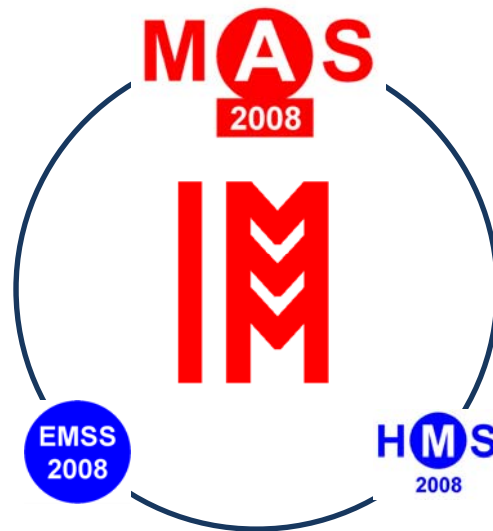# 7TH INTERNATIONAL WORKSHOP ON MODELING & APPLIED SIMULATION

CALABRIA

SEPTEMBER 17 - 19 2008
CAMPORA S. GIOVANNI
ITALY

INTERNATIONAL MEDITERRANEAN AND LATIN AMERICAN MODELING MULTICONFERENCE

EDITED BY:

AGOSTINO BRUZZONE

PRISCILLA ELFREY

ENRICO PAPOFF

MARINA MASSEI

ISTVAN MOLNAR

MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA

M&S CENTER - LABORATORY OF ENTERPRISE SOLUTIONS

DIPTEM, UNIVERSITY OF GENOA

LIOPHANT SIMULATION

THE 7ᵀᴴ INTERNATIONAL WORKSHOP ON MODELING & APPLIED SIMULATION

SEPTEMBER 17-19 2008
CAMPORA S. GIOVANNI (AMANTEA, CS), ITALY

MAS
2008

EDITED BY

AGOSTINO BRUZZONE
PRISCILLA ELFREY
ENRICO PAPOFF
MARINA MASSEI
ISTVAN MOLNAR

THE 7TH INTERNATIONAL WORKSHOP ON MODELING & APPLIED SIMULATION

CAMPORA S. GIOVANNI (AMANTEA), ITALY
SEPTEMBER 17-19, 2008

ORGANIZED BY

MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA

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

DIPTEM – UNIVERSITY OF GENOA

LIOPHANT SIMULATION

SPONSORED BY

ACADEMIC, INSTITUTES AND SOCIETIES SPONSORS

IEEE – INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS

SCS – SOCIETY FOR COMPUTER SIMULATION INTERNATIONAL

MISS – McLEOD INSTITUTE OF SIMULATION SCIENCE

M&SNET – MODELING & SIMULATION NETWORK

IMCS – INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF SIMULATION

NASA – KENNEDY SPACE CENTER

INDUSTRY SPONSORS

MAST – MANAGEMENT AND ADVANCED SOLUTIONS AND TECHNOLOGIES

TONNO CALLIPO

LIQUIRIZIE AMARELLI

GIULIO BARCA PRODOTTI IN PELLE

EDITORS

AGOSTINO BRUZZONE
MISS-DIPTEM, UNIVERSITY OF GENOA
VIA MOLINERO 1
17100 SAVONA, ITALY
agostino@itim.unige.it
www.st.itim.unige.it

PRISCILLA ELFREY
NASA, KENNEDY SPACE CENTER, 32899, FL, USA
priscilla.r.elfrey@nasa.gov
www.ksc.nasa.gov

ENRICO PAPOFF
MECHANICAL DEPARTMENT
UNIVERSITY OF CALABRIA
VIA P. BUCCI – 87036, RENDE (CS), ITALY
epapoff@unical.it
www.ingegneria.unical.it/impiantiindustriali

MARINA MASSEI
LIOPHANT SIMULATION
VIA MOLINERO 1,
SAVONA 17100, ITALY
marina.massei@liophant.org
www.liophant.org

ISTVAN MOLNAR
BLOOMSBURG UNIVERSITY OF PENNSYLVANIA
400 EAST SECOND STREET, BLOOMSBURG PA, USA
Molnar01@aol.com
www.bloomu.edu

