducial system was chosen for its application simplicity, implementation maturity and low cost. The ARCOM divides information in a hierarchic scenegraph which represents all project aspects which are relevant for on the field visualization. ARCOM

databases can be viewed on the field with the ARDI tool. Figure 7 and 8 shows the ARCOM system with multiple annotations being added.

The project is the root of the hierarchy and divides into work environment, afterwards each environment divides in interest areas. Each area is represented by a unique fiducial symbol which can have any number of information layers. The layers are then divided inf the following content groups: images, 3d models, audio, textual description, field comments (which can be photos, text or drawings). Figure shows the full tree of an example project opened in ARCOM. Every element can be accompanied by versioning information and be arranged in to match real objects. Every element can be viewed and arranged independently.

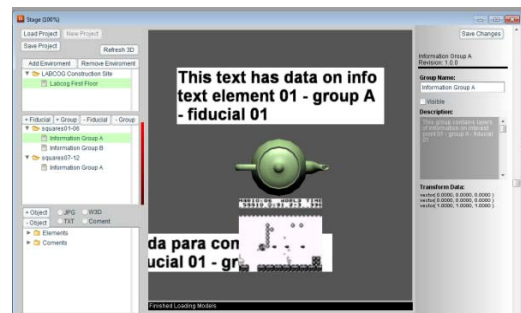


Figure 7: AR Content Manager Screenshot.

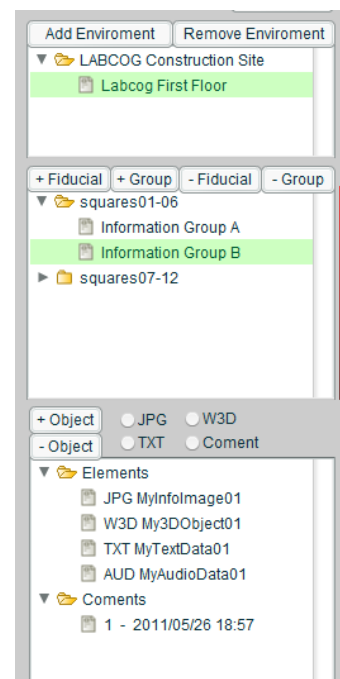


Figure 8: Hierarchical organization of project structure.

The ARCOM is also accompanied by a fiducial panel builder and pattern creator which helps create proper identification for interest areas. The last tool is camera configuration utility which helps quickly identify AR compatible cameras and set thresholds for different lighting conditions. Figure 9, 10, 11 shows screenshots of the AR support tools.





Figure 9: Pattern Builder.

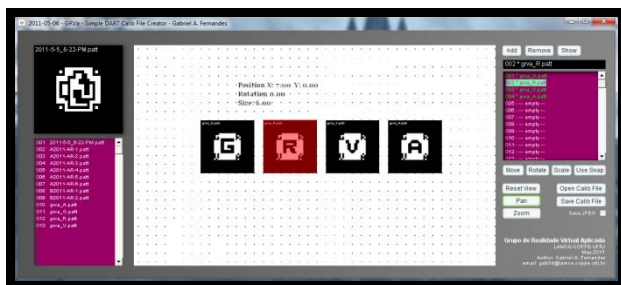


Figure 10: Panel Builder.

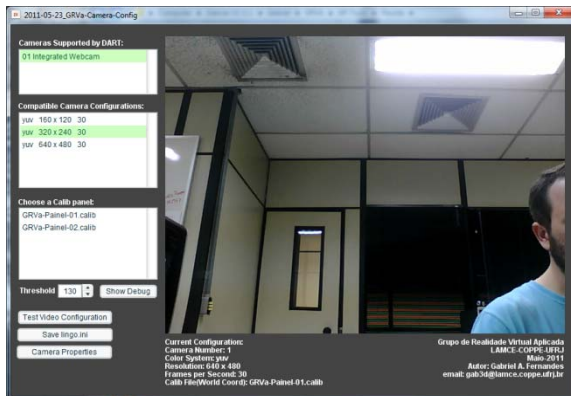


Figure 11: Camera Configuration utility.

## 2.4. AR Detailed Inspection (ARDI)

The AR detailed Inspection is a mobile access terminal for the content manager (ARCOM), its function is to present the visual database for supervision, management and inspection purposes. It is a project navigation solution where data is presented in tree like fashion and allowing users to dynamically add comments to previously defined interest points. These comments and observations are marked for content reviewers in real-time so, if necessary, proper action might be taken. The interface resembles the Outdoor AR application with upgraded layer/tree navigation options, as can be seen in Figure 12. The application is basically a viewer for XML files generated by the ARCOM.

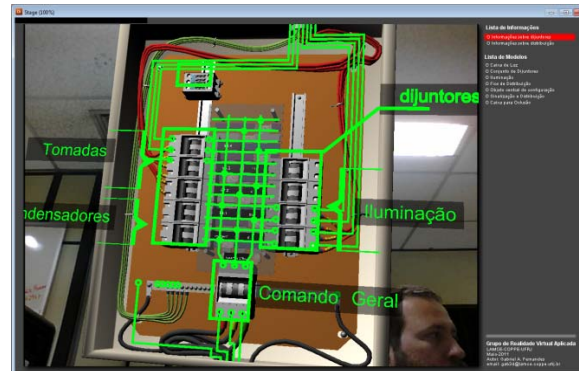


Figure 12: AR Detailed Inspection Screenshot: viewing a power distribution compartment.

## 2.5. Future AR management developments

The overall AR solution is still a work in progress, although much of its current form is based on years of field experiments and trials it still has a great deal of improvements pending. Some of the most immediate demands are development of the multiuser database, migration to a more sophisticated rendering system, support for a wider range of tracking solutions and the pen annotation system. The user database will be migrated to SQL server, where users will be able to access on-demand generated XML files. The rendering engine (in this case not Ogre3D) currently used is simple, although well integrated with 3d content creation tools, has many quality limitations, such as support for advanced shader language. As for the tracking, fiducials are very efficient and easy to use but there are many possibilities in the field of feature and parallel tracking that can be explored. The pen system is an upgrade to allow users to add annotations with pen or finger in the form of hand writing and drawings. Other directions for future developments could be the better use of low precision tracking devices such as cell phones and tablets to view spatially aligned annotations.

## 3. OTHER SUPPORT SOLUTIONS

The following items are ramifications of our project management solutions and make part in the overall NAV experience, but are not directly linked to the Augmented Reality support system. They provide content interaction options which are simpler to apply but still provide important project support.

### 3.1. Telepresence

Viewing large construction projects is quite a stressful task require time and manpower. Covering large areas for inspection is not a comfortable or easy assignment, to overcome costs for this activity the NAV solution has proposed a panoramic telepresence system. This system is composed of two modules a panoramic video receiver and a remote immersive access terminal. With these tools a decision maker can easily overlook large areas in a short time span without having to physically be there. The first module is placed on a work vehicle, small AR Drone, be carried by workers or even be placed on strategic observation points. Although similar to the

remote AR solution, the panoramic system has high resolution cameras which can zoom-in at meter scale details. The remote module can be composed of stereo glasses for individual use or a large display wall in a meeting room.

### 3.2. Project Information Management

The collaborative information management system is a friendly interface to manage changes in project schematics. Usually composed of 2D Cad drawings, project information is analyzed and evaluated in printed form where observations can be easily proposed and registered. Afterwards, in a well organized work flow, these changes are taken to a chain of employees for review before it is ultimately approved and effectively implemented in the project files, becoming the original print version. Due to new tablet PC technologies some more demanding operations already use mobile computing to make on-the-field observations of CAD drawings in a pen and paper style through PDF readers (Figure 13). Figure 14 shows a user interacting with the system through a touchscreen mobile computer. The information management system takes this proposal a step further by including a sophisticated versioning platform where observations can be sent in real-time to a plant server and possible schematics changes are alerted to all interested parties. The result is a more efficient and dynamic error correction and project modification pipeline which could save resources such as materials, time and rework. The system also marks observations in combination with a GPS positioning device which makes it possible to search the database for geo-referenced information. Below is a description of the main solution aspects: User management, geo-reference information retrieval, collaborative versioning and user map.

User management is achieved by an encrypted login system which divides access into five groups: readers, reviewers, editors, project manager and administrators (Figure 15). Each user category has limited access to features and is directed to fulfill a predefined task. The reader can access approved information sheets mainly for task execution. A reviewer can make observations through a note system directly in the virtual printed CAD file, just as he would over paper, but he can only comment the latest version available. An editor can add files and create versioning trees which can be viewed by all parties. A manager can create new project database and manage overall content. An Administrator can make any sort of change on the database and has the privilege of permanently deleting files.

Data access is divided in three search engines: the common word search, the geo-referenced search and a data project tree search. The common word search allows the user to input terms and retrieve files with similar expressions. The geo-reference allows the user to input GPS coordinates and retrieves localized project files, the input can be the current location, manually inputted coordinates or select a point on a map. The tree

allows the user to navigate through the project area files based on a pre-defined hierarchy generated by the project manager. The Collaborative versioning system alerts editors to errors observed on-the-field, then provides ways to update the files through a versioning platform much like a SVN. The user map is an extension screen within the solution which allows the current user to view other registered users positions on the field (Figure 16). Considering that all users have gps devices attached with their tablets. This way the workers trajectories and work paths can be recorded and/or monitored in real time.

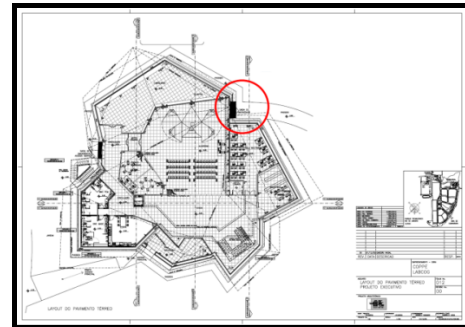


Figure 13: Screenshot of commented project file



Figure 14: User interacting with the plant manager.



Figure 15: Manager Web Portal.



Figure 16: GPS map system for geo-referencing observations and user mapping.

## REFERENCES

- Behzadan, A. H. and V. R. Kamat (2005). Visualization of construction graphics in outdoor augmented reality. Proceedings of the 37th conference on Winter simulation. Orlando, Florida, Winter Simulation Conference: 1914-1920.
- Boun Vinh, L., T. Kakuta, et al. (2010). Foreground and shadow occlusion handling for outdoor augmented reality. Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on.
- Castle, R. O., G. Klein, et al. (2011). "Wide-area augmented reality using camera tracking and mapping in multiple regions." Computer Vision and Image Understanding **115**(6): 854-867.
- Chen, K.-M., L.-L. Chen, et al. (2008). "Development and comparison of a full-scale car display and communication system by applying Augmented Reality." Displays **29**(1): 33-40.
- Hornecker, E. and T. Psik (2005). Using ARToolKit markers to build tangible prototypes and simulate other technologies. Human-Computer Interaction - Interact 2005, Proceedings. M. F. Costabile and F. Paterno. Berlin, Springer-Verlag Berlin. **3585**: 30-42.
- Joonsuk, P. and P. Jun (2010). 3DOF tracking accuracy improvement for outdoor Augmented Reality. Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on.
- Karlekar, J., S. Z. Zhou, et al. (2010). Model-based localization and drift-free user tracking for outdoor augmented reality. Multimedia and Expo (ICME), 2010 IEEE International Conference on.
- Kim, J. and H. Jun (2008). "Vision-based location positioning using augmented reality for indoor navigation." Ieee Transactions on Consumer Electronics **54**(3): 954-962.
- Kim, M. J. and M. L. Maher (2008). "The impact of tangible user interfaces on spatial cognition during collaborative design." Design Studies **29**(3): 222-253.
- Klein, G. and D. Murray (2007). Parallel Tracking and Mapping for Small AR Workspaces. Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on.
- Klein, G. and D. Murray (2009). Parallel Tracking and Mapping on a camera phone. Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on.
- Petit, B., J.-D. Lesage, et al. (2009). Virtualization gate. ACM SIGGRAPH 2009 Emerging Technologies. New Orleans, Louisiana, ACM: 1-1.
- Shin, D. H. and P. S. Dunston (2008). "Identification of application areas for Augmented Reality in industrial construction based on technology suitability." Automation in Construction **17**(7): 882-894.
- Wang, X. and P. S. Dunston (2006). "Compatibility issues in Augmented Reality systems for AEC: An experimental prototype study." Automation in Construction **15**(3): 314-326.
- Wither, J., S. DiVerdi, et al. (2009). "Annotation in outdoor augmented reality." Computers & Graphics **33**(6): 679-689.
- Zhuo, C. and L. Xinyu (2010). Markless tracking based on natural feature for Augmented Reality. Educational and Information Technology (ICEIT), 2010 International Conference on.



# COMPETITION AND INFORMATION: *CUMANA* A WEB SERIOUS GAME FOR EDUCATION IN THE INDUSTRIAL WORLD

Marina Massei, Alberto Tremori  
*MISS-DIPTEM, University of Genoa*

Email {massei, tremori}@itim.unige.it - URL [www.itim.unige.it](http://www.itim.unige.it)

Alberto Pessina, Federico Tarone  
*Simulation Team*

Email {alberto.pessina, federico.tarone}@simulationteam.com - URL [www.simulationteam.com](http://www.simulationteam.com)

## ABSTRACT

This paper describes *CUMANA*, acronyms for Cooperative/Competitive Utility for Management and Advanced Networking skill Acquisition, a model developed by the Simulation Team for educating the industrial and business communities to operate in a competitive-cooperative environment assigning the proper value to information and sharing data with others players, but with the final goal to win the game considering both individual and common objectives. *CUMANA* is an agent driven simulator, distributed via web and based on serious game approach.

Keywords: *agent driven simulation, cooperative-competitive environments, serious game, web simulation*

## 1. INTRODUCTION

Information and competition: two key-words in our society; from every day life to business and industry it is necessary to deal with enormous amount of information and to use these data in the proper way to compete and success: traffic news for not being late in the office, chose the best product in the supermarket to save money, etc.

This is particularly true in business: a new competitor is coming, problems with failures or delays in production, new potential markets; to identify these emerging phenomena by available information is critical for company success.

In today companies are dealing with many challenges therefore a major question could be summarized as: what is main asset that a modern company need to have to successfully compete in international global markets? Surely there are several answers to this inquiry, but it is possible to summarize most of these concepts within few words: a complex network where knowledge is shared quickly and effectively by flowing information form one node to another one; these nodes are representing both internal and external resources.

So it is critical to support training and education of managers and executives in properly approach information sharing and scenario evaluation over competing/collaborative environment. The paper propose a specific web games that combines Business and Serious Games approaches and apply of Modeling

and Simulation (M&S) for this purpose; the solution presented is defined *CUMANA* in reference to the mythological Cumaeen Sybil that was a famous prophetess in ancient Rome providing sentences as forecasts of the future that need to be “decrypted” by decision makers; a famous example comes from Sybil sentence “ibi redibis non morieris in bello” given to a warrior moving to the battle: this Latin sentence change radically its meaning by moving the “comma” (by the way, not included in the phrase), in fact the translation could evolve from “you will go and will be back, not dying in war” to “you will go and you will not return, dying in war”; this example demonstrates that the decision makers need to process information carefully in order to take right decisions and this is exactly the case of the *CUMANA* game.

The paper provide an overview on previous similar experiences developed by the authors with agent driven simulation used for web serious games.

The authors propose the detailed description of *CUMANA* model and architecture as well as the results obtained from its early experimentation phases . The obtained results confirm the potential of using *CUMANA* in different areas as well as the opportunity for further developments of this kind of simulators; in fact, the authors are currently completing *CUMANA* developments in order to summarizes related main findings and they are moving forward for activating new researches based on this simulation framework. *CUMANA* is currently available as a Simulation Team resource for experimentation and training ([www.mastsrl.eu/solutions/cumana](http://www.mastsrl.eu/solutions/cumana))

## 2. INDUSTRIAL AND BUSINESS FRAMEWORK

As anticipated a company is in fact an organization that could be represented by a complex network where data are distributed, processed and actions and decisions are make by different entities; this network involve both *internal* and *external nodes*; the *internal nodes* are the different structures, organization, division of a company (see fig. 1); for instance these includes among the others:

- Research and Development;
- Design and Engineering;
- Production;

- Logistics;
- Services and Maintenance;
- Commercial (Sales and Marketing)
- Administration
- Human Resource Management
- Subsidiaries Branches:
  - Commercial;
  - Industrial and Production;
  - Service.

This is particular true for big companies.

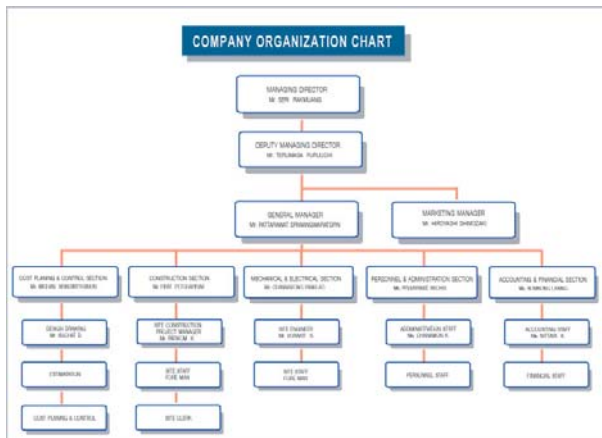


Fig. 1 Sample "Internal Nodes": Company Organization Chart

In addition there are the *external nodes* that are dealing not only with customers or dealers, but even with suppliers, allied companies; in fact today, the presence of a large number of external entities it is very common both for big companies and SME (small and medium enterprises). Among these different kind of external collaborations, not only related to classical commercial activities on markets, it is possible to list among the others the following elements:

- Commercial Agreements;
- Joint ventures,
- Suppliers
- External Service Providers for engineering, installation or maintenance;

In fact, it is possible to consider even all the entities with which there is no kind of hierarchical relationship. Therefore, in this paper it is presented the researches that have lead to the development of a model that reproduces this kind of Industrial and Business Scenario for educating people to be effective and efficient. The name of the model is *CUMANA*, acronyms for Cooperative/Competitive Utility for Management and Advanced Networking skill Acquisition: it re-creates operative conditions of a day by day activity of different divisions of a big company with new events and problems. The goal is to educate people to face and solve problems in cooperation, without forgetting the specific objective assigned to each individual. For this purposes *CUMANA* has been developed following the approach of combining Business Games and Serious Games: the main idea is that the learning experience is based on playing a game

with other players and that this approach support a clear understand of the correct mechanism for operating in the real world. *CUMANA* is a web-based simulation model able to facilitate training of distributed teams.

*CUMANA* Web Game provides the opportunity to play interactively a cooperative/competitive game, in a distributed environment where different "Managers" operate concurrently with benefits and penalties connected to both common and individual objective achievements related to their role in their Corporation.

The accreditation and the early phases of experimentation of *CUMANA* has been run with the support of different international teams, in fact nowadays, the simulator have been involved in playing games and providing feed backs people from

- Boeing, Seattle (Washington)
- University of Nebraska, Omaha (Nebraska)
- CSC Training Center of Excellence, Orlando (Florida)
- University of Genoa (Italy)
- MAST srl Genoa (Italy)

### 3. BUSINESS GAMES AND SERIOUS GAMES

A Business Game is a simulation based exercise about company management; in this context the players have to manage a "virtual entity" (i.e. a company or a division) operating in a competitive market. They have to take decisions and to apply methods in order to achieve their goals. It is correct to define it as a game because different players have to compete to achieve usually a common goal and it is based on simulation because it always implies economic models to simulate company and market behaviour (sometime very simple conceptual models, sometime computer simulators).

According to David Kolb *Experiential Learning Model* (Kolb and Fry 1975) the learning process is composed of four main steps (see fig. 2):

1. concrete experience,
2. observation of and reflection on that experience,
3. formation of abstract concepts based upon the reflection,
4. testing the new concepts,

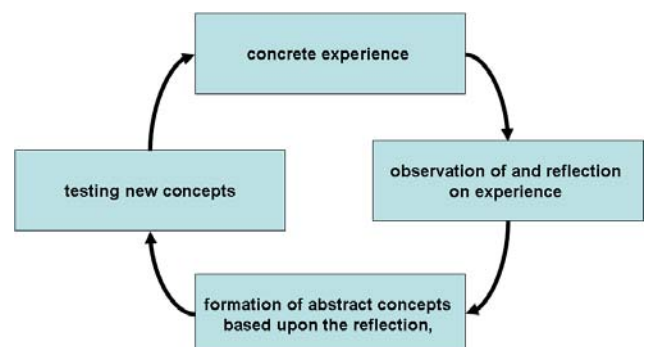


Fig. 2 Kolb Experiential Learning Model

Based on this model, it is clear the importance of business games for management training and/or potential development: they allow to iteratively proceed along all the above mentioned four steps. Business Games are a very powerful tool for developing skills and analytical capacity for decision-making, in knowledge development and also to change individual and personal attitudes. Business Games allow dynamic and active experiential learning reducing risk in the real world operation for new managers that need to face challenges and evolving threats.

An evolution and, at the same time a generalization, of Business Games is represented by Serious Games (Abt., 2002; Michael and Chen, 2005; Bergeron, 2006; Iuppa and Borst, 2006). Based on application of Simulation and new technologies (see fig. 3) they allow to better involvement of users. Serious game is a term used to refer to a software or hardware application developed with game technology and game design principles for a primary purpose other than pure entertainment. Serious games include games used for educational, persuasive, political, or health purposes; often serious games are derived from entertainment computer game engines.