THE INTERNATIONAL MEDITERRANEAN AND LATIN AMERICAN MODELING MULTICONFERENCE, I3M 2008

GENERAL CO-CHAIRS
AGOSTINO BRUZZONE, *MISS DIPTEM, UNIVERSITY OF GENOA, ITALY*
MIQUEL ANGEL PIERA, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*

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

THE 7ᵀᴴ INTERNATIONAL WORKSHOP ON MODELING & APPLIED SIMULATION, MAS 2008

GENERAL CO-CHAIRS
PRISCILLA ELFREY, *NASA, KENNEDY SPACE CENTER, USA*
ENRICO PAPOFF, *MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA, ITALY*

PROGRAM CO-CHAIRS
MARINA MASSEI, *MAST, ITALY*
ISTVAN MOLNAR, *BLOOMSBURG UNIVERSITY, USA*

MAS 2008 International Program Committee

Karlos Alastruey, *University of Navarra, Spain*
David Al-Dabass, *Nottingham Trent University, UK*
Alex Alexandridis, *University of Athens, Greece*
Stefano Armenia, *University of Rome Tor Vergata, Italy*
Krassimir Atanassov, *Sofia University, Bulgaria*
Naamane Aziz, *LSIS, France*
Marian Barbu, *University of Galati, Romania*
Enrico Bocca, *Liophant Simulation, Italy*
Dirk Briskorn, *Christian-Albrechts-Univ. Kiel, Germany*
Agostino Bruzzone, *University of Genoa, Italy*
Franco Cicirelli, *University of Calabria, Italy*
Antonio Cimino, *MSC-LES, University of Calabria, Italy*
Duilio Curcio, *MSC-LES, University of Calabria, Italy*
John Darzentas, *University of the Aegean, Greece*
Fabio De Felice, *University of Cassino, Italy*
Priscilla Elfrey, *Kennedy Space Center, USA*
Maria Pia Fanti, *Polytechnic of Bari, Italy*
Angelo Furfaro, *University of Calabria, Italy*
Claudio Garuti, *Fulcrum Engineering Ltd., Cile*
Norbert Giambiasi, *LSIS, France*
Oke Gulay, *Imperial College London, UK*
Janos Sebestyen Janosy, *KFKI, Hungary*
Helen Karatza, *University of Thessaloniki, Greece*
Jerzy Jozefczyk, *Wroclaw Univ. of Technology, Poland*
Francesco Longo, *MSC-LES, University of Calabria, Italy*
Roberto Lu, *Boeing Company, USA*
Marina Massei, *MAST, Italy*
Giovanni Mirabelli, *University of Calabria, Italy*
Istvan Molnar, *Bloomsburg University, USA*
Libero Nigro, *University of Calabria, Italy*
Enrico Papoff, *University of Calabria, Italy*
Pit Pichappan, *Al Imam University, Saudi Arabia*
Dilip Kumar Pratihar, *Indian Inst. of Technology, India*
Fengxiang Qiao, *Texas Southern University, USA*
Constantinos Soutis, *The University of Sheffield, UK*
Zhou Suiping, *Nanyang Technological Univ., Singapore*
Simon Taylor, *Brunel University, UK*
Walter Ukovich, *University of Trieste, Italy*

Tracks Chairs

Decision Support Systems Applications
Chair: Fabio De Felice, *University of Cassino (Italy)*

Inventory Management Simulation
Chair: Antonio Cimino, Duilio Curcio, *Modeling & Simulation Center - Laboratory of Enterprise Solutions, University of Calabria (Italy)*

Emergency Planning and Management
Chair: Zhou Suiping, *Nanyang Technological University (Singapore)*

Agent and Service based Modelling & Simulation
Chair: Angelo Furfaro, *University of Calabria (Italy)*

Automation
Chair: Naamane Aziz, *Laboratoire des Sciences de l'Information et des Systèmes (France)*

Modeling Information Security
Chair: Pit Pichappan, *Al Imam University, Riyadh (Saudi Arabia)*

Generalized Nets
Chair: Krassimir Atanassov, *Sofia University, (Bulgaria)*

Industrial Engineering
Chair: Karlos Alastruey, *Universidad Pública de Navarra (Spain)*

Wastewater Treatment Processes Modelling & Simulation
Chair: Marian Barbu, *Dunarea de Jos University of Galati (Romania)*

Artificial Intelligence Techniques in Modelling and Simulation
Chairs: Alex Alexandridis, *National Technical University of Athens (Greece)*; Oke Gulay, *Imperial College London (UK)*

Energy Production and Distribution
Chair: Janos Sebestyen Janosy, *KFKI Atomic Energy Research Institute (Hungary)*

Advance in Information and e-Integration in Healthcare Systems and Management
Chairs: Maria Pia Fanti, *Polytechnic of Bari (Italy)*; Walter Ukovich, *University of Trieste (Italy)*

Production Simulation
Chairs: Roberto F. Lu, *The Boeing Company, Commercial Airplanes Material & Process Technology (USA)*

GENERAL CO-CHAIRS' MESSAGE

WELCOME TO MAS 2008

MAS2008 provides the scientific community with a high quality technical/scientific framework within the International Mediterranean and Latin American Modeling Multiconference (I3M). The I3M represents, together with the European Modeling & Simulation Symposium (EMSS) and the International Workshop on Harbor, Maritime & Multimodal Logistics Modeling & Simulation (HMS), a scientific event involving different topics related to Modeling & Simulation.

The main goal of MAS 2008 is the creation of a unique framework for supporting real networking among scientists, technicians and users for increasing opportunities in establishing new cooperation in applied simulation fields.

MAS 2008 edition achieved the largest results in term of attendance and papers selection rate, thanks to the valuable work of MAS International Program Committee that assured high scientific quality and scholarly relevance of the contributions.

In addition to the scientific issues, the wonderful framework of Calabria coast (South Italy riviera) offers a glimpse of two thousand years of history. Calabria was part of Magna Graecia, the site of rich and powerful Greek colonies, where artifacts such as the famous Riace bronze sculptures, recovered from the sea few years ago, are now proposed to the visitors in the National Museum of Reggio Calabria.

Among the region remaining curiosities are Greek-speaking villages, Albanian communities (dating back five hundred years), hundreds kilometers of coast facing onto two seas, the Tyrrhenian and the Ionian. Calabria has several of the most appealing sites in Italy, such as Tropea, a Baroque seaside town, and Scilla, while Cosenza can offer relics of Gothic domination. The nature is outstanding: wonderful beaches, national parks and much more. The Sila is a high-level plain with lakes, rolling hills and woodland, while the Aspromonte is like a rocky backbone to the region.

MAS2008 location makes possible a short but fruitful period of suspension from ordinary activities, giving to all the attendees the opportunity to take the most from the event, not a local or regional Congress, but a worldwide community meeting, well represented both by the attendees and the different Thematic Conference Organizations. This confirms the fact that MAS and I3M are truly international context with very strong liaisons to all excellence centers in Modeling & Simulation.

Next year the event will be held in Tenerife, Canary Island, and we look forward to an even large event. However, we are glad to welcome you in MAS 2008 within the wonderful framework of Calabria.

PRISCILLA ELFREY, NASA - KENNEDY SPACE CENTER,

ENRICO PAPOFF, MECHANICAL DEPARTMENT, UNIVERSITY OF CALABRIA

ACKNOWLEDGEMENTS

LOCAL ORGANIZATION COMMITTEE

ANTONIO CIMINO, *UNIVERSITY OF CALABRIA, ITALY*
DUILIO CURCIO, *UNIVERSITY OF CALABRIA, ITALY*
FRANCESCO LONGO, *UNIVERSITY OF CALABRIA, ITALY*
GIOVANNI MIRABELLI, *UNIVERSITY OF CALABRIA, ITALY*
ENRICO PAPOFF, *UNIVERSITY OF CALABRIA, ITALY*
ERIKA PINTO, *UNIVERSITY OF CALABRIA, ITALY*
CLAUDIA RUBINO, *CHAMBRE OF COMMERCE, CROTONE, ITALY*