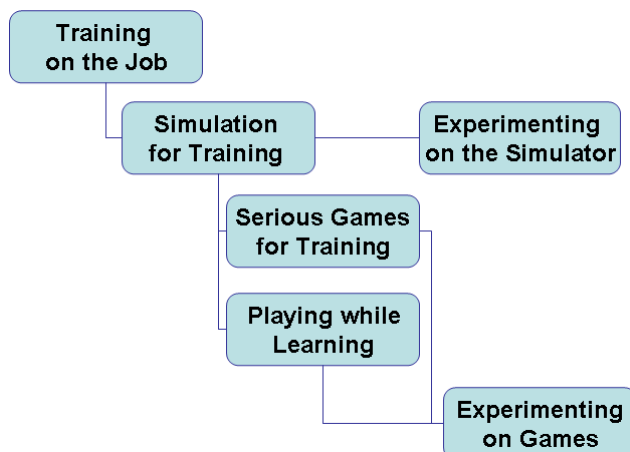


Fig. 3 Serious Games and Evolution in Training

In fact, Serious Games provide an opportunity to improve performances with reduced efforts to professional simulation with great attention to Interfaces, Graphics, Story and Emotional Involvement as well as New Technologies. Obviously the games frameworks to be used in training should be verified and validated in order to guarantee proper fidelity and correct models for avoiding negative training effects; therefore it is critical to define their use focusing on right involvement of the users and on the training objectives.

The potential application areas for these games usually introduce the necessity to add Artificial Intelligence (AI) able to add realism by introducing Human Behaviours; it is common to develop Intelligent Agents (IA) able to drive several of the active entities in the game; in fact this is one of the research areas in which the authors have great experiences as described in the following paragraph.

#### 4. PREVIOUS EXPERIENCES AND EXPERIMENTATIONS

Authors have several experiences in application of M&S in different areas and sectors: Defence, Security, Production, Logistics, ect. (Bruzzone et al. 1998; Bruzzone et al. 2000; Mosca 1994) Among these experiences there are some initiatives that represent a fundamental back ground for *CUMANA* (Bruzzone et al. 2000).

In fact since first middle of '90 the authors was involved in web games based on modelling and simulation for training applied to homeland and financial security personnel (Mosca et al. 1996); the evolution of web based simulation provided further opportunities to create applications in this area (Bruzzone, Page, Uhrmacher 1999).

A major opportunity to experience these approaches in business was provided by *IEPAL Project* (Elfrey., Bruzzone et al. 2005), acronyms for Intensive Educational Program in Advanced Logistics; this initiative was funded by the US Dep. of Education and by European Community FIPSE. IEPAL was a Transatlantic Educational Program in which several students from different Universities was involved operating both within Academic and Industrial frameworks in Joint International MultiDisciplinary Teams. In fact as the title suggests IEPAL was focused on logistics, but it was an important experience in managing educational packages for distributed, international teams. In the project several international Universities, Research Centers and Companies were involved: among the other Magdeburg University (Germany), Genoa University (Italy), Marseille University (France), Stevens Institute (New Jersey), Boston College (Massachusetts), UCF (Florida), NCS (Company and Agency Consortium for R&D in USA), CFLI (Company Consortium on Logistics in Europe).

Based on this experience, and with the support of several partners involved in IEPAL, in 2008 the authors started a new project for studying impact of new technologies for education. The results of these research was *SIBILLA*, Simulation of an Intelligence Board for Interactive Learning and Lofty Achievements, a web based Serious Game studied for Homeland Security Purposes. *SIBILLA* Game (Bruzzone, Tremori et al. 2009) provides the opportunity to play interactively a collaborative/competitive game, in a distributed environment where different "Intelligence Agencies" operates concurrently with benefits and penalties connected to common and individual objective achievements. *SIBILLA* is multiplayer web strategy game that simulate Terrorist Actions organized by different organization directed by Intelligent Agents that plan, prepare and execute attacks on specific location, site, time and threat type

The development of Intelligent Agents for representing the terrorist behavior *SIBILLA* had benefits form other research projects that the authors carried since the '90s in the area of Agent Driven Simulation and Human

Behavior Modelling (Mosca et al. 1995; Avallé et al. 1999). In this paper are briefly mention only the following models: ROSES, PIOVRA and IA-CGF.

ROSES (Bruzzone et al., 2003 Bruzzone et al 2004) was developed for the Research Branch of the main Italian Ship Constructor, and it is devoted to create an Oil Spill Simulator, including countermeasure models: in particular the rescue vessels for deploying countermeasure against the spill were totally autonomous and managed by the first generation of Intelligent Agents.

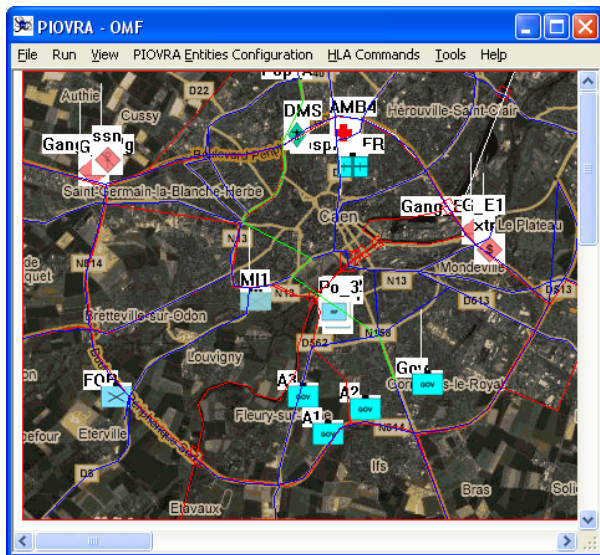


Fig. 4 PIOVRA demonstrator involving Intelligent Agents

In 2001 authors were involved in research in Human Behavior Simulation (Bruzzone 1996, Bruzzone et al., 1998; Bruzzone, Mosca et al., 2001; Bruzzone, Massei et al, 2004;) or reproduction of the Humans by using computer models. Usually this requires to simulate aspects related to Emotions, Rational Thinking, Psychology, Ethology and Sociology with the details required by the specific Modeling & Simulation environment. In 2007 PIOVRA (*Polyfunctional Intelligent Operational Virtual Reality Agents*), an EDA funded Project involving even Italian and France MoD, was completed (Bruzzone et al. 2004; Bocca et al. 2007; Bruzzone, Massei 2007). One of the main goals of PIOVRA was to develop a new generation of Computer Generated Forces (CGF) able to simulate "Intelligent" and "Autonomous" behavior. PIOVRA CGF demonstrated "Intelligence", in term of co-operative and competitive behaviors (coordinating units both during operative actions and situation evaluation) based on their configuration and nature, the perception of the scenario, order received and Rules of Engagement (ROE). Deliverables (see fig. 4) of this project was a new generation of CGF including human factors interoperating within an HLA Federation and resulting in being re-usable units for other interoperable simulators. In fact after this project the authors involved several entities active in the Simulation Team and developed IA-CGF (Intelligent Agent CGF) a new

generation of intelligent agents including special human behaviour libraries, special entities and units as well as non conventional frameworks devoted to simulate specific cases (i.e. earthquakes, civil military cooperation, urban warfare, piracy etc.) (Bruzzone 2008; Bruzzone, Massei 2010). In fact often human factors have a strong impact on Complex System and due to the new advances (including Games) it becomes possible to face this problem: i.e. Behaviour Collection in Massively Multiplayer On-Line Games or web games like *SIBILLA* or *CUMANA*; in fact both simulators get benefits from the use of IA-CGF modules.

## 5. CUMANA: THE SIMULATION MODEL

The *CUMANA* simulator is creating a business scenario with the following educative goals for managers:

- Learning to cooperate with the different players involved in the game by sharing the information received by individual resources in term of:
  - Event Location
  - Event Type
  - Affected Products
  - Affected Company within the Corporation
- Learning to use effectively each info element for taking decisions such as preventive actions to mitigate risks, to maximize opportunities and to avoid a possible future crisis
- To Learn how to properly evaluate the value of available information (directly received and/or shared) in order to share with other players
- Learning to develop trustiness and to estimate and manage reputation while exchanging data and knowledge with other players in respect of individual and common goals
- To develop negotiating skills
- 

The general architecture of the game (see fig. 5) is based on the interaction among the different players connected via web and intelligent agents that are managing every critical event and the distribution of information.

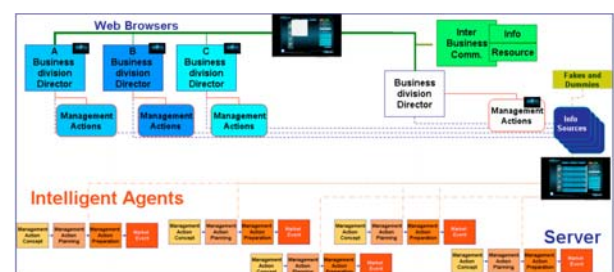


Fig. 5 CUMANA Architecture

In fact *CUMANA* is based on a complex framework of information (see paragraph 5.1) to be distributed among the players based on the stochastic engine and to be received back from players and eventually elaborated and re-used.



The first goal of every player is to forecast critical events on the market. Every event predicted means rise score in the game. For doing that it has been created a simple, but robust, system that creates and direct the events. During simulation execution *CUMANA* takes care of :

- Generating new critical issues: it is randomly assigned to an agent the responsibility of creating and preparing a new critical issue that will arise along time in a critical event (i.e. after n-time); the critical issue and its parameters are randomly generated by Montecarlo techniques; each critical issue need to predicted by players before it is activated
- The agents proceed in planning and preparing the critical event (i.e. launching a new competing product over a geographic region); during their activity, randomly spill of information are generated and distributed among the players; each spill contains a maximum of different bit of information related to this activity usable to predict the incoming event (i.e. where, what, why, who, target), therefore these info are split and distributed among the players partially based on their available resources and on stochastic factors.

For instance we have this event “...a new product will come in the Chinese market, it will have a particularly fancy design that will affected this kind of goods from our company FBKF Paris Ltd...”. Here below a partial scheme questions/Answers:

Questions	Answers
What's happen	New Competitive Product
Which product characteristic is affected	Design
Which geographic area	China
Which company will be attacked	FBKF Paris Ltd

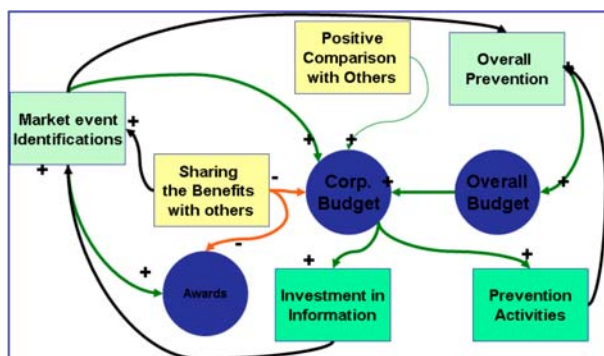


Fig. 6 Awards and penalties

In the flow of information regarding this event are created and distributed also fake information; so every player has some good and some bad data about an event and with this information he has to understand and forecast the whole event.

Usually the first reaction of players is to data mine their data and generate forecasts and actions; therefore the game goal is to educate people to share information to improve the overall performance of the corporation. A mechanism of awards and penalties has been created and implemented in the system (see fig. 6) to assign awards for prediction and penalties for wrong forecast to individual players as well as to the whole system.

## 5.1. Information DB Structure

The information are structured in a DataBase (DB), here below a brief description of the logical structure of information (fig. 7).

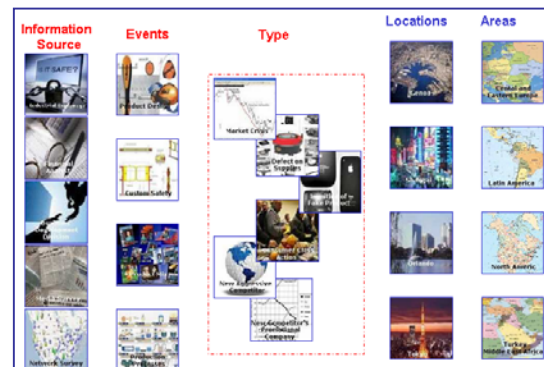


Fig. 7 Structure of Information

- **Info Sources:** different sources of information are available for players that are entitled to allocate their budget on these source for improving their data collection capability; each source have specific probability to intercept/detect info spill on the critical events; these source includes:
  - Customer Reports
  - Financial Analysts
  - Strategic Consulting
  - Internal Report
  - Network Survey
  - Industrial Espionage
  - Media Survey
  - R&D Division
- **Companies:** name and probability for every company that can be affected by the critical event
- **Company Area:** company directorate affected by the event, for instance
  - Product Design
  - Packaging
  - Logistics
  - Quality
  - Service
- **Geographic Area:** geography area affected:
  - Greater China
  - Asia Pacific
  - North America
  - Latin America
  - Western Europe
- **Type of Event:** event to be predicted such as:
  - New Competing Product

- New Competitor's Promotional Campaign
- Market Crisis
- Defects on Supplies
- New Aggressive Competitor
- Event\_PL: planning actions carried out by the agents generating spill of information on critical events such as :
  - Meeting
  - Phone Interception
  - Detection Information on Custom
  - Email Interception
  - Web Meeting
  - Teleconference
  - Information Request
  - Creation Business Unit
  - Detection of Business Plan
  - Accessing Information on Product Design
  - Accessing Information on Marketing
- Event\_PR: preparation actions carried out by the agents generating spill of information on critical events preparation events such as:
  - Information Stolen from a Company
  - Crisis Trade
  - Economy Trade
  - Computer Security Violation
  - Password Sniffing
  - Suspicious Economy Signal
  - Personnel Abnormal Behaviour
  - Business Plane Disappeared
  - Contract Lost
  - Logistic Information Lost
  - New Logistic Planning
  - New Customs Laws
- Events and Events 1: “support” sheets with all the events generated at system initialization collecting information from all the DB fields (see fig. 8).

1	Detection: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design)
2	Phone Interception: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
3	Money Transfer: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
4	Money Transfer: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
5	Money Transfer: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
6	Personnel Abnormal Behaviour: Location (Name)
7	Suspicious Economy Signal: Location (Name), Possible Target (Product Design)
8	Personnel Abnormal Behaviour: Location (Name), Possible Target (Product Design)
9	Personnel Abnormal Behaviour: Location (Name), Possible Target (Product Design)
10	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
11	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
12	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
13	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
14	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
15	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
16	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
17	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
18	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
19	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
20	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
21	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
22	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
23	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
24	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
25	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
26	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
27	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
28	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
29	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
30	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
31	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
32	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
33	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
34	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
35	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
36	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
37	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
38	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
39	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
40	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
41	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
42	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
43	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
44	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
45	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
46	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
47	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
48	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
49	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
50	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
51	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
52	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
53	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
54	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
55	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
56	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
57	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
58	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
59	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
60	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
61	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
62	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
63	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
64	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
65	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
66	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
67	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
68	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
69	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
70	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
71	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
72	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
73	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
74	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
75	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
76	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
77	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
78	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
79	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
80	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
81	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
82	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
83	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
84	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
85	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
86	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
87	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
88	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
89	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
90	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
91	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
92	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
93	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
94	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
95	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
96	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
97	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
98	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
99	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)
100	Teleconference: People Involved (Name), Terrorist Group (Location), Possible Target (Product Design), Possible Attack (New Computing Product)

Fig. 8 Sample of Events Sheet

## 5.2. CUMANA Game

CUMANA Game operates by accessing the web within a browser; the CUMANA GUI (Graphic User Interface) allows to access a specific game and to play; in fact the server is able to manage multiple games with several players concurrently. In figure 9 it is presented a sample of the of the CUMANA desktop. The game include a chat area for sharing info and for their negotiation; in chat it is possible to select what player(s) involve and to arrange money transfer.



Fig.9 CUMANA desktop player

CUMANA provides different frames to use the different resources and manage decision; for instance Prevision Area is where the information are collected for building forecasts; time management in the game could be based on turns or on clock time, the different modes are selected when a new game is initialized; the players need to manage their budget and their cash for allocating resource players.

In figure 10 it is proposed the failure message deliver to player for missing a critical event prediction; this event affect the whole corporation and all player budget are downsized.



Fig. 10 Sample of Missed Prediction Message

Viceversa, figure 11 propose the message for correct critical event prediction; in this case the whole corporation get benefits and all player budget are increased, therefore the overall budget shared evolve to provide a larger slice to the players able to forecast the crisis and prevent it; in this way it emerge a clear individual and common goal for all the CUMANA users over a specific game.

## 6. ACCREDITATION AND EXPERIMENTAL PHASE

Authors have completed a first preliminary experimental phase on CUMANA with the involvement of different international subjects. In this paper are summarized the tests conducted in the USA in:

- Boeing, in Seattle Everett (WA)



- University of Nebraska, in Omaha (Nebraska)
- Training Center of Excellence, Lead Associate at CSC in Orlando (Florida)

For every test were played matches for preparation of players and one final match which results were recorded. Personnel from Simulation Team - DIPTM University of Genoa was directly involved for providing training, tutoring and support during exercises. The results of Boeing test was also used by the company for evaluation of an international team of young engineers during their internship. In Table 1 the results of best players are summarized.

Player	Final Score	Successful Forecasts	Forecasts
Player 1	1450	1	1
Player 2	1325	1	1
Player 3	900	0	0
Player 4	750	0	0
Player 5	700	0	1
Player 6	650	0	2
Player 7	525	0	2

Table 1 Final Results First test in Company

In the second test a team it was involved a group of Industrial Engineering students of the Last Year from University of Nebraska. It was a numerous team and the test was pretty intensive and required a big effort for tutoring people and collecting information.

Player	Final Score	Successful Forecasts	Forecasts
Player 1	1350	3	3
Player 2	1325	3	4
Player 3	1320	3	4
Player 4	1300	2	5
Player 5	1300	2	7
Player 6	1020	1	5
Player 7	1010	1	4
Player 8	1010	1	6
Player 9	990	1	8
Player 10	965	1	7
Player 11	945	0	2
Player 12	900	0	4
Player 13	800	0	5
Player 14	750	0	6
Player 15	600	0	6
Player 16	555	0	2
Player 17	550	0	6
Player 18	450	0	7
Player 19	425	0	8

Table 2 Final Results second test in University

The last test was made in the Training Center of Excellence, CSC. In CSC the team was a mix people coming from the company and academia,



Fig. 11 Correct Forecast

Player	Final Score	Successful Forecasts	Forecasts
Player 1	1020	1	1
Player 2	990	0	1
Player 3	900	0	0
Player 4	880	0	0
Player 5	860	0	0

Table 3 Final Results of third test in Training Center

At the end of every test a questionnaire was completed by participants here below the average results.

Question	Vote
Quality of GUI (0 very bad – 10 excellent)	9,3
Usability (0 very bad – 10 excellent)	8,6
SW Speediness (0 very bad – 10 excellent)	6,6
Difficulties (0 no difficulties – 10 impossible)	2,0
Training Objective (0 very bad – 10 excellent)	8,8
General Vote (0 very bad – 10 excellent)	8,8

Table 4: average results from participants

In fact, despite the limited number of people used in these tests and the preliminary nature of this experimentation, the evaluation of CUMANA results very satisfactory. A lot of comments were also collected about duration of games or about the impact of this tool for educational purposes.

Authors are completing further analysis at the University of Genoa with the support of Italian companies and there are undergoing new tests for completing the evaluation phase of CUMANA and for improving the educational and training capabilities of this framework.

## 7. GAME USE AND FUTURE DEVELOPMENT

In fact, different evolutions and improvements of CUMANA are under investigation. For instance the Authors are studying the possibility to uses CUMANA, but also SIBILLA, in multi-scenario and multi-level games: an example of multi-scenario could be represented by “horizontal” concurrent games with both cooperative and competitive approaches in different Regions (see Fig. 12). Another possible extension in the use of CUMANA over a “vertical” approach with the

creation of chains of command / company multilayer organizations where different players or teams act as different hierarchical levels within a corporation exchanging information among the different levels and the different divisions. A combined approach of the two above described modes is also possible and could be studied and adopted in the future.

Other improvements of this family of Serious Games are under evaluation; in fact the authors are planning new researches to improve emphatic involvement of the player: it is very important to remember that this is one of the key success factor for serious games being the efforts of the users strongly motivated by these aspects that could maximize the educational impact of the games. Other tests are also under evaluation considering the collaboration of different user communities both in geographical and cultural matter. Very close to this concept the idea of using *CUMANA*, but *SIBILLA* as well, as experimental labs for other models. There are also possible developments related to authors' research tracks on Intelligent Agents and Human Modelling with new improvement on the autonomous behavior of entities generating events.

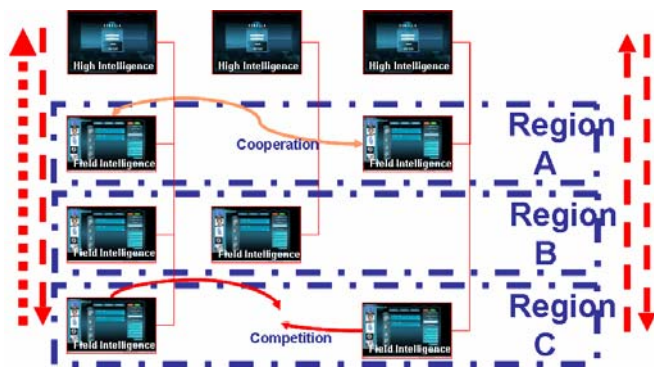


Fig. 12 Multi Level and Multi Scenario Games

On a more technological side there are future ideas of integrating *CUMANA* with Synthetic Environments or other multimedia content (video streaming, audio...) to improve realism of the game; special attention will be devoted to combine these games with Mobile solution using smart phones; in fact with currently technology, it is possible to develop new releases of this game engine for mobile applications such as for smart phones or PDAs for new improvement in the game (i.e. the multilevel approach described above)

## 8. CONCLUSIONS

This paper described the use of an agent driven, web simulator that applying theories of serious games for education and training of businessmen and managers; the main goal is to transfer skills in understanding the value of information and to transfer the concept that sharing data with colleagues empowers the system.

*CUMANA* has been developed based on several previous experiences of the authors in the industrial sector on the use of M&S in training and education and in the development of Serious Games and Agent Driven

Simulators. Authors used great attention first of all in the creation of the conceptual models that could represent the complexity of an extended enterprise with different areas and directorate and with operations over different geographies. The development of the game engine presented on the other hand difficulties due to the choice of a web distributed application and with the strategic goal of a tool that could be used form people behaving to different cultures, both from an academic or applicative and a international point of view. The first experimental phase showed good results form users and authors are using data from these international and multi-disciplinary tests for completing analysis of results of the use of this innovative kind of systems for education purposes. New tests are also under evaluation at Simulation Team and in cooperation with MISS DIPTM in the University of Genoa and in collaboration with industrial partners. The authors are currently involved in several extension of these models considering the great potential of the proposed approach and the results provided by *CUMANA* simulator.

## ACKNOWLEDGEMENT

The authors thanks Mike McGinnis, Judith Converso and Roberto Lu and their institutions and organizations for their support and availability in the preliminary experimentation phase of the project.

## REFERENCES

- Abt C.C., (2002) "Serious Games", University Press of America
- Argyris, C. (1958) "Some Problems in Conceptualizing Organizational Climate: A Case Study of a Bank", *Administrative Science Quarterly*, 2, 4, pp.501-20
- Avalle L, A.G. Bruzzone, F. Copello, A. Guerri, P.Bartoletti (1999) "Epidemic Diffusion Simulation Relative to Movements of a Population that Acts on the Territory: Bio-Dynamic Comments and Evaluations", *Proc. of WMC99*, San Francisco, January
- Axelrod R. (1981) "The Emergence of Cooperation among Egoists." *American Political Science Review* 75 (June): 306-18
- Bergeron B, (2006) "Developing Serious Games", Charles River Media
- Bennis, W. (2002). *Nobody in Charge: Essays on the Future of Leadership* Jossey-Bass.
- Bavelas, A. (1950) "Communication Patterns in Task-Oriented Groups", *Journal of the Acoustical Society of America*, vol. 22, issue 6, p. 725
- Bocca E., Pierfederici, B.E. (2007) "Intelligent agents for moving and operating Computer Generated Forces" *Proc. of SCSC*, San Diego July
- Bruzzone A.G. (1996) "Object Oriented Modelling to Study Individual Human Behaviour in the Work Environment: a Medical Testing Laboratory ", *Proc. of WMC'96*, San Diego, January

- Bruzzone A.G., Giribone P. (1998) "Quality Service Improvement by Using Human Behaviour Simulation", Proc.of ESM98, Manchester UK, June
- Bruzzone A.G. M.E. (1998) "Upgrading Civil Protection Systems: The impact of New Techniques on Emergency Management Training", Proceedings ITEC98, Lausanne, April 28-30
- Bruzzone A.G., Page E., Uhrmacher A. (1999) "Web-based Modelling & Simulation", SCS International, San Francisco, ISBN 1-56555-156-7
- Bruzzone A.G., Vio F., Capasso S. (2000) "MOSLES: Modelling & Simulation for Transport and Logistic Educational Support", Proceedings ITEC2000, The Hague, April
- Bruzzone A.G., Mosca R., B.M., B.C., Spirito F., Coppa A.C.M. (2001) "Advanced Modeling for Customer Behaviour Analysis ", Proceedings of MIC2001, Innsbruck Austria, February
- Bruzzone A.G., Mosca R. (2003) "Modelling & Applied Simulation", DIPTM Press, Genoa, ISBN 88-900732-3-3 (197pp)
- Bruzzone A.G., Orsoni A., Giribone P. (2003) "Fuzzy and Simulation Based Techniques for Industrial Safety And Risk Assessment", Proceedings of ICPR, Blacksburg VA USA, August 3-7
- Bruzzone A.G., Viazzo S., B.C., Massei M. (2004) "Modelling Human Behaviour in Industrial Facilities & Business Processes", Proc. of ASTC2004, Arlington, VA, April
- Bruzzone A.G., Massei M., Simeoni S., Carini D., B.M. (2004) "Parameter Tuning in Modelling Human Behaviours by Using Optimization Techniques", Proceedings of ESM2004, Magdeburg, Germany, June
- Bruzzone A.G., Figini F. (2004) "Modelling Human Behaviour in Chemical Facilities and Oil Platforms", Proceedings of SCSC2004, San Jose'
- Bruzzone A.G., Petrova P., B.M., B.C. (2004) "Poly-functional Intelligent Agents for Computer Generated Forces", Proceedings of Wintersim2004, Washington DC, December
- Bruzzone A.G., Williams E.(2005) "Summer Computer Simulation Conference", SCS International, San Diego
- Bruzzone A.G., (2007) "Challenges and Opportunities for Human Behaviour Modelling in Applied Simulation", Keynote Speech at Applied Simulation and Modelling, Palma de Mallorca
- Bruzzone A.G., Massei M. (2007) "Polyfunctional Intelligent Operational Virtual Reality Agent: PIOVRA Final Report", EDA Technical Report, Genoa
- Bruzzone A.G. (2008) "Intelligent Agents for Computer Generated Forces", Invited Speech at Gesi User Workshop, Wien, Italy, October 16-17
- Bruzzone A.G., A. Tremori (2008)"Safety & Security in Retail: Modeling Value Chain Dynamics "Proceedings Of BIS2008, April 14 - 17, 2008, Ottawa, Canada
- Bruzzone A.G., Elfrey P., Cunha G., Tremori A. (2009)"Simulation for Education in Resource Management in Homeland Security" Proceedings of SCSC2009, Istanbul, Turkey, July
- Bruzzone A.G. (2009) " Intelligence and Security as a Framework for Applying Serious Games", Proceedings of Serixgame, Civitavecchia, November
- Bruzzone A.G., Massei M. (2010) "Intelligent Agents for Modelling Country Reconstruction Operation", Proceedings of AfricaMS 2010, Gaborone, Botswana, September 6-8
- Cartwright D, Zander A. (1960) "Group dynamics: Research & theory", Tavistock, 1960, London
- Computer Science and Telecommunications Board-US National Research Council (1977) "Modeling and Simulation: Linking Entertainment and Defense", *National Academy Press*,
- Elfrey P. (1982) "The Hidden Agenda", Wiley & Sons, NYC
- Elfrey P., Bruzzone A.G., R.Mosca, A.Naamane, C. Frydman, N.Giambiasi, G.Neumann, D.Ziems, S.Capasso, R.Signorile, S.Ghosh, A.Nisanci, B.C.B.M. (2001) "Logistics & Education In IEPAL International Cooperation", Proceedings of HMS2001, Marseille October 15-17
- Iuppa N, Borst T, (2006) "Story and Simulations for Serious Games: Tales from the Trenches" Focal Press
- Kolb. D. A. and Fry, R. (1975) Toward an applied theory of experiential learning. in C. Cooper (ed.) Theories of Group Process, London: John Wiley
- Michael D, Chen S, (2005) "Serious Games: Games That Educate, Train, and Inform " Course Technology PTR
- Moreno J.L. (1947) "The Theatre of Spontaneity", Beacon House, NYC
- Mosca R., Giribone P. & Bruzzone A.G. (1994) "Simulation & Automatic Parking in a Training System for Terminal Container Yard Management", Proceedings of ITEC94, The Hague, April 26-28
- Mosca R., Giribone P. & A.G.Bruzzone (1995) "AI Techniques to Optimize the distribution of Fire-Fighting Squads during an Accident", Proc. of Simulators International XII, Phoenix, April 9-13
- Mosca R., Bruzzone A.G. & Costa S. (1996) "Simulation as a Support for Training Personnel in Security Procedures", Proc. of Military Government and Aerospace Simulation, New Orleans LA, April 8-11
- Nylen D., Mitchell J.R., Stout A.T. (1967) "Handbook of staff development and human relations training: materials developed for use in Africa", NTL Institute for Applied Behavioral Science, Washington D.C.

# Modeling Logistics & Operations in a Port Terminal for Environmental Impact Evaluation and Analysis

Marina Massei Francesca Madeo

Simulation Team University of Genoa, Italy

{marina.massei, francesca.madeo}@simulationteam.com

Francesco Longo

Simulation Team MSC-LES Univ. of Calabria, Italy

francesco.longo@simulationteam.com

**Keywords:** Simulation, Sustainability, Logistics, Environmental Impacts

## ABSTRACT

The research proposes a new methodology to analyze and evaluate the Environmental Impacts (EIs) and efficiency on Operations & Logistics by using Modeling and Simulation (M&S). In particular the authors developed models for analyzing these aspects and create the GreenLog (Green Logistic) simulation framework based on web technologies in order to define and evaluate environmental impacts (i.e. air emissions, power consumption, etc.) over supply chains and logistics networks. GreenLog allows to estimate quantitatively all the different EIs in order to evaluate different strategies and solutions for improving the supply chain in term of efficiency and sustainability, including productivity and costs aspects. As matter of fact, in this context, it is critical to identify effective and competitive solutions able to reduce EIs (i.e. resources consumption, garbage/waste disposals, noise, discharges, spills) without losing efficiency and competitiveness.

In this paper the authors propose a case study, in order to evaluate EIs within a port terminal and to compare different solution over different KPIs

## 1. INTRODUCTION

In recent years, people environmental attitudes are changing due to major attention and sensitivity towards environmental issues that are increasing health consciousness and motivation to preserve environment for the next generations (for instance consumers, for ecological reasons, prefer to stop buying products from companies that cause pollution) [1] [36].

Therefore many companies are moving toward more sustainable structures and infrastructures by valorizing their own environmental performances and sustainable solutions in order to correspond to users needs and requirements. In addition new governmental laws or international pacts (such as Kyoto Protocol, White Book on Transportation, ISO 14000 Series, European Regulation 761/2000-EMAS, Ecolabel) press the different economic sectors in increasing their regards for the environment [7] [26]. In particular as concern for the environment issues, companies must take into account climate changes, air pollution, noise, vibration and accidents as well as logistics costs in order to achieve a more sustainable balance among economic, environmental and social objectives [20].

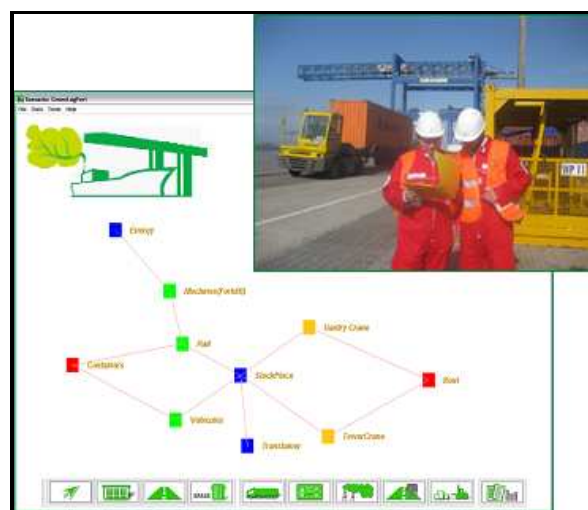


Figure 1. SimulationTeam staff member working on Collecting Data for GreenLog Port Simulator

A critical aspect is represented by Logistics due to the fact that it is related to all the activities required for moving products through the whole supply chain, from the raw materials acquisition to the product realization and distribution to the point of consumption and the associated reverse logistics [6]. As matter of fact, logistics represent a complex framework due to the many different activities and interactions with all the internal processes and external supply chain operations and it strongly impacts on environment by causing indirect costs and resources consumption [5] [14].

Therefore it is fundamental to identify supply chain management practices and strategies that allow environmental impacts and energy consumption reduction, but generally these solutions are evaluated based on experts feelings or qualitative analysis.

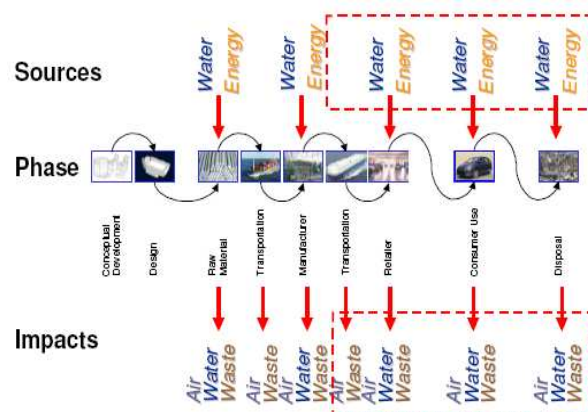


Figure 2. Source and Impact of Supply Chain



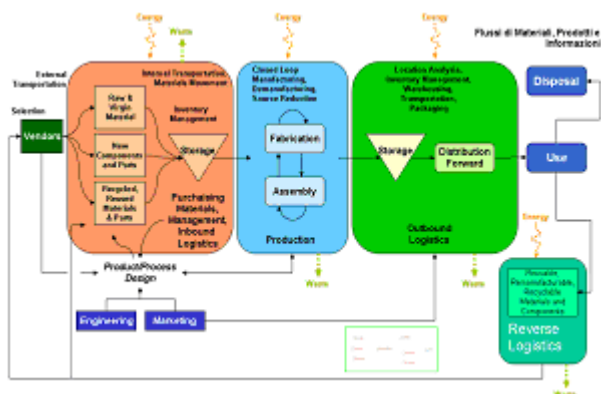


Figure 3. Logistics & Green Logistics Complex Framework

The authors propose a decisions support tool able to provide quantitative analysis and measures of the environmental impacts of the whole supply chain [1][17], by considering air emissions as well as waste disposals (i.e. rubber due to tire trucks consumption) terrain degradation, noise, spills, dusts etc. In addition the sustainability analysis includes many KPIs (Key Performance Indexes) such as that one related to costs analysis in order to find an efficient solution that considers both environment, quality, efficiency and economic aspects of the logistics networks[10].

In fact complex phenomena, different factors and processes (i.e. international regulations) need to be considered for environmental impacts estimation of the whole supply chain, so Modelling and Simulation is an ideal approach for investigating these complex system, with many components that interact each other [4] [9]. The authors propose an application for EIs evaluation in a Port Terminal by using the Green Logistic simulator (GreenLog) developed by the Simulation Team since 2008 to face these issues. GreenLog was developed originally within the project “Italian Green Logistics Initiative” that involved Industry, Academy and Governmental Institutions in order to provide innovative solutions for “greening” improvements for the logistics and for measuring environmental impact and costs efficiency on a specific supply chain.



Figure 4. GreenLog Web Interface

## 2. THE GREENLOGISTIC SIMULATOR

Simulation Team developed GreenLog Simulator for Analyzing Environmental Impacts in Industrial and Logistic networks including for instance Production and Distribution Centers, Storage and Terminals [8].

GreenLog is a Web Based Simulation Engine including models related to performance, costs and Environmental Impacts of the Productive, Logistics and Transportation Elements along the Supply Chain. GreenLog application is owned by Simulation Team and the GreenLogistics framework is available on MISS DIPTM web portal ([www.itim.unige.it/greenlogistics](http://www.itim.unige.it/greenlogistics)) where it is possible to access to several services:

- Company Qualitative and Quantitative Questionnaire
- EIs Self Measure of the Company Logistics
- Green Level based on automated configuration of the specific simulation model
- Supply Chain Simulation for measuring impacts and performances.

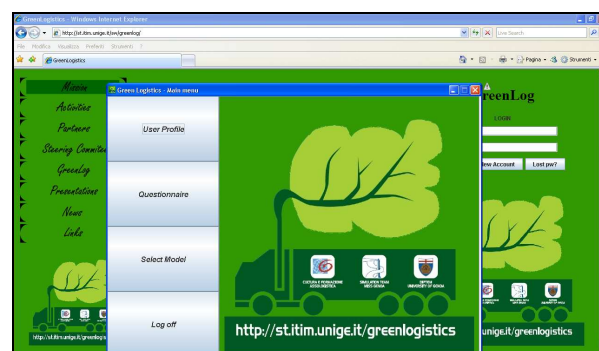


Figure 5. GreenLog Web Services

As further development, since 2008, the authors developed different modules focusing on specific aspects ([www.mastsrl.eu/solutions/greenlog](http://www.mastsrl.eu/solutions/greenlog)):

- GreenLog Port, devoted to support estimation of Environmental Impact in Ports Garbage & Port Waste, Dredging, Dust, Noise, Ship Air Emissions, Air Quality, Hazardous cargo, Bunkering, Port development, Ship Discharge
- GreenLog Ship, devoted to analyze the Environmental Impact of the Ship for supporting monitoring, alternative evaluation, saving and benefits from different solution in use, handling, operating as well as in Ship Design
- GreenLog Crane, devoted to analyze the Environmental Impact of Cranes and Handling Devices considering Operative Costs and Environmental Impact; to estimate the benefits provided by innovative solutions in term of power saving, oil spill reductions, better safety procedures and higher performances
- GreenLog Heavy Haul devoted to analyze the Environmental Impact of Trucks and Heavy Hauls considering Operative Costs and Environmental Impact.

These researches propose opportunities to apply GreenLog in different case studies; for instance Greenlog Port was applied to model EIs some Italian Port Terminal.

Figure 6. Example of GreenLog Web Questionnaire for Companies

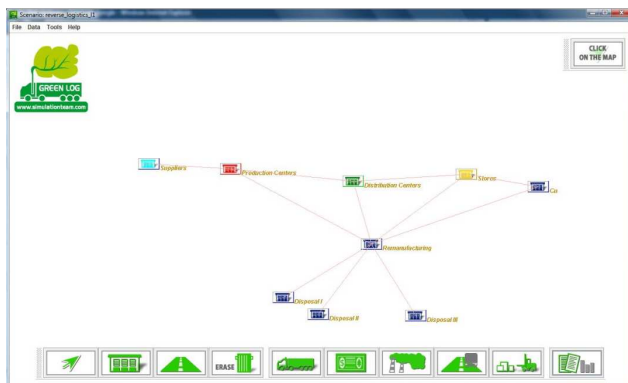


Figure 7. Scenario in GreenLog Simulator

Therefore the Green Log Simulation Models are based on the following main objects devoted to reproduce infrastructures, resources, processes and performances:

- Logistics Node (i.e. a terminal or a distribution center, a production center etc.)
- Logistics Link between two nodes
- Vector (i.e. truck, portainer, ship etc)
- Equipment (i.e. HVAC, Forklifts, etc.)
- Logistics Flows (goods and products)
- Environmental Impact (EI)

GreenLog users are allowed to create and configure the objects by Green Log GUI (Graphical Users Interface) that provides charts and diagrams reports and EIs visualization on the logistic network, in order to identify critical factors. In addition users are able to attribute Environmental Impacts to a node or to a link or to a vector and they could include emissions, disposals, power and water consumption etc.

Each EI can be constant or dependent on distance, time, flow volumes, flow type. Vectors characteristics (ships, trucks, trains, transainers, gantry cranes, reach stackers), Node information and Links definition (i.e. terminal to yard or to transhipment) are input data for the simulator. It is possible to visualize impacts over the whole supply chain by graphics and quantitative reports and to find out the critical path by graphical representation, in order to make hypothesis and find a solution to mitigate the Environmental Impacts.

### 3. ENVIRONMENTAL SUSTAINABILITY IN A PORT TERMINAL

The concept of sustainability is related to the policies and strategies for using resources in order to not deplete the resources themselves. Among most popular definitions of sustainability it is well referenced that one presented at United Nation Conference in 1987, defining "sustainable developments" as those that "meet present needs without compromising the ability of future generations to meet their needs"(WECD, 1987) [18].

Therefore economic and social development should take into account of natural land, water, and energy resources saving as integral aspects of the development.

In a Port Terminal the sustainability concerns with innovative solutions and technologies for greening existing port operations and pursuing green development opportunities. In particular the main Environmental Impacts are related to:

- Definition and adoption of Sustainable Engineering Design and Construction Guidelines
- Energy and Water Conservation Policy
- Waste disposals from ships and terminal operations
- Air Quality
- Training on Sustainability

It is possible to face these issues by implementing pollution prevention measures, by identifying, for instance, waste streams to be recycled and new technologies and solutions for minimizing negative impacts on the environment and surrounding communities, by applying clean air Actions, by green buildings based on renewable energy etc.