Index

Inventory Management Simulation

Logistics, Manufacturing & Supply Chain Management

Decision Support Systems Applications

Advance in Information and E-Integration in Healthcare Systems and Management

Industrial Engineering Modeling and Simulation

Agent and Service based on Modeling & Simulation

MAS General Session

Author's Index

A Study of Materials Management System in a Large Mining Organization

Dr. Sharif, Dean of Faculty Affairs Office, IIT Kanpur-
208016. (India)
Email: sharif@iitk.ac.in

ABSTRACT

Inventories are indispensable to run the plant efficiently. It helps to maintain both material inputs and the manufactured output at an optimum level of requirement. While inventories are necessary to absorb input and demand variations, a balance is needed between blocked capital in the inventories and impact on production. The materials management department is expected to provide operational convenience with a minimum possible investment in inventories. In this paper I would discuss the Inventory Management system followed in a Large Mining Organization owned by Government of India. This study addresses the overall inventory management system followed at one of the units of this organization. The study focuses on understanding and analyzing current inventory system being practiced in the organization. Further, a set of recommendations to improve the system are also made.

Key Words: Inventories, Inventory Management System, Current Inventory System, Optimum level

1. INTRODUCTION

Inventory management covers effective management in all its aspects including materials planning, programming and purchasing, inventory control, receiving, warehousing and store keeping, materials handling and disposal of scrape & surplus stores. The importance of materials management can be emphasized in India where the cost of material accounts for nearly 60% of the production / total cost. The facts highlight the importance of the efficient and effective use of materials particularly in the present time of credit squeeze and unstable market conditions.

Inventories are very important to the management of an industrial undertaking as they have direct bearing upon the profits of the concern. Since profits are affected by the element cost, and inventories constitute nearly two third of the total cost of the production, their effective control is now the primary concern of modern management.

2. OBJECTIVE OF THE STUDY

The objective of this study is to critically examine and understand the practices and performance of Materials Management functions adopted in an Industrial Organization. This case study helps to know the functioning of an organization and also highlights the various problems being faced within the organization. This study addresses the overall inventory management system followed at one of the units of this Manufacturing organization. The study brings out the detailed analysis of inventory management system followed in the above organization, the observations, issues and valuable suggestions.

3. ORGANIZATION

The organization is a Manufacturing one and is a vertically integrated metal producer in India engaged in a wide spectrum of activities ranging from exploration, Mining, Beneficiation, Smelting, Refining and Continuous Cast Rod manufacturer. Types of stores maintained at this organization are –

- Raw materials
- Spares
- Operational supplies
- Consumables and other supplies.

4. THE METHODOLOGY USED FOR THE STUDY

- Collection of Data using field visits, direct interviews and telephonic interaction.
- Case study method
- Contents analysis and summary of observations
- Issues and valuable suggestions

5. CURRENT PURCHASE AND INVENTORY SYSTEM AT THIS ORGANIZATION

The following are the salient features of current procedures of inventory system at this minining organization –

5.1. Stores and Inventory Management

Inventory Policy:
The organization maintains 3 months inventory on staggered delivery basis at its central stores. Issues are made to the site stores of the consuming departments against their demand.

Spares Inventory System:
To maintain and run the machine / equipment, the organization maintains spare parts inventory of about

25%. The spare parts inventory is divided in three groups as follows:

- Running maintenance inventory
- Over running (preventive maintenance) inventory
- Over hauling inventory

Stock Items:
The following items are kept as stock items at this organization.

- PVC cables
- Copper winding wire
- Nuts and volts
- Bulbs, tube lights, choke and starter
- PVC tape
- HRC fuse

Supply to shop floor:
Besides central stores, site stores are created to meet the day today requirement of shop floor. Central stores make supplies to site stores as per
their requirement and site stores issue the material to shop floor people.