Figure 8. Environmental Impact Parameters

Figure 9 Nodes and Link Modeling



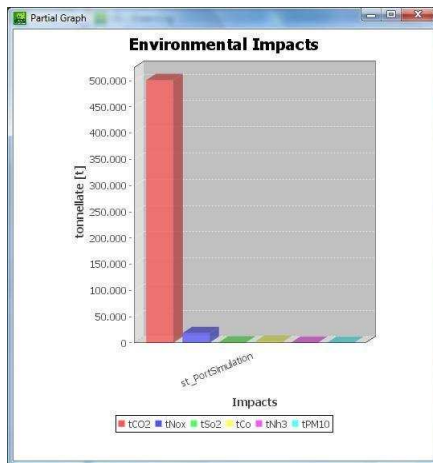


Figure 10. Environmental impacts

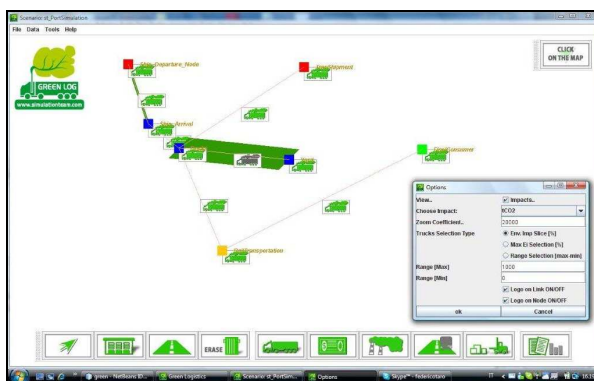


Figure 11. Impacts in Port Terminal Scenario

The authors propose a case study by analyzing:

- the environmental benefits of a new technology at the gate for control operations in order to reduce trucks waiting time and so air emissions at the gate
- the effect of rail infrastructures improvement
- the benefits of using low sulphur fuel for ships, instead heavy fuel while in operating in the port terminal

In particular the first aspect is very critical in terms of EIs. In fact the required waiting time for control operations at the gate could vary from 45 minutes to more than several hours by causing negative impacts on air quality due to trucks air emissions.

Therefore new technologies and methodologies are adopted to reduce control operation time such as the container pre-clearing that allows to know the container required information before its arrival.

The authors compared in their analysis traditional solutions, for instance based on a mechanical seal, respect new ones (i.e. use of an electronic seal based on reusable RFID tag) as well as different management strategies (i.e. sharable common tracks among different port terminals at the same gate).

Another important aspect is related to the shift from trucks to trains for import/export activities; in fact this is relevant from an environmental point of view. Obviously ship fuel is an important issue too and even already widely in evaluation; this third aspect has a great impact on Sulphur emissions.

As it will be described in the next paragraph, the authors developed a Port Terminal Model and analyzed EIs by using GreenLog Simulator in two different cases.

#### 4. PORT TERMINAL SIMULATION MODEL AND EXPERIMENTS

The authors developed a conceptual model for the Port Terminal to be analyzed by defining entities and environmental impacts characteristics.

A scheme of the proposed conceptual model is represented in figure:

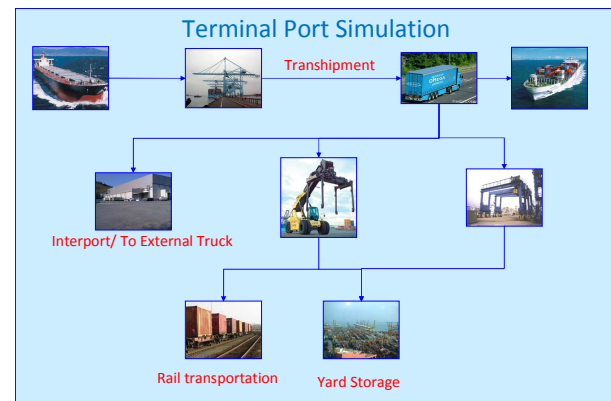
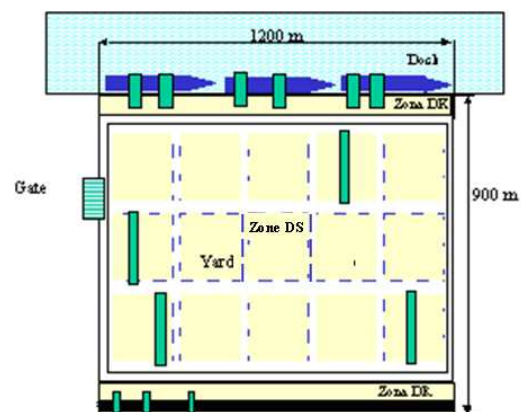


Figure 12. Port Terminal Simulation

The Port Terminal includes:

- Zone DK for Dock Operations
- Zone DS: Yard / Storage Area
- Zone DR: Rail Terminal



The flows are managed in accordance with the following directions:

*Import*

From Ship To Feeder Ship

Ship  $\Rightarrow$  DK  $\Rightarrow$  DS  $\Rightarrow$  DK  $\Rightarrow$  Ship

From Ship To Train

Ship  $\Rightarrow$  DK  $\Rightarrow$  DS  $\Rightarrow$  DR  $\Rightarrow$  Train

From Ship To Truck

Ship  $\Rightarrow$  DK  $\Rightarrow$  DS  $\Rightarrow$  Truck\_Out

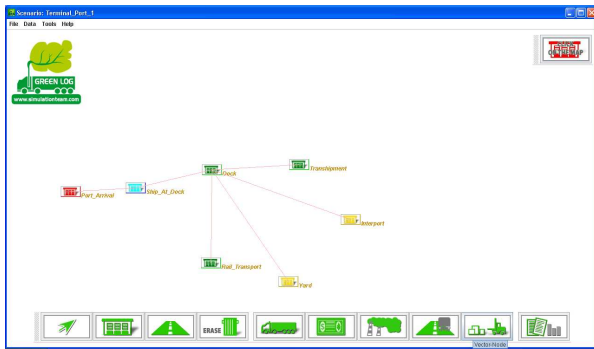


Figure 13. Port Terminal Model

#### Export

From Feeder Ship to Ship

Ship  $\Rightarrow$  DK  $\Rightarrow$  DS  $\Rightarrow$  DK  $\Rightarrow$  Ship

From Train to Ship

Train  $\Rightarrow$  DR  $\Rightarrow$  DS  $\Rightarrow$  DK  $\Rightarrow$  Ship

From Truck To Ship

Truck\_Out  $\Rightarrow$  DS  $\Rightarrow$  DK  $\Rightarrow$  Ship

The authors used a real case study of an Italian Port Terminal as reference in order to evaluate the current EIs and the EIs by considering new solutions allowing to improve port terminal sustainability:

- improvement of rail infrastructures

In particular the first scenario considers:

- Flow of 700'000 TEU
- Lower use of rail transportation (30% of the total flow)
- Heavy Fuel Ship Engine

Scenario 1			
Flow	700000 TEU/anno		TEU/anno
Interport	60% to Truck		420000
Rail Transportation	30% to Train		210000
Transshipment	10% to Feeder Ship		70000
Waiting Time at Gate	1 h		
Heavy Fuel			

The second scenario take into account:

- Flow of 700'000 TEU
- Increase Rail transportation (60% of the total flow)
- Low Sulfur Fuel Ship Engine

Scenario 2			
Flow	700000 TEU/anno		
Interport	30% to Truck	210000	TEU/anno
Rail Transportation	60% to Train	420000	TEU/anno
Transshipment	10% to Feeder Ship	70000	TEU/anno
Waiting Time at Gate	20 min	0.333333	h
Low Sulfur Fuel			

The Authors created the Conceptual Model by using the GreenLog Objects and implemented it in GreenLog Simulator by defining the general architecture based on the following critical elements:

- Ship\_Arrival that represents the entrance operations in the port
- Transshipment to load containers from a ship to a ship
- Dock that refers to the dock operations
- Yard
- Rail Transportation
- Multimodal Terminal for representing containers to external truck

The different critical nodes are connected in order to represent the whole port terminal model:

The main characteristics for handling operations present in the terminal are summarized in the following table:

	Capacity[kg]	Productivity	speed[m/s]	Hp	Number
Portainer	50000	25	0.75	470	4
Transainer RTG	46000	25	2.2	750	3
Reach Stacker	46000	20	7	242	4
Truck 1 Out	45000	13	15	371	20
Transainer RMG 1	50000	25	1	750	1
Transainer RMG 2	50000	20	1	750	1

The authors model stochastic ships using for average reference vessel the following features:

- Average Capability 8000 TEU
- Average Length 332.0 m
- Average Width 43.2 m
- Maximum Draught 16.0 m



Figure 14. Ship of Reference

Concerning with EIs the authors focused this analysis on the following elements:

- Air emissions (CO, NO<sub>x</sub>, PM10)
- Power Consumption
- Tire Consumption
- Dust
- Noise
- Waste

In addition the model allowed to estimate the following KPIs:

Terminal throughput  
Hour Productivity  
Container Handling Costs  
Year Profitability

The Emission Factors used in the model was set based on from European standards for Diesel Engine and from European Reports about Rail System Environmental Impacts [12] [27]:

Emission Factors for Representative Diesel Engines		
CO Emissions Factor	1.5	g/kwh
NO <sub>x</sub> Emissions Factor	12.5	g/kwh
PM Emissions Factor	0.3	g/kwh

Emissions	Ship Heavy Fuel [g/kWh]	Ship Low Sulfur Fuel[g/kWh]
Nox	15.2	17.5
CO	0.9	0.69
CO2	691	676
SO2	9.5	0.26
PM	0.67	0.14

Average Rail Energy Consumption Analysis for European Trains	
German Trains	22.78 KWh/trainKm
French Trains	17.95 KWh/trainKm
Spanish Trains	16.58 KWh/trainKm
European Average Energy Consumption	19.10 KWh/trainKm

Rail Emissions for Diesel railway Locomotives	
Emission Power Specific g/kWh Fuel Specific - g/kg	
CO	1 - 10 5 - 40
HC	0.5 - 4.0 3 - 25
NOx	6 - 16 30 - 70
Particulate	0.2 - 1.2 1 - 6
SO2	0.2 - 2 1 - 10

By running on this scenario the Green Log Simulator it is possible to extract the following results summarized in the following Green Log reports and charts.

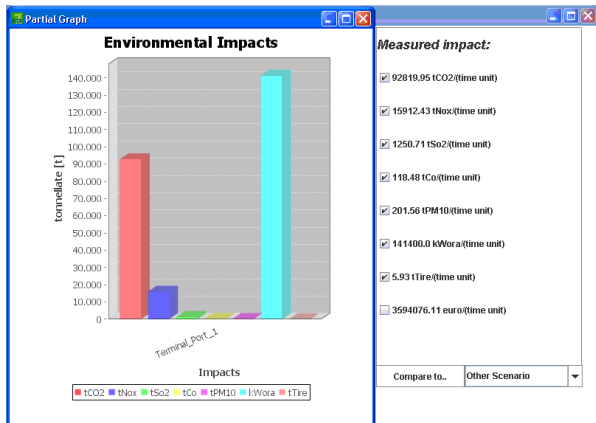


Figure 15. Port Terminal First Scenario Results

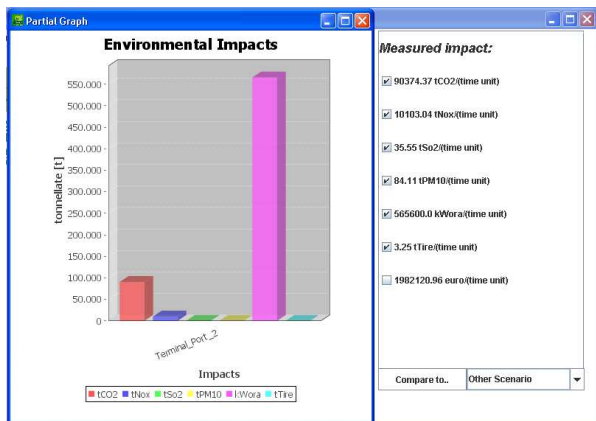


Figure 16. Port Terminal Second Scenario Results

From the analysis of the obtained data it emerge that second scenario confirms the possibility to guarantee high reduction of CO2 and SO2 emissions.

The first aspect could be correlated to the reduction of trucks waiting time at the gate and on the improvement of rail infrastructures by using electrical locomotives; the second one is related to the use of low sulphur fuel for ships in the second scenario.

On the other hand the electricity consumption increase in the second scenario due to the increasing use of electrical locomotive transportation.

Some comparisons are reported in the following diagrams.

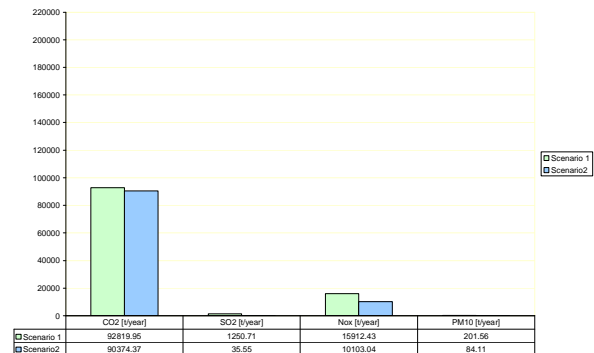


Figure 17. Air Emissions Comparison

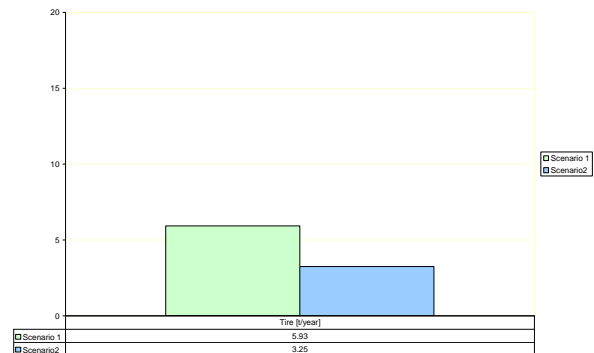


Figure 18. Tire Consumption Comparison

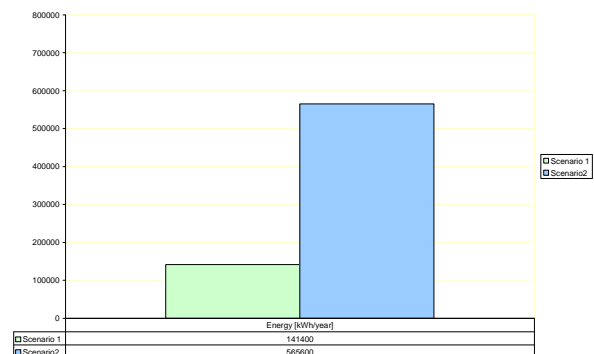


Figure 19. Power Consumption Comparison

## 5. CONCLUSIONS

The research proposes a model of a Port Terminal devoted to support its logistics, operational as well as environmental efficiency; in fact in order to analyze and evaluate Environmental Impacts related to these complex phenomena the use of simulation represent a strategic advantage able to guarantee good benefits; in fact this approach move qualitative greening concepts to quantitative results by running the simulator, so it become possible to analyze and to compare different solutions such as the use of low sulphur fuel instead heavy fuel or the use of electronic seal instead the mechanical one. The proposed results highlights the

relevance of modelling and simulation as support for EIs analysis in logistics and for decision making in term of strategies and solutions adoption for sustainability improvement in a Port Terminal. As matter of fact, the authors present GreenLog Simulator as decision support tool for Environmental Impacts Evaluation in different scenarios (maritime as well as terrain). The future development of these researches will be focusing on the evaluation of combined KPI able to provide an overview of different solutions in term of combined efficiency/cost/safety/quality and sustainability in complex logistics networks as well as in port operations

## REFERENCES

- [1] Amico Vince, Guha R., Bruzzone A.G. (2000) "Critical Issues in Simulation", Proceedings of SCSC, Vancouver, July.
- [2] Arnold O. (2006) "Efficiency of Battery Channel", Technical Report French Ministry of Environment, Paris.
- [3] B. Hayman, M. Dogliani, Ivar Kvale A. Magerholm Fet, "Technologies for reduced environmental impact from ships – Ship building, maintenance and dismantling aspects", July 2000.
- [4] Barros F., Bruzzone A.G., Frydman C., Giambiasi N. (2005) "International Mediterranean Modelling Multiconfernece - Conceptual Modelling & Simulation Conference", LSIS Press, ISBN 2-9520712-2-5 (pp 232).
- [5] Bruzzone A., Mosca R., Revetria R. (2001) "Supply Chain Management over the web in Aerospace Industry by using Simulation: WILD", Proceedings of Virtuality2001, Turin November
- [6] Bruzzone A.G. (2002) "Supply Chain Management", Simulation, Volume 78, No.5, May, 2002 pp 283-337 ISSN 0037-5497
- [7] Bruzzone A.G. (2004) "Introduction to Modeling and Simulation Methodologies for Logistics and Manufacturing Optimization", Vol 80, No.3 pg 119-120 ISSN 0037-5497.
- [8] Bruzzone A.G., Massei M., Tarone F., Longo F. (2009) "Environmental Sustainability: A Case Study In The Automotive Sector", Proceedings of INGIND17 Summersim, Monopoli, Italy, September
- [9] Bruzzone A.G., Mosca R., Revetria R., Coppa A. (2001) "Simulation Application in Distributed Logistics", Proceedings of 6th ISL2001, Salzburg, July 8-10.
- [10] Bruzzone A.G., Williams E. (2006) "Summer Computer Simulation Conference", SCS International, ISBN 1-56555-307-1 (504 pp).
- [11] Colville, R.N., Hutchinson, E.J., Mindell, J.S. and Warren, R.F.(2001) "The transport sector as a source of a air pollution", Atmospheric Environment, 35(9), 1537-65.
- [12] Cooper D. (2003) "Exhaust emissions from ships at berth", Atmospheric Environment 37 3817–3830, Elsevier
- [13] Cox, A.(1999) "Power, value and supply chain management". International Journal of Supply Chain Management, 4(4), 167-75.
- [14] Daly H.E. and Cobb J.B.Jr., (1989) "For the Common Good: Redirecting the Economy Toward Community, the Environment and a Sustainable Future", Beacon Press, Boston (1989).
- [15] Denton, T. (1998) "Sustainable development at the next level" Chemical Market reporter, 253(7) 3-4.
- [16] EDCG (2000) "A study on the economic valuation of environmental externalities from landfill disposal and incineration of waste", Technical Final Report European Commission DG Environment, Bruxelles, October.
- [17] Gifford, D. (1997 "The value of going Green" Harvard Business Review, 75, 11-2.
- [18] Harris, J.M., Wise, T.A., Gallagher, K.P. & Goodwin, N.R. (eds.). (2001), "A Survey of Sustainable Development: Social and economic dimensions", Island Press, Washington
- [19] Hibbert, L.(1998) "Sustainable activity" Professional Engineering, 11, 32 -3.
- [20] J. Sarkis, Greening the Supply Chain, London Springer-Verlag, 2006, Part I.
- [21] J.H.J. Hulskotte, Denier van der Gon, Fuel consumption and associated emissions from seagoing ships at berth derived from an on-board survey, 2009 Elsevier.
- [22] Jokat, J. and Cabos, C. (1998). 'NOISE FEM, the Practical Application of a New Prediction Method for Noise Levels in Ships', Proceedings of PRADS 98 (1998).
- [23] Khoo, H.H., Spedding, T.A., Tobin, L., Taplin, D.(2001) "Integrated simulation and modelling approach to Decision Making and Environmental Protection" Environment, Development and sustainability, 3(2), 93-108.
- [24] Matthias, R. (1999) "Strategies for promoting a sustainable industrial ecology, Environmental Science & Thecnology, 33(13), 280-82.
- [25] Merkuryev Y., Bruzzone A.G., Merkuryeva G., Novitsky L., Williams E. (2003) "Harbour Maritime and Multimodal Logistics Modelling & Simulation 2003", DIPTTEM Press, Riga, ISBN 9984- 32-547-4 (400pp).
- [26] De Jonge E., Hugi C., Cooper D., (2005) Report 2005 for European Commission Directorate-General-Environment Directorate C - Unit C1, Entec Uk limited Windsor House. Gadbrook Business Centre.
- [27] Report 2009 "Emissions Standards: European Union Heavy-Duty Diesel Truck And Bus Engines" 2009 Global Sourcing Guide, produced in cooperation with DieselNet
- [28] Seamann, R.(1992) "The environment and the need for new technology: empowerment and ethical values" Columbia Journal of World Business, 27, 186-93.
- [29] Spedding, T.A., H.H., Taplin, D. (1999) "An integrated simulation approach for teaching industrial ecology" Proceedings of the HKK Conference" Waterloo: Canada

- [30] Starcrest Consulting Group, Port of Los Angeles inventory of air emissions, Technical Report, September 2007.
- [31] Stead, J.G., Stead, E. (2000), "Eco-enterprise strategy: standing for sustainability" *Journal of Business Ethics*, 24(4), 313-29.
- [32] Street, A.C. (1986) "The Diecasting Book" 2nd Edition, Portcullis Press.
- [33] Taplin, D.M.R., Spedding, T.A., Khoo, H.H. (2001) "Environmental security: simulation and modeling of Industrial Process Ecology" Paper presented at the Strasbourg Forum, Council of Europe, Strasbourg: France.
- [34] U.S. DoD, Estimating energy and water consumption for shore facilities and cold iron support for ships, UFC 3-401-05N 16 January 2004.
- [35] U.S. EPA, "Draft Cruise Ship Discharge Assessment Report", Technical Report, December 2007.
- [36] V. Yılmaz, C. Aktaş, C. Yağizer, Investigating The Effects Of Environmental Sensitivity And Environmental Behavior On Ecological Product Buying Behavior Through Structural Equation Modeling, *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* (20) 2010 / 2 : 127– 140
- [37] Velasco H. (2008) "Sustainability: The matter of time horizon and semantic closure", *Ecological Economics* Volume 65, Issue 1, 15 March 2008, Pages 167-176.

# MODELLING A TRACEABILITY SYSTEM FOR A FOOD SUPPLY CHAIN: STANDARDS, TECHNOLOGIES AND SOFTWARE TOOLS

De Cindio B. <sup>(a)</sup>, Longo F. <sup>(b)</sup>, Mirabelli G. <sup>(c)</sup>, Pizzuti T. <sup>(d)</sup>

<sup>(a)</sup>Department of Modeling for Engineering, University of Calabria, Italy

<sup>(b) (c) (d)</sup> Mechanical Department, University of Calabria, Italy

<sup>(a)</sup> [bruno.decindio@unical.it](mailto:bruno.decindio@unical.it) <sup>(c)</sup> [f.longo@unical.it](mailto:f.longo@unical.it), <sup>(b)</sup> [g.mirabelli@unical.it](mailto:g.mirabelli@unical.it), <sup>(c)</sup> [teresa.pizzuti@unical.it](mailto:teresa.pizzuti@unical.it)

## ABSTRACT

The aim of this article is to define the state of the art of the international regulatory framework for the food industry, standards, technologies and software utilized for modelling the food supply chain, analyzing for them the advantages and the disadvantages.

Traceability systems can help to know food origin and monitor means of transport and distribution, create “hallmarks” and improve commercial potential (external or supply chain traceability).

On the other hand, for a single company, there are many advantages in adopting a traceability system such as avoiding food adulteration, improving control systems, guaranteeing high levels of quality, increasing health and safety using transparency during the entire process chain.

Keywords: tracing, tracking, food supply chain, decision support system

## 1. INTRODUCTION

After the serious incidents that have invested the food sector (BSE, dioxin contamination in food, blue mozzarella, divulgation of the Escherichia Coli mutation, etc.), several institutions have promoted the introduction of control systems able to effectively trace not conforming (good) products and to find the risk factors that led to dangerous conditions for human health by compromising food hygiene. Currently, although important legal regulations have been introduced to define the general principles of food safety, the food sector is continuously exposed to risks and dangers of food fraud. To reduce this exposure is desirable the widespread adoption of efficient traceability systems that take into account the specific features that characterize each type of food supply chain.

The food supply chain is the complex structures that contribute to the production, distribution, marketing and supply of a food product. The food supply chain is typically formed by five basic entities: the raw materials producer, the processing company, the carrier, the distributor and the retailer (Gandino et al. 2009). Each entity plays a specific role. The raw materials producer

(first processing company) sows and grows the agricultural products and sells them to the processing company; the processing company transforms the raw materials; the carrier conveys food from a company to another; the distributor handles food commodities; the retailer sells food to the final consumer.

A well organized food supply chain should have the ability to reconstruct the history of each product and follow the food through the various processing steps, identifying and recording the materials used and the operators involved, correctly combining this information to the single product package introduced in the market (Figure 1).

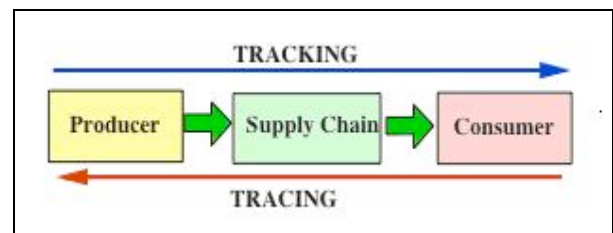


Figure 1: Tracking and tracing system.

This result can be conveniently achieved if each company of the supply chain adopts an internal system for controlling and recording information (internal traceability) and if transitions between the actors are regulated and managed in a coherent and shared form (external traceability). In this way is possible to trace the path followed by a food product that moves from “farm to fork”.

Tracking and tracing terms indicate two mirror processes, but they are often used in an interchangeable way even though they have different meanings. Tracking is the process by which the product/food is followed by upstream to downstream in the supply chain by recording data in each production stage. On the other hand, tracing is the reverse process of tracking. Through tracking systems it is possible to trace the global history of the product and the responsibilities at different processing stages. At the end of the production process, a third party or public control authority, processing the stored information, will reconstruct the



history of the product and identify the critical control point (Bernardi et al. 2007). Traceability should not be confused with the obligation to provide consumers with information that characterize the final product, since this information are present in the product label.

The operations required by the traceability management system can be divided into two main activities which refer to internal traceability and supply chain traceability. The *internal traceability* is realized by internal procedures, different for each business, that allow to trace the origin of materials used, the process operations and the food destination. The *food supply chain traceability or external traceability* is guaranteed by the integration and coordination of the tracking procedure adopted by each operator of the chain, and represents the ability to follow the path of a specific unit of product along the production chain.

## 2. REGULATORY FRAMEWORK

The regulatory framework for food traceability is wide and diversified. The traceability term was first introduced by the Codex Alimentarius Commission (1999) as the “ability to trace the history, application or location of an entity by means of recorded information”.

The Regulation (EC) 178/2002 represents the main regulatory reference for the food legislation. The regulation introduces for the first time the traceability concept for all the food products. Regulation EC/178/2002 defines traceability as the ability to trace and follow food, feed, and ingredients through all stages of production, processing and distribution”. The Regulation specifies that all operators of the sector must have appropriate systems and procedures that enable the competent authorities to retrieve the necessary information and to select the companies which have been provided. Therefore, each operator can independently choose different tools and methods to achieve this goal. The critical aspect related to the Regulation (EC) 178/2002 is that it obliges each operator to only record the information related to their immediately preceding suppliers and immediately successive client (Charlier et al. 2008). It doesn’t introduce any prescription on the internal traceability that enable to trace the path followed by each single unit of raw material and ingredient utilized in the production process in the company.

This information gap can be filled by implementing voluntary internal traceability systems. A voluntary traceability system requires to record additional information that enable companies to effectively monitor each production phase.

In Italy, The Italian National Unification Agency (UNI) has introduced the standards (1) UNI 10939 “Traceability systems in the agricultural chain: general principles for design and development” (April 2001) and (2) UNI 11020 “Traceability System in the agro-food industries: principles and requirements for development” (December 2002). In 2008, this standards where placed by the (3) UNI EN ISO 22005:2008 that

defines “General principles and basic requirements for system design and implementation”.

Currently, the European regulatory framework requires food and feed labelling and identification documented by recorded information. This goal is achieved by introducing European Directives and specific nation laws. In this context, the European Community has focused its attention to certain foods, such as fresh and processed meats (Reg. CE 1760/2000), milk, eggs (Reg. CE 2295/2003), fish products (Reg. CE 2065/2001), genetically modified foods (Dir. 2001/18/CE) and it has introduced specific regulations for them. Other regulations where introduced for lipids (Re. CE 20.7.98, n. 1638 art. 4 bis), grape and vine transformation (D.M. 29.05.2001), olive oil. In Table 1 the main regulations for the traceability in the food sector are summarized with attached their definition.

Table 1: European and Voluntary standards for the food sector

European Standards	
Directive 93/43/EEC on the hygiene of food	It defines general rules of hygiene for food and the procedures for verification of compliance with these rules.
EC Regulation 1760/2000	It establishes a system of identification and registration of cattle and defines a system of mandatory labeling of beef and beef products.
EC Regulation 2065/2001	It sets out the procedure for implementing Regulation (EC) 104/2000 for informing consumers on the products of the field of fisheries and aquaculture, providing traceability of fish products.
Directive 2001/18/EC on the deliberate release of GMOs	It requires to the Member States to adopt measures to ensure traceability and labeling for GMOs.
EC Regulation 178/2002	It sets out general principles of food law, establishes the European Food Safety Authority and defines procedures in matters of food safety. It introduces for the first time in a horizontal manner, and therefore applicable to all types of food, the instrument of traceability.
EC Regulation 2295/2003	It defines the procedures for implementing Regulation (EEC) 1907/90 to ensure traceability of eggs, the control of their origin and of the production method.
EC Regulation 1830/2003	It defines the rules on traceability and labeling of products containing GMOs or formed by them.
Voluntary standards	
UNI 10939:2001	It provides general principles for the design and development of traceability systems in the agricultural sector.
UNI 11020:2002	It defines principles and specific requirements for the development of a system of traceability in the agro-industries.
ISO 22000:2005	It defines the requirements for the design and implementation of a system of food safety management in any company in the agro-food industry.
UNI EN ISO 22005:2008	It defines the principles and specifies the basic requirements for the design and implementation of a food traceability system.

## 1. TRACEABILITY STANDARDS AND TECHNOLOGIES

Auto-identification technologies available for food traceability are: bar code, Radio Frequency Identification (RFID) technologies, Near Field Communication (NFC) systems, Real Time Locating Systems (RTLS) systems.

The bar code is the most common and widespread technology for encoding data. In this technology the information are present in form of sequence of vertical bars characterized by different spacing and thickness. (See Figure 2). The decoded data is stored through the use of optical systems (optical scanners) that, reading the sequence of symbols, allow to obtain the desired information.



Figure 2: Typical linear bar codes

Recently, the linear bar codes have been joined by the two-dimensional bar codes. The latter adopt a matrix representation and encode information in ordered sequences of white and black modules. (See Figure 3). Such codes can contain much more information than that encoded in a linear bar code, in a more compact and with redundancy. This latter feature allows the system to read the complete information even if a piece of code is illegible or is damaged (torn) through the passing from one point in the chain to another. In fact, even if the image is damaged or irregular due to the effects of light or reflection, it is possible to reconstruct and decode the code through the use of appropriate algorithms. The reading of the two-dimensional code is possible through the use of imager or camera-based scanner that capture an image of two-dimensional using the same principle of industrial vision systems. The disadvantages in the use of two-dimensional codes are related to the shortcomings of the technologies used to capture images that are seriously affected by dirt and/or inadequate lighting conditions. The most common formats for the two-dimensional bar codes are: the standard QRCode, (especially prevalent in Japan and Asian); Datamatrix standard used in Europe, and the EZCode standards that are "proprietary solutions" present in various forms (Spain, Italy, USA, Mexico).

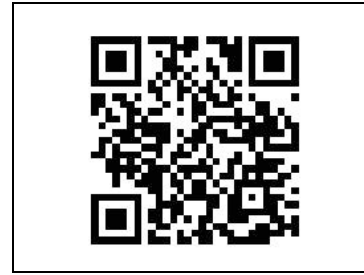


Figure 3: Typical two-dimensional bar codes

Currently, the market is characterized by a significant growth of two-dimensional barcode and a progressive development of self identification systems based on RFID. The radio frequency identification technology, called "RFID System", uses some electronic components called reader and transponder. The reader is the device installed at the control center and has the ability to interrogate the transponder, send and receive data, interfacing with the corporate information systems. The transponder is an intelligent electronic device that is applied to the object that must be tracked and/or monitored. It consists of a tag, an antenna and a support that includes the other components. The tag is the electronic component used to store the information, the antenna allows the tag to receive and transmit information, while the packaging ensure their adequate protection from bumps and weather. The antenna enable the communication between the reader and the tag and allow the tag to receive the energy necessary for its operation. The information exchange takes place via a radio signal generated by and reflected from the tag reader.

Depending on the type of power required, the tags can be classified as active or passive. The first are characterized by autonomous internal power, or battery, and enable greater power transmission and greater working distance. Passive tags do not have instead internal energy sources and are characterized by a smaller overall size. The transmission frequency is the most important technology parameter for the correct application of RFID technology. There are four different frequency bands for transmission:

- LF (Low Frequency) 125 kHz - 131 kHz;
- HF (High Frequency) 13.56 MHz;
- UHF (Ultra High Frequency) 433 MHz and 866 MHz - 915 MHz;
- MW (MicroWave) 2.45 GHz and 5.8 GHz.

The last frequency band is not really used because of the low number of applications and the high costs of employment.

The use of different frequency bands affects the ability of communication, and the construction and environmental conditions in which the tag is able to work. In fact, the higher is the operating frequency the higher is the distance of employment, the amount of information that can be transferred per unit time, the moving speed of the object to be traced and the

manufacturing cost, while the operational sensitivity is significantly conditioned by the presence of metals, liquids and electromagnetic activity. Table 2 shows, for different operating frequencies, the spectra, the reading distance, the type of power, the bit rate and areas of application (Puddu et al. 2011)

Table 2 - Operating frequencies of RFID

Operating frequency	125-135 kHz	13,6 MHz	860-960 MHz	2,4 GHz
Spectrum	low frequency (LF)	high frequency (HF)	ultra-high frequency (UHF)	microwave
Magnitude order of the operating distance	0,5 m	1 m	3 m	1 m
Power	passive	passive	passive, active	passive, active
Bit rate	up to 1 kbit/s	25 kbit/s	100 kbit/s	250 kbit/s
Application examples	animal tracking, access control, containers, vehicles identification	smart card, logistics, ticketing, baggage handling	logistics: pallet and objects, control	supply chain and logistics

A recent evolution of RFID systems is represented by the Near Field Communication (NFC). NFC technology provides wireless connectivity (RF) two-way short-range. NFC operates on a frequency of 13.56 MHz and can achieve very high transmission rate with up to 424 kbit/s. The basic elements involved in transmission are the initiator (or the first device to interrogate) and the target. Initiator and Target have symmetrical roles, and, once the communication has been initiated, they are alternate equally in the transmission. Unlike traditional RFID systems, NFC technology allows therefore a two-way communication in which initiators and targets create a peer-to-peer network in which they can both send and receive information.

A further RFID evolution is represented by the RTLS systems. This systems, using the same technology used by RFID, can identify a product, locate its position, track their movements over time. A typical active RTLS uses active tags placed on the good you want to locate, reader devices that detect the information sent by the tag and, finally, hardware and software devices that, processing the received information, are able to determine the position of the product under observation. (See Figure 4). RTLS systems may use different communication standards, the most common are: GPS Protocol, WiFi, Zibee, RFID and Ultra Wide Band (UWB). The choice of this standards depends on the type of application: for this purpose Zibee technologies, RFID, WiFi, and the Ultra Wide Band can be classified as indoor localization systems, while the GPS and active RFID are outdoor localization systems.



Figure 4 - Structure of a traceability system using RTLS systems

The most commonly used systems are bar codes and RFID; the use of RTLS is limited to those products of high value or particularly dangerous. The choice of the appropriate technology depends on many factors such as the value of assets, the size of objects, the nature of the products, the type of packaging, the amount of data to be stored, the supply chain characteristics, construction and operating costs. It is clear that the choice of RFID is the most appropriate for valuable and critical products, while bar codes are preferred for products characterized by low value, small size and low hazard of perishability.

Therefore, the possibility to use radiofrequency identification is connected with the adoption of the same frequency band for the data transmission common to all the operators of the supply chain. As a consequence, for the large adoption of RFID systems in the food supply chain is indispensable to identify a free band at global level. Finally, some privacy problems occurs during the phase of information exchange between different operators, even if some researches are trying to avoid security and privacy risks (Sanchez et al. 2001; Juels A. 2006).

The choice of the technology depends on the specific application that can ensure the lowest total cost of ownership (TCO) and faster return on investment (ROI). Of course, bar codes are more convenient for individual packages of product (consumer units) even if the adoption of RFID systems are preferable for multi-packs (packaging unit and pallets).

### 3. SOFTWARE TOOLS FOR THE DESIGN, IMPLEMENTATION AND MANAGEMENT OF TRACEABILITY SYSTEMS FOR THE FOOD INDUSTRY

The use of software for modeling and managing processes either directly or through web services is a valuable tool for implementing a traceability system. A system for storing and disseminating data concerning the traceability must include a software for modeling the food supply chain, a data server to store information

and an access via the web that make the data accessible simultaneously from multiple locations.

The processes and actors involved in the supply chain and the relations existing between them can be modeled through different techniques including Petri nets, the Structured Analysis and Design Technique (SADT), techniques for Integration Definition (IDEF) and Event-driven process chain (EPC). These techniques, along with other methodologies for process modeling (UML Activity Diagram, UML EDOC Business Processes, IDEF, ebXML BPSS, Activity-Decision Flow (ADF) Diagram, RosettaNet, Loveme, and Event-Process Chains (EPCs)), were revised in 2001 by the Business Process Management Initiative (BPMI), which defined a new standard notation, the Business Process Model and Notation (BPMN).

The BPMN allows to reconstruct the process diagrams (BPD - Business Process Diagram) by means of graphs or networks of "objects". These objects represent the activities of the process and are linked by control flows that define the logical relationships, dependencies and order of execution. In general, BPMN is used for two different operating configurations. The first refers to processes that take entirely place in a company. In this case, processes are private and the internal activities are not directly visible from the outside (internal tracking). The latter refers to "collaborative" processes between two or more business entities (companies, organizations, units, etc.). Each entity develops its own process, which will exchange information with other industry players. These exchanges take place especially when a product moves from one operator of the supply chain to another and it is necessary to keep track of this transition.

For traceability information it is possible to create a computer-based system that integrates the database of the business processes and creates a "front end" displayed in a Web browser that enables the retrieval of information via internet. There are different software for modeling processes according to the BPMN standard that directly create a user interface for displaying the flow of information. The most commonly used are:

1. Aris (IDS Scheer GmbH Prof.)
2. Tibco Business Studio 2.0
3. Intalio
4. WebRatio.

The ARIS Platform has been mainly used in the past for modeling processes according to the scheme EPC scheme (Event-Driven Process Chain) (Bevilacqua et al. 2009). It is a web based application that allows to design in a "user friendly" way the business processes. The information captured by the Aris tool set is stored in a database according with the Entity-relationship (ER) model. Aris allows to different servers to immediately visualize the processes. Moreover it allows to quickly access to the information, simplifying data management.

In accordance with the BPMN 1.0 standard, Tibco Business Studio supports multiple processes in separated pools that represent the environment within the process is developed, even if they are only connected by streams of messages. It does not support the execution environment created in some constructs such as Tibco Process Modeler such as intermediate posts, intermediate events and error message, solutions multi pool connected with streams of messages.

Intalio is software based on the BPMS standard (Business Process Management System) which, starting from business process, generates executable platform independent processes that can be directly managed by users. Intalio is built on the Eclipse platform and it is characterized by a modular architecture. It is able to translate any BPMN diagram in an executable model using the Business Process Execution Language (BPEL). In particular, Intalio | BPMS AJAX tool represents an integrated environment used for the real-time development of user interfaces. It is based on open source version of TIBCO General Interface TM that allows to communicate using XML, SOAP, Java Script and other accessible http services. Moreover this tool offers the possibility to directly create user-maintainable web interface.

Web Ratio is an useful tool for supporting the modeling and design of a traceability system. It allows to model the food supply chain according to the BPMN standard, create the Entity Relationship model and visualize the Web interface without the necessity to write code. Starting from a BPMN diagram, it is able to generate a complete web application according to the WebML standard. In fact, using Web Ratio is possible to create a working system without the need to master the WebML standard. The WebML model represents the basic structure for a web application and uses a unique and fixed data model for any web project. The data model is shown in the form of ER. Web Ratio allows a parallelism between the BPMN elements and the entity/relationship data model. Moreover this tool allows to save information about the status of the elements and actors of the workflow system. The use of such software, which allow process modeling and management of information flows along the supply chain can efficiently implement a system of traceability and management of a typical food chain.

#### 4. CONCLUSIONS

The development of efficient traceability systems in food chains has assumed considerable importance in recent years. The ability to trace and track every single unit of product depends on the supply chain traceability system which in turn depends on the internal data management system and the information exchanged between the actors. The technological tools, such as bar code and RFID, available for guaranteeing the traceability of product can be advantageously used in a

comprehensive traceability system that includes an internal software infrastructure for modeling the supply chain, defining the points for data capture and generating of a front end for the retrieval of information both locally and directly on a web browser.

Of course, to ensure traceability and implement an effective traceability system is necessary to model the supply chain using a notation or a language understandable by analysts and computer experts. A notation that can be advantageously used for processes modeling is the BPMN standard, that allows the reconstruction of processes through basic graphics elements.

The software infrastructure for the storage and dissemination of data on traceability should also include a data server to store information that can be retrieved by simple queries.

Currently, there are different software tools that enable the supply chain modeling and the creation of complete web applications; however, to ensure system interoperability and communication between the different actors, it is necessary to identify a standard for encoding information not only for common companies operating in the single chain, but for all reference operators.

Finally seems trivial to note that managing a system of traceability after the definition of the reference context (food chain), technology and software infrastructure usable, as well as organizational and management issues, led to the definition of the realization costs and of the costs of exercise that necessarily have to be compared with the benefits expected in order to verify the actual economic sustainability.

## REFERENCES

- Bernardi, P., Demartini, C., Gandino, F., Montrucchio, B., Rebaudengo, M., Sanchez, E.R., 2007. Agri-Food Traceability Management using a RFID System with Privacy Protection. The 21st International Conference on Advance Networking and Applications, 68-75.
- Bevilacqua, M., Ciarapica, F.E., Giacchetta, G., 2009. Business Process reengineering of a supply chain and a traceability system. A case Study. *Journal of Food Engineering* 93, 13-22.
- Charlier, C., Valceschini, E., 2008. Coordination for traceability in the food chain. A critical appraisal of European regulation. *Eur J. Law Econ* 25, 1-15.
- Gandino, F., Sanchez, E.R., Montrucchio, B., Rebaudengo, M., 2009. Opportunities and constraints for wide adoption of RFID in agri-food. *Proceedings of IJAPUC*. 2009, 49-67.
- Juels, A., 2006. RFID security and privacy: a research survey. *IEE Transaction on Industrial Electronics*, vol. 24, n. 2, 382-394
- Puddu, E., Mari, L., 2011. Un'introduzione ai sistemi Rfid. *Principali caratteristiche tecnologiche e funzionali*. *Tutto\_Misure*. Anno 13, nr. 1.