Warehouse Policy:
Items have been distributed in different groups centrally (In Central Stores) keeping in mind the requirement of the indenting departments. The six digits and ten digits codification system is followed. All items are codified. Items are stocked nature wise and tagging is done for easy identification. Apart from central store there exist site stores also. The items are issued from central stores to the site stores of the indenting department.

5. 2. Inventory Costing Policy:
For costing the materials Weighted Average method is adopted. This method is quite straight forward, it takes the weighted average of all units available for sale during the accounting period and then uses that average cost to determine the value of cost of goods sold and ending inventory. For example, let's say that a bakery produces 200 loaves of bread on Monday at Rs.1 each, and 200 more on Tuesday at Rs.1.25 each. The Weighted average cost for inventory would be Rs1.125 per unit, calculated as (200 x Rs1 + 200 x Rs1.25) ÷ 400 = Rs.1.125

Rate Contract Policy:
Following items are under rate contract:
Safety items:- like Gum boots, leather shoes, helmet cells, belts and Globes
Explosives:- like drill rods, drilling spares and

Automobile spares are covered under rate contract.

Control System:
The measures taken for cost reduction and controlling the purchase costs are as follows:

- Cost reduction through entering into rate contracts.
- Cost reduction by cutting consumption after scrutinizing the indenter's requirement by a well-established Inventory Control Cell.
- Cost reduction by negotiating the rates.

5.3. Procurement Cycle:
Figure 1.0 describes the procedural cycle at this organization.

vendors. Offers are opened on a given date and time by a committee consisting of a purchase officer, a finance officer and one more representative from the purchase department. The vendors are allowed to be present while opening the quotations. A comparative chart is prepared and sent to the indenting department for technical recommendations. Normally, lowest offer is preferred unless it is technically not suitable. After recommendation, the draft supply order is prepared by the purchase department and submitted to the finance department for financial concurrence and subsequent Executive Director's approval. On receipt of the approval of the Executive Director, the purchase order is placed to the firm. Regular follow – up is done to receive the supply in time. If the supplier fails to supply the material in time, he is penalized for late delivery.

Fig. 1.0

As can be seen from the above schematic diagram of purchasing, the consuming department raises the indent and takes non availability certification from the stores. The indenter generally suggests names of the suppliers if known to him. The indent is than submitted to the inventory control cell where the indent is scrutinized and material code number is allotted to the item under procurement. The indent is than passed on to the purchase department for procurement of the item. Purchase department makes an entry in the register and than the enquiry is floated to the prospective suppliers. Four kinds of enquiries are floated depending upon the nature and value of the item. (1) Limited tender enquiry (2) Single tender enquiry (3) Open tender enquiry and (4) Global tender enquiry. Sealed quotations are called from the

When the material is received, it is inspected by the inspection team from the indenting department. Entries are made in the daily receipt book (DRB). Entries are also made in the computers. Receipt voucher (RV) is prepared and sent the finance for settlement of the bill of the supplier.

5.4. Selective Inventory Control System

Adoption of Inventory Control Techniques:-
Automatic indenting is done for 'C' class (general consumable) items. However, for 'A' and 'B' class items, manual reviewing is done on regular basis.

FSN classification is adopted in this organization to control the obsolete items whether spare parts, or raw materials or components. Items are classified as Fast moving, Slow moving and Non-moving – FSN - according to their consumption pattern. Items issued and consumed in the last one year are placed in the Fast moving category, items consumed after a year but within three years are placed in the Slow moving category, and items consumed after three years or not consumed at all are placed in the Non moving category. FSN classification is done at the time of accounts closing. It excludes the insurance items/ equipments.

ABC Analysis: Every year computer section carries out an ABC analysis for all stock items furnishing the items in descending order of their annual consumption value and classifying them under the following heads:

- A Items: Initial 70% of consumption value.
- B Items: Following 20% of consumption value.
- C Items: Following 10% of consumption value.

To ensure that the inventory level of the stock item is kept at reasonable level, the safety stock of one month is kept in the case of 'A' items, 2 months for 'B' items and 6 months for C items. Normally, annual ordering is done in the case of 'A' items with phased delivery schedule to ensure that minimum level as prescribed above of the safety stock is maintained. In the case of certain 'B' items also the annual ordering is resorted to. The ordering in the case of 'C' items is done on the basis of the indents raised by the computer section. However, there is built in provision for increase or decrease in quantity of consumption especially in the case of 'A' items by regular checks.

5.5. Administrative System

Man power Management:
The Materials department is headed by Assistant General Manager (AGM). Besides AGM, the department has many other officers. Each officer is assigned to procure a particular kind of material. Say for example, one is responsible to procure electrical items; the other is responsible for plant machinery and so on. Besides materials background, few persons from technical background too are placed in Stores and Purchase department.

6. OBSERVATIONS
On the basis of the current system of Inventory Management in this Large Mining Organization, following observations are made

6.1. Ordering and Inventory Management System

No System for Determination of Order Quantity Based on Inventory Cost:
The objectives of inventory management are to provide the required level of customer service and to reduce the sum of all costs involved. To achieve these objectives, two basic questions must be answered; one, how much should be ordered at one time? And two, when should an order be placed?
Management must establish decision rules to answer, how much should be ordered at one time and when should an order be placed? Economic Order Quantity (EOQ) basically answers these questions. The advantages of EOQ are many. However, EOQ is not being followed at this organization. Inventory carrying and inventory ordering costs do not form part of the ordering policy which is likely to result in higher working capital requirement and inventory cost.

Ordering System:
More beaureocracy is involved in ordering which results in time delays. The lead-time is very high (for consumable items, 180 days and for spares, 270 days approximately). Material is not made available in time to the consumers. Shop floor and maintenance people seem to feel dissatisfied. No efforts seem to be made to develop the vendors. Vendor development and management are being given considerable importance in modern materials management. Therefore, an organization should move towards developing strong and productive relationship with its vendors.

Poor Control on purchase costs:
Controlling purchase cost will lead to profitability. However, methods adopted to control the purchase cost are not found adequate. Cost reduction by purchasing in lot size is not followed. Consolidation of purchases as well as good vendor relationship can assist in cost reduction and better cost management.
Cost price analysis is not done which is very important especially for items which are procured on single quotation basis. Only few items are covered under rate contract.

Poor Buyer – Seller relations:
The vendors are not happy because their payments are inordinately delayed. In many cases, it was found that the suppliers were not paid even after one to two

years of the receipt of the materials. This has resulted in losing many reputed suppliers. The enquiry was floated to many registered suppliers but the response was very poor. As a result, the company was finding it difficult to have competitive rates and the material was being procured on high prices. After getting the order, the vendors do not supply the material in time due to their apprehensions of not getting payments after supply. In many cases the materials department could not get released the dispatch documents of the suppliers from the banks and the suppliers had to take back their goods unwillingly. This has resulted in poor buyer – seller relations. The purchase department does a lot of correspondence to get the supply in time resulting in waste of man-hours and follow up cost.

6.2. Procurement Cycle

Repeat Order not in place:
There is no system followed to repeat the earlier order in spite of the rates and specifications of the item being same. As the lead-time is very high (for consumable items, 180 days and for spares, 270 days approximately) at this organization, repeat order may prove fruitful.

6.3. Administrative System

Poor Man Power Management:
Baring few persons like AGM, Chief Manager, and Manager, none of the persons working in the Stores and Purchase department have the materials management degree. Many persons from technical departments are placed in Stores and Purchase department who do not have sound knowledge of materials management.

7. SUGGESTIONS:
Following legitimate suggestions are being made to improve the current system

7. Ordering and Inventory Management System
Inventory Management System:
Inventory carrying and inventory ordering costs do not form part of the ordering policy, which is likely to result in higher working capital requirement and inventory cost. Management must therefore, establish decision rules to answer as to how much should be ordered at one time and when should an order be placed?
For achieving this objective, Economic Order Quantity system should be followed so that Inventory carrying and Inventory ordering costs are at minimum to avoid higher working capital requirement.