Sanchez, E.R., Gandino, F., Montrucchio, B., Rebaudengo, M., 2009. Public-Key in RFIDs: Appeal for Asymmetry. *Security in RFID and sensory networks*. CRC Press, 195-216.

## AUTHORS BIOGRAPHY

Prof. Bruno de Cindio is 65 years old and is full professor of Thermodynamics at University of Calabria. He got a Chemical Engineering MS degree at the University of Naples "Federico II" where he was appointed as associated professor till about 1994 when he moved to University Calabria. His scientific activity was focused on polymers during the first 10 years, when he shifted his interest to food rheology and processing. He had in charge the direction of a small group of scientist for about three years starting in 1981. Thereafter he was directing till 1991 an industrial research center for foods close to Naples, then he definitively moved to Calabria. He published more than hundred papers dealing mainly arguments related to rheology, polymers and food processing. He is nowadays directing a group of about ten young researchers of the Laboratory of Rheology and Food Process at the University of Calabria. He was now elected president of the Italian Society of Rheology.

Francesco Longo was born in Crotone in 1979. He took his degree in Mechanical Engineering, *summa cum Laude*, in October 2002 from the University of Calabria. He received his Ph.D. in Mechanical Engineering from University of Calabria in January 2006. He is currently Assistant Professor at the Mechanical Department of University of Calabria. His research interests include Modeling & Simulation for production systems design and supply chain management. In these areas specific subjects regard: Modeling & Simulation for production systems and supply chains design and management; Supply chain security; The inventory management problem along the supply chain; simulation tools for training procedures in complex environment. Workplace and Workstation Effective Ergonomic design within manufacturing and production systems. He is Director of the Modeling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES), a laboratory operating at the Mechanical Department of University of Calabria. The MSC-LES is member organization (as one of the two Italian centers) of the MS&Net (McLeod Modeling & Simulation Network).

Giovanni Mirabelli was born in Rende in 1963 and he took the degree in Industrial Engineering at the University of Calabria. He is currently researcher at the Mechanical Department of University of Calabria. His research interests include ergonomics, methods and time measurement in manufacturing systems, production systems maintenance and reliability, quality and logistic. He has published several scientific papers participating as speaker to international and national

conferences. He is actively involved in different research projects with Italian and foreign universities as well as with Italian small and medium enterprises. He currently teaches Industrial Plant and Management of Production Systems respectively for students of Management and Mechanical Engineering.

Teresa Pizzuti was born in Cariati (CS) in 1985. She took her degree in Management Engineering, Logistics specialization, *summa cum Laude*, in July 2010 from the University of Calabria. She is currently PhD student at the Mechanical Department of University of Calabria in the Transport, Logistics and Transformation sector.

Her research activities concern the study and application of innovative methods for improving quality and logistics of industrial products, referring in particular to agro-food products. She collaborates with the Industrial Engineering Section of the University of Calabria to research projects for supporting innovation technology in SMEs.



# Command Feedback and Response (CFR) – The evolution of Command and Control in an Immersive and Interactive environment (C2I2).

Marco BIAGINI <sup>(a)</sup>

Michele TURI <sup>(b)</sup>

Genova University  
DIMS (Department of Mathematic Engineering and Simulation)  
PhD Course

<sup>(a)</sup> [m.biagini@liophant.org](mailto:m.biagini@liophant.org)

<sup>(b)</sup> [m.turi@liophant.org](mailto:m.turi@liophant.org)

## ABSTRACT

According to the NATO NEC C2 Maturity Model, there are two key factors about evolution of C2 operational function in the 21st century:

- the extreme uncertainty of next decades military missions space;
- the DOTLMPFI military transformation approach, related to Institutions and Actors in the Information Age, about the ability to leverage new information technologies in a IIIM context.

This complex and dynamic environment will affect next generation of digital C2I systems that should be integrated with analysis capabilities and M&S technology in a full interoperable environment.

The authors introduce a new theory about the leverage of the C2 operational function referred to strong requirements to implement more agile capabilities to support operational planning and decision making processes.

This paper, applying the **Command, Feedback & Response** theory, will present a research and study activity about the impact of an immersive/interactive environment, that the authors named “C2I2 bubble”, to improve the efficiency and the effectiveness in the cognitive-sensorial and social domains.

In this sector the authors have considered the most advanced technology for information visualization, using Augmented Reality techniques and advanced HMI for remote collaboration.

research activities to design the next generation of Command and Control.

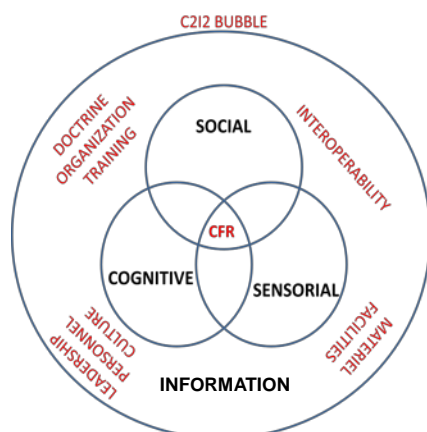
Taking into account these trends, the authors develop their research considering what they named CFR. The authors define CFR as the evolution of next generation Command and Control Operational Function applied to a “*Crowdsourced environment*” driven by “*Virtual Cross Functional Teams*”. They looking at the benefits of dispersing leadership throughout “virtual” teams and “virtually dynamic” organizational structure that can be changed in real time (self-synchronized) to response faster to the mission evolutions. This environment is enabled by the support of portable devices, immersive and virtual technologies meshed with augmented reality using standardized interoperable languages.

From an operational point of view, CFR represent the moving from traditional command and control organization and processes to the true collaboration and team working in a combination of the two operating simultaneously.

In a military environment Crowdsourcing can be referred to the increasing ability, especially facilitated by new media technologies, to leverage the knowledge, talent, expertise and interest of large groups of people — Commanders, Analysts, Staff employees, Augmentees, Specialist and Warfighters, in a boundless-like Command Post, where the “*people needs to know*”, are far beyond their physical location.

People belonging to different staffs and organizations can be put together joining a Virtual Cross Functional Team passing from a hierarchical Command and Control – based organization to a much flatter one running off of social networking groups.

This concept, far beyond the traditional staff/line management (that remain however the pillar of a military type of organization structure), can leverage human resources and dynamic organizations to implement more agile capabilities to support operational planning and decision making processes. To work properly this concept need to be used with a positive approach to information sharing (intelligence paradox) in a Network Centric Value Chain (fig. 3), represented in the information cognitive, physical and social domains, SAS-065 (2010).



## 1. INTRODUCTION

The rising of extensive use of social networking, the new generation of mobile devices and the latest M&S trends and technologies, associated to the virtual worlds and augmented reality and ubiquitous Internet access, are addressing the most studies and

By the way CFR takes into consideration not only the differences between the “line/staff” and the “crowdsourced management organization”.

In complex endeavors SAS-065 (2010), the authors want to show the idea that the concept of “Unity of command” could be improved by the CFR concept of “Multidimensionality of command”.

In addition, in crowdsourced organizations based on virtual cross functional team, the commander could be able to leverage resources creating self-synchronized relationship with subordinated parallel, and upper level of command gathering the feedbacks and addressing responses as orders to the assigned units.

## 2. STATE OF THE ART TO DEVELOP A REVISITED C2 CONCEPT

The developing of CFR theory takes into consideration the NCW and NEC paradigms applying Web 2 and Web 3 technologies making virtual and realtime dynamic the C2 relationship between people and the belonging to organizations in an EDGE C2 level. NATO NEC C2 approach is basically characterized by a robustly networked collection of entities having wide spread and easy access to information sharing information extensively interacting in a rich and continuous fashion and having the broadest possible distribution of decision rights. The objective of EDGE C2 is to enable the collective to self-synchronized, SAS-085 (2010). Starting from the scientific and technology state of art, the authors define the boundaries of a CFR model taking inspiration from the following concepts:

### 2.1. Crowdsourced, Cross Functional Team Organizations and a related Maturity Model.

Crowdsourcing is the idea to call the collaboration of an undefined group of people, it gathers those who are most fit to perform tasks, solve complex problems and contribute with the most relevant and fresh ideas, Howe (2006).

The terms coming from the act of outsourcing tasks, traditionally performed by an employee or contractor, to an undefined, large group of people or community (a “crowd”), through an open call.

But crowdsourcing is also the *channeling the experts desire to solve a problem and then freely sharing the answer with everyone*, Van Ess, (2010).

Crowdsourcing also has the potential to be a problem solving mechanism for government and no profit use, Brabham (2008).

For this reason it is possible to think about a Government Crowdsourced organization as a distributed problem solving model. Even if in the classic use of the term, problems are broadcast to an unknown group of solvers in the

form of an open call for solutions, this model can be adapted and implemented into a traditional military staff/line management organization introducing the concept of Cross Functional Team (CFT) (fig. 1).

To face today's complex challenges, the authors need to incorporate a wide range of capabilities, skills, and different perspectives. Cross functional teams are regarded as a means to manage social collaboration and concept creation.

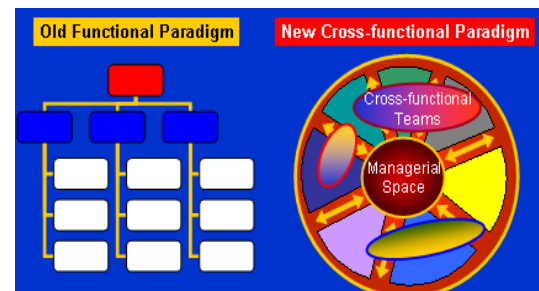


Fig. 1 functional paradigms, [http:// www.1000ventures.com](http://www.1000ventures.com).

This kind of organization has been used for the first time in a military environment, at IJC Headquarters (ISAF Joint Command, Afghanistan). After it was adopted in 2010 during deployment of HQ MNC NE (Headquarters Multinational Corps Northeast) in the same operational theatre.

CFT can be thought as an evolution of a "traditional" military staff organisation. In a traditional staff organization all the principal functional areas are devoted to work together only for a specific efforts and each area specialists add up their studies and conclusions to the general plan. Coordination and synchronization is demanded to the leadership, Palm (2011). All processes at the HQ are maintained by CFT and the role of functional areas is to provide "in-depth" knowledge of their respective areas. CFT were created as a direct answer to the Afghanistan complex environment and latest NATO's to develop Comprehensive Approach and create a better holistic situational understanding, to enable combined planning and conduct comprehensive counter insurgency operations. CFT enable the HQ's to be more fast, flexible and agile. Actually NATO is conducting (2011) an investigation thorough the CFT concept in different training events (Little Eagle I 2011, Exercise Compact Green 2011, Bold Eagle Exercise 2011). The implementation of Cross-Functional Team concept requires a change of everybody's mindset because it is a mesh between the staff and line management organization. Even though the basic military planning, execution and analysing principles remain untouched, there are new requirements

for information sharing, intensity of social interaction and team work. The positive effect of different teams is based on the theoretical perspective of information processing states that diversity in teams will increase the range of perspectives and enhance opportunities for knowledge sharing, and improve the outcome in terms of quality and creativity.

CFR implements the Cross Functional Team Concept from a hierarchical to a flatter perspective of Command and Control in a Crowdsourced environment using the Virtual Cross Functional Team approach, supported by the latest social networking technologies and portable devices.

The above concepts well fit a Command and Control organization shaped on the CFR that, according to the N2C2M2 model, can fulfill the EDGE C2 level.

The object of EDGE C2 is to enable the community to be self-synchronized. It requires that a solid, shared understanding would exist across the contributing elements connected by a robust network. The N2C2M2 was developed to create a conceptual model, and then to test it in an experimental environment using ELICIT. ELICIT is an online multiuser platform for conducting experiments in information sharing and trust", SAS 065 (2010).

The application of this model could be used to evaluate the maturity level of a C2 approach in a complex environment to address new mission challenges.

This way of thinking about C2 is compatible with current NATO Allied Command Transformation (ACT) thinking on Future Capable Forces which puts the emphasis on Mission Command within federated complex coalition's environments.

The model defines five levels of maturity that are linked to the five NNEC capability levels ranging from conflicted C2 to Edge C2.

Each level corresponds to a different region within the C2 Approach Space, it contains the different possible approaches to accomplish the functions that are associated with command and control (fig 2). This approach space can be viewed from two perspectives.

First, it can be used to think about C2 within existing organizations. Second, it can be used to think about how a disparate set of independent (inter-dependent) entities, that is, a collective, can achieve focus and convergence. The latter is taking into consideration by the developing CFR concept.

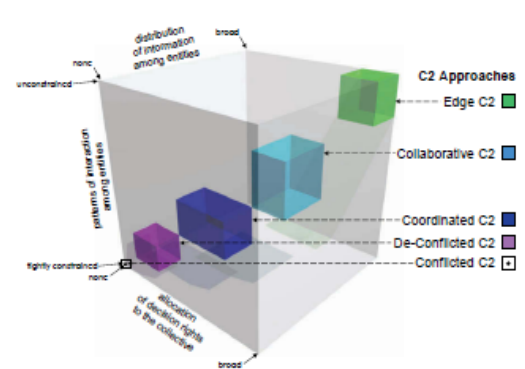


Fig. 2 C2 approach spaces, SAS 065 (2010)

## 2.2. From common languages to a C2 crowdsourced interoperable architecture

In a crowdsourcing environment is also fundamental to communicate in a common understandable and effective way. The state of the art of last upcoming C2 technology, regarding common languages to achieve interoperability and immersive and interactive devices to best perform and visualize information, are following explained.

Web 2.0 and Web 3.0 technologies make the core of the social networking environment.

About the adoption of a common understandable and standardized language, for the CFR concept the authors are focusing on the C-BML (Coalition Battle Management Language). C-BML define an unambiguous language to describe a commanders intent and generate executable descriptions of a mission that can be used by human forces in real operations supported by C2 systems, by simulated forces in simulated operations, and by robotic forces in real or simulated operations, <http://c-bml.org>.

C-BML is under standardization by Simulation Interoperability Standards Organization (SISO), in the C-BML Product Development Group. The resulting language is intended to be applicable not only to simulation systems, but also to operational command and control systems, and robotic systems.

C-BML aims to guarantee interoperability with other standards and languages like MIP bl.3 (JC3IEDM defined by NATO Stanag 5525 currently under ratification) and MSDL (Military Scenario Description Language). It is a language defined by XML schema used to transfer scenario operational data from C2 systems to simulation systems. C-BML implements the Command and Control lexical Grammar (C2LG) and currently it interoperate with the Operations

Intent and Effects Grammar (OIEG) ver. 1.1.

C2LG is an implementation of a Domain Specific Grammar. C2LG it is considered necessary because there is no grammatical approach within the C2IEDM (MIP bl. 2) and JC3IEDM (MIP bl. 3) that represent Command and Control information in such a way that is needed to from a BML, Schade and Hieb (2006-2008).

OIEG is a language designed to support the exchange of Intent and related information objects in Command and Control (C2) systems and simulators. OIEG is built upon Lexical Functional Grammar and also separates the vocabulary from the grammar. OIEG is intended to be used as exchange mechanism and not for direct reading.

The state of the art of this implementation is performed by SCEMANTA tool, developed by SAAB. It is able to transfer data from an operational scenario designed into a C2 system (SITAWARE) to a constructive simulator (SWORD) via MSDL. In addition it is able to address orders and tasks from a C2 system versus the simulated entities and then it is able to address task and reports from the simulated entities to the C2 system (C-BML <=> JC3IEDM).

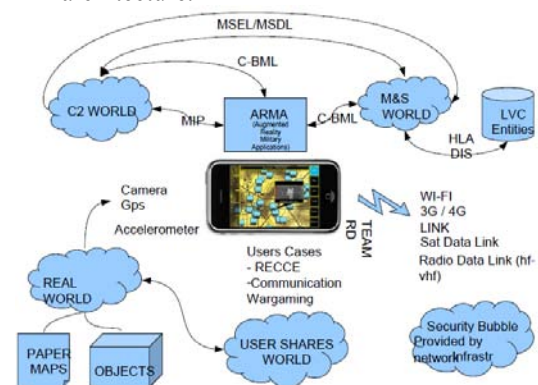
The simulation architecture could be run in standalone or in a HLA federation (i.e. HLA 1516 evolved as state of the art technology) with or without external intelligent agents, able to drive special behaviors and addressing them to simulate special purpose constructive entities.

An architecture like the above mentioned, could be used to test and experiment a C2 organization maturity level, using the ELICIT platform that is related with experimentation, teaching, and analysis to investigate the cognitive and social impacts of C2 approach and organizational structure (e.g. information sharing, trust, shared awareness). In this way it could be possible to analyze the level of maturity achieved by an organization taking the advantages of the latest C2 and M&S technologies.

From the perspective of the development of the CFR concept, the authors want to introduce the possibilities offered by the latest technologies to improve the efficiency and the effectiveness in the cognitive sensorial and social domains of a crowdsourced immersive interactive environment.

Starting from the assumption that to improve the level of awareness shared by users experienced in a crowdsourced environment, it is important to stimulate

their cognitive and sensorial apparatus using immersive and interactive techniques. The best way to this kind of stimulation can be achieved using a CAVE system (Cave Automatic Virtual Environment) that put users in an immersive virtual reality environment. An alternative to CAVE could be performed by holographic projectors. The above solutions require fixed installation and are not so easy to deploy. The widespread of more and more powerful portable devices like Smart phones and tablet pc's equipped with camera, 3D and augmented reality applications, make possible to increase the user level of immersion and interaction in a crowdsourced environment thanks to ubiquitous internet access. The authors believe that augmented reality represents the key factor and the best way to achieve an immersive and virtual experience implementing solutions that comprehend virtual world's synthetic environments. In the followed schema is represented a logical architecture.



### 3. CFR IN A COUNTERINSURGENCY SCENARIO

The reference scenario to show a CFR model is thought around the ISAF Mission, fulfilling and augmenting the Afghan Mission Network concept.

A possible CFR scenario could be designed at tactical level with an observer equipped with a special Augmented Reality (AR) mounted optical system. He could acquire a target (insurgents, terrorist, riot groups, etc.), using target acquisition methods (e.g. via passive AR technologies). He could send (automatically) a light-weight text information (e.g. in a C-BML format) to the **Crowdsourced Command Center** where it could be processed (in real time) to create the related target entities in a virtual immersive tactical 3D map, using for example, an MSDL schema (e.g. avoiding to send the real video image).

The virtual cross functional team can access and assess this information directly in the crowdsourced environment, exchanging pieces of information, suggesting task and orders, to find the right solution



to the crowdsourcer. He can directly visualize information through the optical devices, using the same light-weight text technology with AR techniques.

In addition a crowdsourced integrated Operational Analysis tool could submit the available data collected in the virtual immersive environment providing to the crowdsourcer the suggestion for the best TTP's or weapon system that could fit the commander's intents in the best way to accomplish the task, according to the desirable effects. So far enabling the warfighter to accomplish the assigned tasks in the best way.

In this context, looking at a complex operational environment characterized by the employment of dispersed forces that operate isolated in a self-synchronized way, with different tasks and missions (e.g. combat forces, combat support forces, training and mentoring forces, civil-military cooperation forces, etc.), a crowdsourced environment can optimize the flow of information. Virtual cross functional team, transversal to the hierarchical chain of command, can then fulfill and address the needs of knowledge in different situations.

An example could be found looking at the CIMIC cooperation where the needs of knowledge and awareness of common shared situation is not referred only to a military dimension but also to the civilian one focusing on local problems. More or less is the same when we are thinking about the exchange of information between coalition and host nation forces and several other Government and Non Government Organizations as characteristics of a JIIM environment.

#### 4. COMMAND FEEDBACK AND RESPONSE – A WAY TO RE-THINK COMMAND AND CONTROL

Rethinking Command and Control does not mean discarding everything about what we have learned up to now. The idea is to draw a model revisiting assumptions and building a new concept upon what remains valid. Taking inspiration from the youngest generation and from the latest social and technological trends that are changing the world (thinking about the North African and Middle East uprising) and the effects and impacts of social networks to share and broadcast information, it is possible to understand the power of a crowdsourced environment run with a social networking approach to problem solving.

Applying crowdsourcing to a military organization this principle can work with some adjustment. For understandable reasons of security the "crowd" cannot be "a so open community" as the traditional concept. The crowd is made by all people that can access to the mission community.

Users (the Virtual Cross Functional Team members) — that the authors consider the "military crowd" —

in this case form a transversal online community, not hierarchically organized, that submits solutions.

In this case the military crowd also can sort through the solutions, finding the best ones.

These best solutions are then owned by the entity that broadcast a task or an operational problem (Commander's intent and Mission Statement) in the first place — the crowdsourcer (what the authors call the Commander) and the winning individuals (users that suggest the adopted solution), in the crowd should be rewarded as the motivations of individuals to participate albeit belonging to different levels of command.

From a technological perspective a crowdsourced organization based on a virtual cross functional team needs a robust network, metaverses technologies and devices that can stimulate users using augmented reality techniques and immersive interactive technologies in a virtual world's synthetic environment. The CFR crowdsourced model is based also on the assumption that because of technological advances have allowed for cheap consumer electronics, the gap between professionals and amateurs has been diminished, Howe (2006).