Ordering System:
To avoid more beaureocracy in ordering which results in time delays leading to idleness of man and machines, ordering system should be simplified. This may include computerization of all purchase functions, which would reduce routine clerical activity by the automatic preparation of documents, e.g. purchase requisition, draft purchase orders (DSOs), Purchase orders, acknowledgement forms and progress letters. Strong and productive relationship with the vendors should be developed. Cash purchases for small value items may be introduced. This will result in shorter delivery time as well as purchases can be made at very short notice.

Control of Purchase costs:
Controlling purchase cost will lead to profitability. Purchasing in lot size will result in Cost reduction. Consolidation of purchases as well as good vendor relationship and Vendor development can assist in cost reduction and better cost management. More items should be covered under rate contract. Cost reduction can also be achieved by extensive negotiation with the vendors.

Vendor Relations:
The organization should develop good relations with mutual interest with its vendors as good vendor relationship can assist in cost reduction and better quality management. Good relationship can go a great way to enhance the good name and prestige of the company. When it is difficult to get a scarce item, a supplier for the sake of his good relations, will undertake to supply it and might put himself to inconvenience if need be. It is therefore, as much necessary to develop goodwill between a company and its suppliers as between it and its customers.
The company should ensure that the payments should be made to the suppliers in time. The practice of making payments for so many supplies of a vendor in one go at a later date (say after 6 months or one year) should be avoided. This will not only remove the burden of large amount payments in one time by the company but also enhance the good buyer–seller relations.

7.2. Procurement Cycle

Repeat Order Policy:
A repeat order policy be introduced with the conditions that if the rates and specifications of the item are same, an order can be repeated within a stipulated period, say, within 6 months or so. As the lead-time is very high at this organization, repeat order may prove fruitful. Purchases through repeat order facilitate immediate purchases for user's

satisfaction. A repeat order would ensure timely purchases, which would help avoid production losses.

7.3. Administrative System

Man Power planning:
A policy of professional management be evolved in order to place right persons on right jobs. Persons without specialized back ground at times are unaware of the right kind of procedures / practices and as a result may add up to procurement cost and litigation in some cases. The function of the Materials management is a specialized one and therefore, it should be carried out by the persons who have the expertise in materials management field. Hence, the materials department must have the persons with specialized background. They should have the degree / diploma in materials management. From the current policy of the generalized pool of manpower, a specialized function based well trained manpower management is needed. Also, a policy of continuous up gradation of skill to keep pace with the changing technology need to be evolved.

CONCLUSION:

In this paper we have studied inventory management system in a Large Mining Organization. Supply chain is very important for this organization to avoid disturbance in the manufacturing process. However, it is observed that current inventory management practices are not able to achieve this objective. As the requirement of the shop floor and maintenance people is not met in time, the production losses are witnessed. Production targets are not achieved. To overcome this problem, it is very important to have a better inventory management system in place. As it may not be possible due to cash flow problem for the organization to change the system in totality, suggestions are made which can be incorporated without any significant investment and are likely to improve the management of the inventory system.

REFERENCES:
George W. Plossl (1985) – Production and Inventory Control: Principles and Techniques – Englewood Cliffs. N.J. USA. P. 17.

Inventory Valuation For Investors June 5, 2002: FIFO And LIFO, | By **Investopedia Staff**, (Investopedia.com) (Taken from Web Page by referring google search)

J. R. Tony Arnold and Stephen N. Chapman (2001): *Introduction to Materials Management,*(Fourth Edition) Prentice Hall International, Inc. New Jersey P. 255.

K S Menon : *Purchasing and Inventory Control* (1979).- Y P Chopra, for AH Wheeler & Company (P) limted, Allahabad -211001 P. 321.

PC Sharma: *Materials Management* – Kitab Mahal, 15, Thorn hill Road, Allahabad (1984) P. 40.

P Gopalakrishnan and M Sundaresan (1994): *'Materials Management' An Integrated Approach*– Prentice –Hall of India Private Limited New Delhi P. 60-66.