Translating this assumption in the military environment, the basic idea is that talented and skilled people can be found at all levels of command and in all kinds of duty and often not related with their ranks. More often they can share their experience and skills giving their contribution to the mission success.

A crowdsourced environment implemented with immersive and interactive technologies represents a crowdsourced command center where a Commander supervises command and control activities in a CFR paradigm. This environment is also named by the authors "C2I2 bubble".

It can be represented with the Network Centric Value Chain schema (fig. 3), by the bounded area from the overlapped domains.

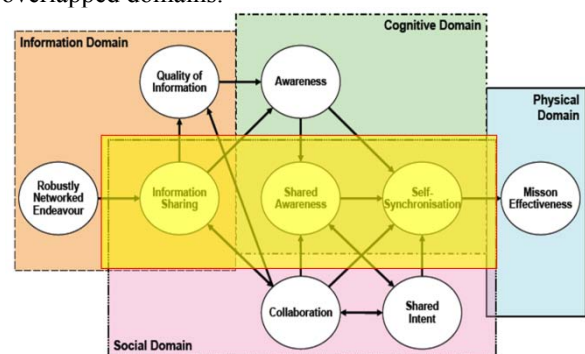
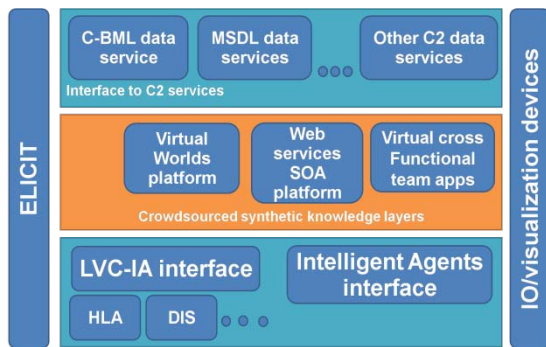


Fig. 3 C2I2 Bubble in the Network Centric Value Chain, SAS 065 (2010)

From the CFR perspective it could be possible to experiment the CFR concept building an architecture implementing the ELICIT platform to test and evaluate the CFR concept and its level of maturity, focusing on information, cognitive, and social domain phenomena.



## 5. CONCLUSION

Crowdsourcing and the Cross functional Team internet-working are the enabler of the CFR theory. CFR needs to be enhanced with emerging M&S technologies like the last Augmented Reality and Virtuality techniques to be more effective.

About the balance between Information Sharing and Information Security is important to refer to the intelligence paradox introduced by the authors. The paradox explain that there are *two main approaches to the information assurance*:

1. *protect an information and not share it to people that could need it, thinking about the damages or the disadvantages that could occur if it leaks (negation of information as a hidebound process)*
2. *share an information and not protect it, thinking about which damages could be avoided or which advantages could be achieved if it is aware by people that could need it (affirmation of information as a bounteous process).*

To achieve a crowdsourced organization model comparable to an EDGE C2 maturity level, within virtual cross functional team, it is imperative to adopt a bounteous process. The complex endeavours requires Risk analysis and Information Value Assessment as key factors for an effective full bounteous process.

The adoption of CFR approach it's only a matter of decision making procedures in the "mind and in the hands" of the Decisor. He should act as the responsible in the evaluation of information risk assessment.

In a crowdsourced organization model, if the principle of the affirmation of information is not agreed and supported or not practiced, the whole environment doesn't work. In this case team working and cross functional team cannot be effective, causing overwhelming and/or lack of information.

Main issues, regarding a crowdsourced organization model in a military environment, are:

- the needs to access the skilled community (the virtual cross functional team), where people are pre-emptive assessed and accredited or dynamically self-assigned;

- the cyber security of networks;
- the Human Culture factor related to its resilience nature to change.

The technology, in this case, is a support that can facilitate and optimize the functionality of the proposed model.

## 6. REFERENCES

- Alberts D. S., & Hayes R. E., (2006), "Understanding Command & Control - the future of command and control", CCRP Publication Series.
- Baddeley A., (2010), "The future of Command and Control", Military technology review MILTECH - 4.
- Biagini M., Joy B. (2011) "an open virtual worlds platform" Proceedings of MindSh@re, SET 2, Rome (Italy) July.
- Bruzzone A.G. Tremori A., Massei M. (2011) "Adding Smart to the Mix", Modeling Simulation & Training: The International Defense Training Journal, 3, 25-27
- Bruzzone A.G., (2007) "Challenges and Opportunities for Human Behaviour Modelling in Applied Simulation", Keynote Speech at Applied Simulation and Modelling, Palma de Mallorca, August.
- Bruzzone A.G., Frydman C., Cantice G., Massei M., Poggi S., Turi M. (2009) "Development of Advanced Models for CIMIC for Supporting Operational Planners", Proc. of I/ITSEC2009, Orlando, November 30-December 4
- Bruzzone A.G., Massei M. (2010) "Intelligent Agents for Modelling Country Reconstruction Operation", Proceedings of Africa MS2010, Gaborone, Botswana, September 6-8
- Bruzzone A.G., Massei M., Madeo F., Tarone F. (2011) "Simulating Marine Asymmetric Scenarios for testing different C2 Maturity Levels", Proceedings of ICCRTS, Quebec, Canada, June
- Bruzzone A.G., Tarone F. (2011) "Innovative Metrics And VV&A for Interoperable Simulation in NEC, Urban Disorders with Innovative C2", MISS DIPTM Technical Report, Genoa
- Child J., (2011), "Display Tech Advances Overhaul Command Control Systems", COTS journal of military electronics and computing, May.
- Cosenzo K., (2010), (Army Research Laboratory, USA), CCRP, "Adaptive Automation for Human Robot Teaming in Future Command and Control Systems".
- Don Kroening, NATO RTO SAS 039, (2009), TRAC FLVN, "Methods and Tools NATO Code of Best Practice (COBP) for C2 Assessment", US.



- Downs RM, Stea D., (1973), "Cognitive Maps and Spatial Behavior: Process and Products", Aldine Publishing Company, Chicago.
- Graff D., Koria M., Karjalainen T., (2010), "Modelling Research into Cross Functional Team Effectiveness", Helsinki, Finland.
- Iannotta B., (2011), "Dueling devices", C4ISR Journal, july.
- Le Roux W., (2010) "The use of Augmented reality in Command & Control Situation awareness", (Council for scientific and industrial research – South Africa).
- Le Roux W., (2011), "the use of augmented reality in command and control situation awareness", Council for Scientific and Industrial Research, South Africa, August.
- Massaro C., Wallis S., (2009), "self growth learning modalities", US.
- Moffat J., (2011), CCRP Publication series "Adapting Modeling & Simulation for Network Enabled Operations", feb.
- Moffat J., NATO RTO SAS 039 (2009), "Formulating the Problem and the Strategy for Solution NATO Code of Best Practice (COBP) for C2 Assessment" UK.
- Parasuraman R. (2010), (George Mason University, USA), Michael Barnes (Army Research Laboratory, USA).
- SAS-065 Membership, (2010), "NATO NEC C2 Maturity Model" CCRP Publication series, Feb.
- Schade, U. and Hieb, M. (2006) Development of Formal Grammars to Support Coalition Command and Control: A Battle Management Language for Orders, Requests, and Reports. In: 11th International Command and Control Research and Technology Symposium (ICCRTS), June.
- Schade, U. and Hieb, M. (2008) A Linguistic Basis For Multi-Agency Coordination. In: 13th International Command and Control Research and Technology Symposium (ICCRTS), June.
- Self R. L., (2010), from <http://www.rebeccaself.com/crowdsourcing/> "From command & control to Crowdsourcing", september.

## APPENDIX A

### Acronyms

C2: Command and Control  
C2I: Command Control Intelligence.  
C2I2: Command and Control Immersive and Interactive.  
CFR: Command Feedback and Responce  
CFT: Cross Functional Team  
DOTLMPFI: Doctrine, Organization, Training, Materiel,  
Leadership and Education, Personnel, Facilities,  
Leadership, Interoperability.  
ELICIT: Experimental Laboratory for Investigating  
Collaboration, Information-sharing and Trust  
HMI: human/machine interface.  
HQ: Headquarter  
JC3IEDM: Joint Command, Control and Consultation  
Information Exchange Data Model.  
JIIM: Joint Interagency Intergovernmental Multinational  
M&S: Modelling & Simulation.  
N2C2M2: NATO NEC C2 Maturity Model  
NCW: Network Centric Warfare  
NCW: Network Centric Warfare.  
NEC: Network Enabled Capability  
TTP's: Tactics, Techniques and Procedures

---

# MODELLING AND SIMULATION OF THE STERILIZATION PROCESS OF POUCH PACKAGING IN AN ASEPTIC LINE

Paolo Casoli<sup>(a)</sup>, Gabriele Copelli<sup>(b)</sup>, Michele Manfredi<sup>(c)</sup>, Giuseppe Vignali<sup>(d)</sup>

<sup>(a) (d)</sup> Department of Industrial Engineering, University of Parma, Viale G.P. Usberti 181/A, 43124 Parma Italy

<sup>(b)</sup> Interdepartmental Center SITEIA.Parma, University of Parma, Viale G.P. Usberti 181/A, 43124 Parma Italy

<sup>(c)</sup> Interdepartmental Center CIPACK, University of Parma, Viale G.P. Usberti 181/A, 43124 Parma Italy

<sup>(a)</sup> [paolo.casoli@unipr.it](mailto:paolo.casoli@unipr.it), <sup>(b)</sup> [gabriele.copelli@unipr.it](mailto:gabriele.copelli@unipr.it) <sup>(c)</sup> [michele.manfredi@unipr.it](mailto:michele.manfredi@unipr.it),  
<sup>(d)</sup> [giuseppe.vignali@unipr.it](mailto:giuseppe.vignali@unipr.it)

## ABSTRACT

Computational Fluid Dynamic (CFD) models were used in this paper to analyze and improve the sterilization process of spouted pouch with a mixture of air and Vaporized Hydrogen Peroxide (VHP) for an aseptic line. In the first part of the work the process of sterilization of spouted pouches has been analyzed using a mixture composed by vaporized hydrogen peroxide and hot sterile air. Simulations allowed reaching the best combination between nozzle position and geometrical shape of output section. In the second part of the work was carried out the simulation of the flow of sterile air in order to obtain the value of flow rate which could ensure the complete removal of the sterilizer with the lowest air consumption. The results obtained ensure a significant reduction of sterile air consumption. Results of simulations have been validated with empirical tests.

Keywords: Aseptic filling, Packaging sterilization, CFD, Design optimization

## 1. INTRODUCTION

Sterilization of food packaging is one of the most critical phases of aseptic processing (Ferretti et al., 2006; Robertson, 2006). During last ten years many experiments and studies have been performed in order to identify the better sterilizing agent for each combination of product/packaging (Robertson, 2006). Chemicals and physical methods (pulsed light,  $\beta$  and  $\gamma$  radiation) have been mainly investigated as sterilization principle (Farkas, 1999; Riganakosa et al., 1999).

Regarding chemical methods, the main problem is due to the temperature of the chemical agent allowable during the packaging treatment. Hydrogen Peroxide reaches a good efficacy only having temperature higher than 70°C, so if the packaging material to treat has a lower point of glass transition (i.e for PET 69°C), it is not possible to use liquid solution, at the risk of damage

the packaging. For this reason an addition of peroxyacetic acid in percentage of about 1% has been frequently used, in order to decrease the temperature of action of the hydrogen peroxide solution. However this addition increase the cost of the solution, so many company tried to use only Hydrogen peroxide in vapor condition (Yun an Sastry, 2007).

The two techniques mainly adopted in the last ten years was Vaporized Hydrogen Peroxide (VHP) and Condensing Hydrogen Peroxide (CHP). The first method does not produce vapor condensation on the inner side of the packaging, instead CHP method want create this condensation in order to be more powerful on microbial reduction. These methods has been tested by several authors for many food packaging in order to optimize the sterilization process (Klapes and Vesley, 1990).

Sterilization of flexible containers shows, however, additional problems during the removal of the sterilizer agent, which is complicated by the small size of the exit hole and by the type of material (Castle et al, 1995; Abdul Ghani, 2001). For this reason it was decided to analyze and simulate both the VHP sterilization process and the removal of sterilizing agent of pouch packaging. The work is therefore divided into two parts.

In the first part of the work the process of sterilization of spouted pouches has been analyzed using a mixture composed of VHP and hot sterile air, while the aim of the second part has been the evaluation and optimization of the removal process of the sterilizing mixture using sterile air. This last phase is used to obtain a low residual value of hydrogen peroxide inside the packages compatible with the values imposed by regulations. Especially, the maximum residual value of hydrogen peroxide fixed by FDA is 0.5 ppm.

Concerning the sterilization process the first goal was to determine the optimal relative position between the sterilization nozzle and the envelope, ensuring an

adequate speed of the flow in each zone of the inner pouch volume. The second aim was to optimize the flow of VHP by varying the nozzle shape.

The objective of the removal process of the sterilizing solution was to evaluate the best configuration of the nozzle, relative position and configuration, and the flow rate value that minimizes the amount of sterile air necessary to remove the residual hydrogen peroxide.

## 2. MATERIALS & METHODS

The process analysed is the sterilization of the packaging pouch in a pilot plant simulating the working of an aseptic line.

The system analyzed is composed of a sterilization station with a single nozzle spraying the sterilizing solution inside the pouch, and of a removal station of this sterilizing agent. In both of these stations, the nozzles have an overall height of 200 mm and an outlet section with a diameter of 2.5 mm. This form can be modified as required. The pouch pack has a very complex geometry, with a total height of 170 mm, a maximum width of 90 mm, a maximum depth of 53 mm and an overall volume of 340 cm<sup>3</sup>.

The sterilization is obtained from a mixture of air and hydrogen peroxide. The hydrogen peroxide at 30% of concentration is vaporized in a plate maintained at 200°C. This solution is subsequently mixed with a flow of sterile hot air that has a rate of 2000 Nl/h. The concentration of hydrogen peroxide which should arrive in each envelope is 4000 ppm. The temperature of the whole system is maintained at 55°C to avoid the problem of condensation. The condensation of sterilizing mixture would make it impossible to remove and obtain a residual value in compliance with the regulations.

After the sterilization phase of the pouch, the sterilizer is removed through a flow of sterile air. The removal of hydrogen peroxide is carried out with a nozzle equal to that used in the sterilization phase. The temperature of the air is about 60°C and the process starts 15second after the end of the previous phase. The injection of hot air has to be continued until the amount of Hydrogen peroxide inside the filled pouch decrease under 0,5ppm.

The simulation process has been carried out by means of a CFD commercial code, "Flow Simulation© (Dassault Systèmes SolidWorks Corporation). 3D CAD model of both the nozzle and the pouch have been created by SolidWorks© software.

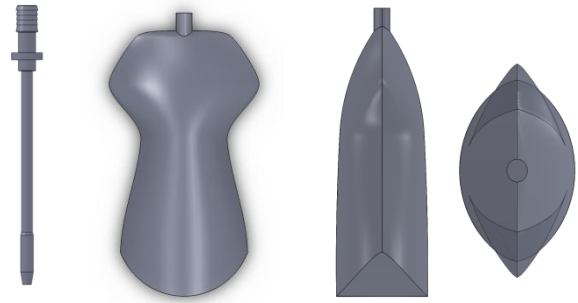


Figure 1: 3D CAD models of nozzle and pouch

The fluid adopted for the simulation was only air because Flow Simulation can't simulate a fluid composed by a mixture of air and hydrogen peroxide. However, this approximation does not create significant problems, being the volume of air approximately 99% of the composition of the mixture.

The turbulence model adopted was the Standard k-ε. The k-ε turbulence models is one of the most used in order to analyse this kind of problems (Ferziger and Peric, 2002; Margaris and Ghiaus, 2006; Bottani et al, 2008). It is part of the Reynolds Averaged Navier-Stokes models (RANS), which consider the average time of the speed to which add terms of fluctuation. In particular, the k-ε is a model with two equations, which means that it includes two additional equations to the classical ones to represent the properties of the turbulent flow.

In this model, the turbulent viscosity  $\mu_t$  is computed according to the following equation:

$$\mu_t = \rho C_\mu \frac{k^2}{\varepsilon}$$

where  $\rho$  is the density while turbulent kinetic energy  $k$  and its dissipation rate  $\varepsilon$  are derived from the resolution of the following set of equations:

$$\begin{aligned} \frac{\partial}{\partial x_j} (\rho u_j k) &= \frac{\partial}{\partial x_j} \left[ \frac{\mu_{eff}}{\sigma_k} \frac{\partial k}{\partial x_j} \right] + \mu_t \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \left( \frac{\partial u_i}{\partial x_j} \right) - \rho \varepsilon \\ \frac{\partial}{\partial x_j} (\rho u_j \varepsilon) &= \frac{\partial}{\partial x_j} \left[ \frac{\mu_{eff}}{\sigma_\varepsilon} \frac{\partial \varepsilon}{\partial x_j} \right] + C_{\varepsilon 1} \frac{\varepsilon}{k} \mu_t \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \left( \frac{\partial u_i}{\partial x_j} \right) - \rho C_{\varepsilon 2} \frac{\varepsilon^2}{k} \end{aligned}$$

$$\mu_{eff} = \mu + \mu_t$$

being  $\mu$  the molecular viscosity,  $\mu_t$  the turbulent viscosity of the fluid.

### 2.1. Mesh setting for the volume of the fluid

The whole body of the pouch is divided into a finite number of volumes on which the analysis is carried out. In "Flow Simulation©" these volumes are cube-shaped.

The number of cells used in the simulations was determined starting from a coarse meshing gradually refined, evaluating the changes in the results. The final mesh was determined when increasing the fineness of the mesh there were not significant improvements in the results.

The mesh was realized initially by creating a uniform subdivision, with a subsequent thickening in the critical areas of the fluid volume. In particular, finer mesh was used near the outlet section of the nozzle,

where is foreseeable that the shear rates would be higher, and close to the wall of the pouch, in order to simulate accurately the flow boundary layer. Since the pouch has a double symmetry, the simulations were carried out in a quarter of volume. This allowed us to optimize simulations run time using mesh with a high level of refinement.

The final mesh consists of about 150.000 cells in the first processing. In the process of sterilizing removal the number of cells has been reduced because the analysis is time-dependent and requires more computational resources.

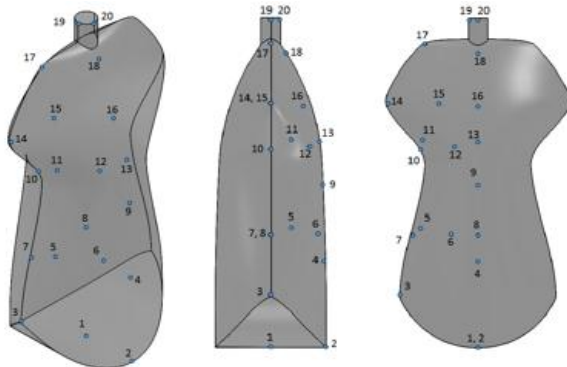


Figure 2: mesh adopted in simulation

With the aim of identifying the values of the flow parameters in significant points of the volume, 20 representative points have been selected, in critical positions, for the sterilization process. Each simulation permits the evaluation of the parameter values in each point.

## 2.2. Simulation settings

The simulations carried out for the sterilization process were stationary. It is not necessary to show the behaviour during the initial phase, but represent the process at the end of transitory time when the dynamics of the flow becomes stationary. Not considering the initial transitory time does not cause an excessive loss of information.

Two series of simulations were carried out. In the first series eight simulations has been designed with the aim to obtaining a better position of the nozzle. The second series consist of six simulations performed to get an optimized form of the nozzle.

The second series of the simulations concerns the optimization of the sterilizing agent removal. The objective of this phase is to assess the time required by the hot air to remove the sterilizing agent, composed by air and hydrogen peroxide.

Time-dependent simulations have been realized to evaluate the time needed to remove the sterilizing solution and replace it with sterile air. These simulations show the dynamic of the evacuation with a time step of 0.1 seconds. At the beginning the whole volume was full of the mixture. The removal of the hydrogen peroxide starts when the nozzle, with the same shape of the one used in the first simulations,

begins to blow sterile air inside the pouch. The flow of sterile air taken in the first round of simulations is 2000 NI/h. In order to carry out the simulations it was necessary to adopt a simplification. Since the code "Flow Simulation©" is not able to simulate a mixture of air-hydrogen peroxide, only air at 2 different temperatures was used, in order to distinguish the two fluids (mixture sterilizing and sterile air). Secondly the coefficient of thermal conductivity of the air has been reduced to avoid having results skewed by the phenomenon of heat transfer.

Two series of simulations have been designed. The first series consists of six simulations to obtain a better position and shape of the nozzle. In the second series seven simulations have been performed, with different flow rate of sterile air to find out which flow rate permits to reduce the amount of sterile air consumed.

In this case it was necessary to decrease the mesh refinement level in order to reducing the run time simulation without compromising the results reliability

## 3. RESULTS

### 3.1. Optimization of sterilization process

In the first phase eight simulations were carried out with different nozzle positions.

The first simulation was performed with the nozzle at a distance of 15 mm from the bottom of the pack . The next simulations were performed with step of 15 mm.

Table 1: simulations realized with different nozzle position

Simulation Number	Distance nozzle – bottom pouch (mm)	Simulation Number	Distance nozzle – bottom pouch (mm)
1	15	5	75
2	30	6	90
3	45	7	105
4	60	8	120

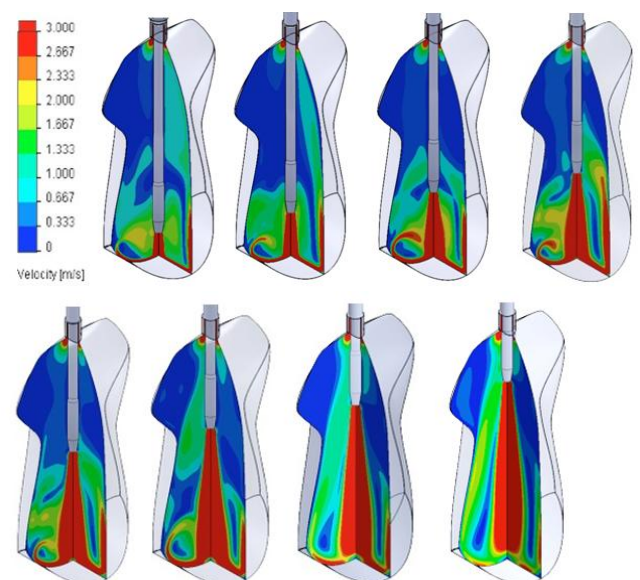


Figure 3: Velocity of the sterilizer

As clearly marked in Figure 3, the conclusions that can be assumed from this first phase of simulations are the following:

- In the first simulations, where the nozzle is in a lower position in the box, the flow of the sterilizing agent reaches a high rate near the bottom of the pouch. In the upper part instead there are large areas at very low speed. The tests performed with the nozzle in upper positions have however good velocity in the lower part and a more uniform flow at the top. The simulations 6, 7 and 8 are those that distribute more evenly the speeds.
- In all simulations carried out, the lateral areas of the upper part of the pouch never reaches significant rate.

For the reasons just mentioned the best position to put the nozzle during the sterilization phase appears to be closer to position 7, in which the nozzle has a distance of 105 mm from the bottom of the envelope. The problem to be solved is related to the low speed in the zones of the upper side. This problem can be solved by changing the nozzle shape. Figure 4 shows simulation 7: as can be seen from the related chart the speed of the fluid in the circled points is very low.

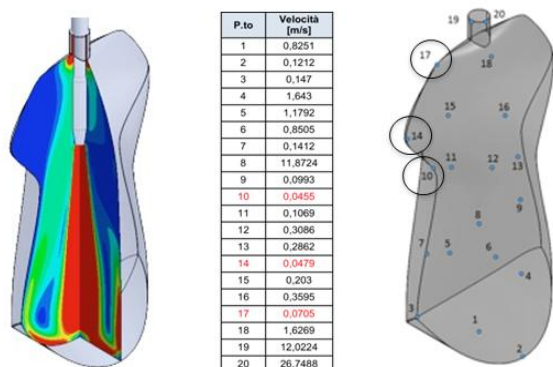


Figure 4: Air velocity: simulation 7

The problem of low velocity of the mixture in the lateral zones of the upper side of the pouch could be solved by introducing holes on the surface of the nozzle. Some simulations have been made introducing 4 holes of a radius of 2mm, which have been placed at a distance of 15 mm from the exit section of the nozzle.

Two parameters, that could be potentially critical in determining the velocity profile of the mixture, are the radius of the outlet section and the thickness of the nozzle close to the holes. Six simulations has been realized, all with the same position of the nozzle as in simulation 7, varying the two parameters.

Table 2: simulations with different nozzle geometry

Number of simulation	Radius of the outlet section (mm)	Thickness of the nozzle in correspondence to holes (mm)
1	2,5	2,1
2	2,25	2,1
3	2	2,1
4	2,5	1
5	2,25	1
6	2	1



Figure 5: Nozzle modified by the insertion of 4 holes

Simulation 3 shows the best solution found. The nozzle has the outlet section with a radius of 2 mm and a thickness close to the holes of 2.1 mm. The reduction of thickness near to the holes, did not produced improvements because the speed did not achieve significant values particularly in the frontal area of the central part of the pouch.

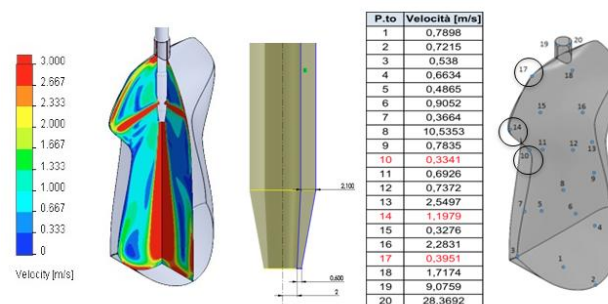


Figure 6: map of speeds and local values of speed of simulation 3, which has guaranteed the most uniform flow

### 3.2. Removal process of the sterilizer

Six simulations has been performed in the first phase as summarized in Table 3. In these simulations the flow rate of sterile air is set at 2000 NI/h.

Table 3: simulations with each nozzle position and geometry

Number of simulation	Nozzle position	Nozzle typology
1	high	nozzle without lateral holes
2	Middle	nozzle without lateral holes
3	Low	nozzle without lateral holes
4	high	nozzle with 4 lateral holes with radius of 2 mm
5	Middle	nozzle with 4 lateral holes with radius of 2 mm
6	Low	nozzle with 4 lateral holes with radius of 2 mm

Simulations were also carried out with a change in the exit section and the thickness of the nozzle, as in the



previous simulation series. The results were very similar to simulations 4, 5 and 6, with a slight worsening in terms of removal time. The upper position has a distance between the nozzle and the bottom of the envelope of 110 mm, the middle position of 70 mm and the lower position 30 mm.

The best solution is that of simulation 5 in which the removal of the sterilizing agent is achieved in 4,6 seconds.

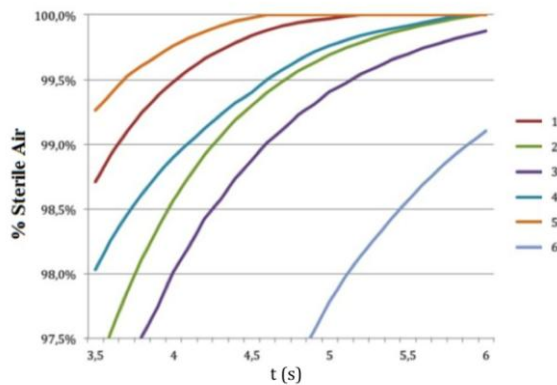


Figure 7: Comparison of percentages' trends of sterile air in the 10 most critical points for the different simulations

Figure 8 shows the trend in the removal process over the time that picks out the replacement of the sterilizer with sterile air. The colour blue corresponds to the sterilizer and the red to the sterile air. As previously explained the simulated process is not the same process of removing the sterilizer with sterile air, because it is not permitted by the software. We have simulated the two fluids with air, and to distinguish the two fluids we have assigned them different temperatures. The VHP assumed a temperature of 55°C and the sterile air at temperature of 200°C. Values of kinematic viscosity and density were set not dependent on the temperature and equal to air values at 55°C Depending on the temperature detected in any point is possible to go back to what percentage of the two fluids is present in that instant. In the coloured map is assigned the blue colour to the areas with a temperature below 198°C and the red colour for temperature between 198°C and 200°C.

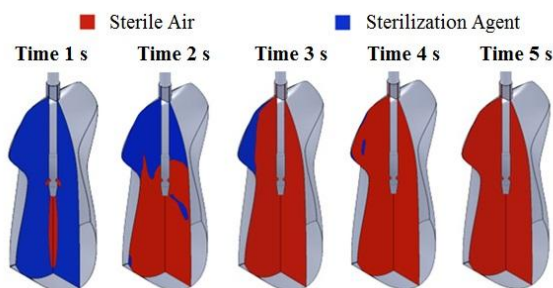


Figure 8: dynamics of the removal process of the sterilizer in the configuration of simulation 5

In the simulations just explained the air flow rate was set at 2000 NI/h as in all the tests.

The objective of this second phase is to determine the optimal flow of sterile air. The optimal range is the one that can get the best compromise between the amount of air used and the size and complexity of the removal station. The objective is therefore to have a flow that minimizes the amount of air needed and a low rinsing time in order to reach a high rate and consequently a reduced number of nozzles. To do this, 7 additional simulations were carried out with the same shape as in simulation 5, with different flow rate of sterile air. The rate values are those reported in Table 4.

Table 4: set flow rate in the last simulations

Number of simulation	Flow rate (NI/h)
1	1500
2	1750
3	2250
4	2500
5	2750
6	3000
7	3250

The obtained time values of sterilizer removal are summarized in Figure 9.

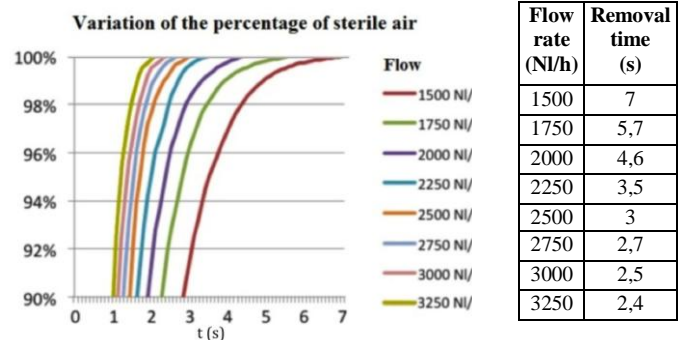


Figure 9: time of sterilizer removal

The needed amount of sterile air is given by the product of flow rate and minimum removal time of the sterilizing agent. The minimum amount of sterile air is achieved with a flow rate of 2750 NI/h (Figure 10). Values slightly higher are obtained with flow rates of 2500 and 3000 NI/h. It is possible to conclude that the optimal range of flow rate is between 2500 and 3000 NI/h.

Flow Rate (NI/h)	Amount of sterile air (NI)
1500	2,91
1750	2,77
2000	2,56
2250	2,18
2500	2,08
2750	2,06
3000	2,08
3250	2,17

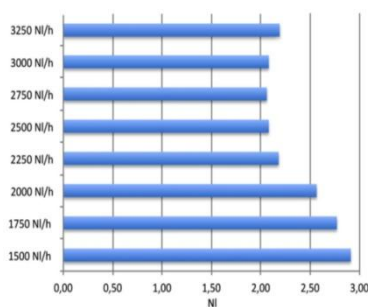


Figure 10: amount of sterile air required for the process

The increase in flow rate, compared to the previous case, permits a reduction of sterile air and allows to reduce the number of nozzles required for this process. The reduction of nozzles can be calculated as follows.

$$\text{Reduction of nozzles [\%]} = \frac{4,4-2,7}{4,4} \times 100 = 39\%$$

The values of 4.4seconds and 2.7seconds are respectively the removal time in the case of 2000 NI/h and the removal time in the case of 2750 NI/h.

The time required to remove the sterilizer has been validated by empirical tests. The test has been realized by executing the normal process of sterilization and then injecting sterile air with the flow rate of 2750 NI/h. The flushing time of sterile air inside the envelope was:

- 3 seconds in the first test
- 5 seconds in the second test

For each test carried out at the end of this process the bag was filled with 330 ml of water without mineral salts, and shacked to ensure that all the residual hydrogen peroxide could dissolve in the solution. To measure the residual value a tool based on reflectometry technique has been used. After this a reagent strip was put inside the pouch for 15 seconds and subsequently inserted into the detector. The concentration is calculated as a function of the redox reaction that occurs. The strips are able to measure the residual values between 0.2 ppm and 20 ppm. The following table shows the concentration reported in ppm.

Table 5: residual of hydrogen peroxide in the empirical tests

Flow rate	Time	Residual
2750 NI/h	3 s	0,6 ppm
	5 s	0,2 ppm

The data obtained are coherent to the one calculated, even if they show a longer time. In fact, the number of ppm after 3 seconds is slightly higher than the maximum allowed (0,5 ppm), while in the calculation carried out after 2.7 seconds the condition of sterility would be reached. After 5 seconds the hydrogen peroxide is completely removed (the indicator always measures a minimum of 0.2 ppm).

#### 4. CONCLUSION

The aim of this study was to analyze and improve the

sterilization process of spouted pouch for an aseptic line.

The first part of the research was aimed at determining the position and the geometry of the nozzle, which could ensure a flow of sterilizing solution with significant velocity values in each part of the pouch, avoiding areas where the speeds are too low. Simulations allowed reaching the best combination between nozzle position and geometrical shape of output section.

In the second part the removal process of sterilizing agent has been analyzed in order to estimate the time required for this operation. Inside the pouch was introduced hydrogen peroxide at a concentration of 5000 ppm and the maximum residual value that may remain in the finished product is 0.5 ppm. Among the simulations carried out the one that has achieved the best results has the nozzle located 70 mm above the bottom of the envelope, with 4 lateral holes that have the same characteristics of those included in the solution obtained in the previous phase. This solution guarantees the complete removal of the sterilizer in the shortest possible time (4.6 seconds), assuming a flow of sterile air of 2000 NI/h.

The last phase of this research has tried to optimize the flow of sterile air in order to obtain the value of flow rate which could ensure the complete removal of the sterilizer with the lowest air consumption. The optimal value of flow rate was 2750 NI/h and the optimal range is between 2500 and 3000 NI/h. Taking as flow rate 2750 NI/h a significant savings of required quantity of sterile air has been obtained. In particular, the consumption for each pouch has become 2.08 NI instead of 2.44 NI that was the amount of the case previously chosen. As a second important advantage, this choice reduces the number of nozzles and consequently the size and complexity of the carousel.

For the sterilization process it would be important to perform experimental tests to evaluate the effect of the obtained improvement. New experimental tests should be carried out. for the process of sterilizer removal in order to calculate the residual concentration of hydrogen peroxide with more precise equipments that allow to obtain data with a lower margin of error.

#### REFERENCES

- Abdul Ghani, A.G., Farid, M.M., Chen, X.D., Richards, P., 2001. Thermal sterilization of canned food in a 3-D pouch using computational fluid dynamics. *Journal of Food Engineering* 48 (2), 147-156;
- Bottani E., Rizzo R., Vignali G. (2008). Numerical Simulation of Turbulent Air Flows in Aseptic Clean Rooms. In: Petrone Giuseppe, Cammarata Giuliano. *Recent Advances in Modelling and Simulation*. (pp. 633-650). ISBN: 978-3-902613-25-7. VIENNA: I-TECH (AUSTRIA).
- Castle, L., Mercer, A.J. and Gilbert, J. 1995. Chemical migration from polypropylene and polyethylene aseptic food packaging as affected by hydrogen peroxide sterilization. *J. Food Prot.* 58, 170–174.

- Farkas, J. 1998. Irradiation as a method for decontaminating food: A review. *International Journal of Food Microbiology*, 44 (3), 189-204
- Ferretti, G., Montanari, R., Rizzo, R., Vignali G., 2006. Ionising Radiation for Food Packaging Sterilisation. *Proceeding of 3rd Central European Congress on Food*, Sofia, Bulgaria, May 22-24;
- Ferziger J.H. and Peric M., 2002. Computational methods for fluid dynamics, Springer-Verlag, Berlin/Heidelberg, Germany.
- Jun, S., Sastry, S., 2007. Reusable pouch development for long term space missions: A 3D ohmic model for verification of sterilization efficacy, *Journal of Food Engineering*, 80,1199–1205;
- Klapes, N.A., Vesley, D., 1990. Vapor-phase hydrogen peroxide as a surface decontaminant and sterilant. *Applied and Environmental Microbiology*, 56 (2), 503-506
- Margaris D.P. and Ghiaus A.G., 2006. Dried product quality improvement by air flow manipulation in tray dryers, *Journal of Food Engineering* 75 542–550.
- Riganakosa K.A., Kollerb W.D., Ehlermannb D.A.E., Bauerb B. and Kontominasa M.G., 1999. Effects of ionizing radiation on properties of monolayer and multilayer flexible food packaging materials *Radiation Physics and Chemistry*, 54, (5), 527-540
- Robertson G.L., 2006. *Food Packaging: Principles and Practice*, Second Edition CRC Press, Boca Raton FL (USA)

## **AUTHORS BIOGRAPHY**

Paolo Casoli is associate professor in fluid machinery for food industries at the Faculty of Engineering of the University of Parma

Gabriele Copelli is a Research Fellow of the Centro Interdipartimentale SITEIA.PARMA

Michele Manfredi was born in Parma (Italy) in 1986, has achieved a master degree in mechanical engineering of food industry, and in 2011 he won a scholarship regarding the optimization of food packaging.

Giuseppe Vignali is an Assistant Professor in Mechanical Plants for food industries at the Faculty of Engineering of the University of Parma

## **Authors' Index**

- Abdelhadi 8  
Abdelkrim M. N. 96, 101  
Abdelkrim N. 96  
Abdul-Hassan 1  
AbouRizk 330  
Aichinger 65  
Aickelin 31  
Aiello 70  
Alix 406  
Al-Otaibi 1  
Aririguzo 126  
Armenzoni 281  
Aubrun 101  
Avai 113  
Axelos 412  
Azar 232  
Baldino 173, 242  
Bao 226  
Barreiro 195  
Becker 289  
Ben Mabrouk 322  
Benini 166  
Bennett 383  
Biagini 495  
Böhm 80  
Bosse 189  
Bruno 90  
Casoli 503  
Castro 107  
Challouf 101  
Concha 370  
Coppelli 503  
Corrêa 361  
Crespo Pereira 375  
Cunha 464  
Dal Maso 113  
Danciu 353  
Daudin 412  
De Cindio 173, 242, 488  
De Felice 397  
de Souza Santos 361  
del Rio Vilas 375  
DellaValle 412, 419  
Demidovs 216  
Dias 389  
Donald Combs 185  
Dong 226  
Ekyalimpa 330  
El-Said 56  
Emond 383  
Enea 70  
Falcone 202  
Faucher 452  
Fernandes 464  
Ferretti 281  
Flores 427  
Forcina 202  
Fouad 56  
Fox 22  
Frick 255  
Frydman 442  
Gabriele 173, 242  
Gamberini 437  
Gebennini 437  
Görges 160, 273  
Gradl 347, 353  
Grassi 437  
Gschaider 149  
Gutierrez 427  
Hasegawa 249, 295  
Hecker 289  
Hegmanns 303  
Henderson 90  
Hillbrand 259, 266  
Holsteyns 149  
Hong 330  
Horton 189  
Hosseinzadeh 45  
Hussein M. A. 289  
Hussein W. B. 289  
Ikhlef 16  
Jafarnejad 45  
Jedermann 195  
Jiménez 50  
Junk 149  
Kádár 155  
Kadkhodae 120  
Kadri 8  
Kangarani 45  
Kansou 419  
Karwowski 370

Kerschl 61	Nogales 80
Kienegger 347	Núñez 107
Kopytov 216	Oancea 80
Krcmar 347, 353	Olmos 427
Krull 189	Oueslati 322
Kuhlmann 149	Pacitto 202
Kuhn 80	Palafox-Albarrán 195
Lang 195	Palau 80
Laws 22	Pan 314
Lechner 149	Paul 314
Leitão 389	Pedrazzoli 113
Lescheticky 80	Peito 389
Liu 226	Pereira 389
Longo 480, 488	Perrot 412
Lopes 464	Pessina 471
Lupi 173, 242	Petrillo 397
Madeo 480	Petukhova 216
Malisuwan 207, 211	Pfeiffer 155
Mami 322	Pinaton 442
Manfredi 503	Pizzuti 488
Marçal de Queiroz 361	Pogány 61
Marchini 281	Popovics 155
Márquez 107	Prieling 149
Marzouk 56	Qian 226
Massei 471, 480	Raciti Castelli 166
Mateos 50	Ramis 370
Matthies 160	Rauh 80
Mayer 347, 353	Rego Monteil 375
Meglouli 16	Renard 412
Menachof 31	Renwick 139
Menassa 232	Revoredo-Giha 139
Menshikov 40	Rimini 437
Merah 8	Rios Prado 375
Min Kim 90	Robla 195
Mirabelli 488	Rodríguez 107
Monostori 155	Rovere 113
Montanari 281	Roy 80
Mota 464	Ruiz-García 195
Mouss L.H. 8	Saad 126
Mouss M.D. 8	Sautot 412
Mucherino 309	Schlager 80
Muriana 70	Schmid 259, 266
Najafi 120	Schöech 266
Nandhakwang 211	Scholz-Reiter 160, 273
Ndiaye 419	Schreiber 347
Nerakae 295	Schuöcker D. 65
Ngoc Hieu 249	Schuöcker G. 65

Sebedio 412  
Sepulveda 370  
Seta 173, 242  
Severinghaus 185  
Shahbazi 136  
Sherman 31  
Siebers 31  
Silvestri 202  
Solari 281  
Solis 314  
Staudegger 149  
Steiner 149  
Steininger 179  
Steward 361  
Tarone 471  
Tellili 96, 101  
Toniato 166  
Toth 303  
Tovia 427  
Tremori 471  
Turi 495  
Urtubia 309  
Uysal 383  
Valerio 427  
Veeraklaew 207, 211  
Vén 155  
Vesecky 22  
Viale 442  
Vignali 503  
Vita 149  
Wang 226  
Wittges 347, 353  
Zacharewicz 406  
Zeltmate 338  
Zhang 226