AUTHOR'S BIOGRAPHY:
I would like to take this opportunity to introduce my self. My name is Dr. Sharif and I have been working with Indian Institute of Technology Kanpur (IITK) India, a world class Institution in the field of Technical Education. Also, previous to joining IIT Kanpur, I had worked with Hindustan Copper Limited (A Public Sector Undertaking owned by Government of India), as a Senior Materials Officer. I have wide knowledge of Materials Planning, Purchase, Stores and Inventory Management, Stores verification and Accounting, Rate Contract, Customs, Excise, Sales Tax, Negotiations, Drafting and Monitoring of Contracts, Vendor development, Imports, Inventory Control and General Administration. As far as my qualification is concerned, I have done my Masters in Management Science (Materials) from Pune University (India) and Ph. D in 'A Critical Appraisal of Inventory Control Techniques and Practices' from CSJM University Kanpur (India). I have published some good papers in national and international journals. My book on "A Critical Appraisal of Inventory Control Techniques and Practices" is ready for publication. I have attended many workshops and conferences in the past.

MODELLING AGRICULTURAL PRODUCTION SYSTEMS
USING AN ACTION-FLOW-STOCK ONTOLOGY

François Guerrin[a,b]

[a]Cirad, UR 78 Risque environnemental et Recyclage, F-97408 Saint-Denis (Reunion)
[b]Inra, UR 875 Biométrie et Intelligence Artificielle, F-31326 Castanet-Tolosan (France)

[a]guerrin@cirad.fr

ABSTRACT

This contribution depicts an approach enabling agricultural production systems to be represented at various spatial-temporal scales and organization levels: single plot or workshop, whole farm, group of farms within a territory, supply chain. This approach is based on an 'Action-Flow-Stock' ontology according to which each production unit is represented by a stock subject to filling-emptying actions controlled by conditions derived from states or events locally observed on some processes. Whereas stocks are computed as continuous variables, actions are represented by dynamical discrete functions. The models built with this ontology, whatever the formalization framework used to implement them (systems dynamics, multi-agents systems, timed automata), aim at dynamically simulating both the material flows and the human activities within agricultural production systems. The intended use of such models is to help agricultural stakeholders design and assess actual or potential management strategies in the light of the new concerns brought by the sustainability issues.

Keywords: Hybrid dynamical systems; Action representation; Agricultural production systems simulation; Integrated impact assessment.

1. INTRODUCTION

Our approach meets the objectives traced by Thornton and Herrero (2001) about mixed crop-livestock production systems to deal with the challenges brought by the needs to develop food production in the next decades worldwide. According to these authors, a modelling framework is necessary to assess the impacts of the various change factors agricultural systems are increasingly subject to. What they call for is, actually, building simulation models to support agricultural stakeholders design and assess 'attainable' (in practice) management strategies rather than, simply, 'feasible' (in theory) strategies.

With this aim, Thornton and Herrero (2001) established a long list of requirements this modelling framework should fulfil among which its capacity of representing agricultural actual practices is first place. According to these authors, this is the criterion that best defines models' applicability. In effect, the issues the farmers are facing to have more a systemic dimension than a mere technical one. Management is thus considered as the key-factor of agricultural complexity and one of the most challenging research issues (Thornton and Herrero, 2001). But taking into account actual human practices in production systems models, namely in agricultural ones, is precisely the missing point (Garcia et al., 2005).

Therefore, one of the major scientific stakes is methodological. What has to be built is a modelling framework allowing one to:

- represent a production system at different scales and levels of analysis;
- integrate the various pieces of knowledge multiple stakeholders have on this system;
- simulate the dynamics of interactions between human activities and material flows;
- assess the impact of actual practices on the viability and sustainability of the system;
- design and test new management strategies with agricultural stakeholders to improve the overall system's performances according to various criteria.

Some of our recent achievements go clearly in this direction:

- designing flow management models at the sub-farm level (Guerrin, 2001; Courdier et al., 2002; Guerrin, 2004; Hélias et al., 2008);
- integrating biophysical and flow management partial models at the whole-farm level (Vayssières, 2008);
- devising a generic action modelling framework (Guerrin, 2005);
- integrating all this modelling endeavour within a global approach.

Whatever the formalism they rely on (systems dynamics, multi-agents system, timed automata), our models share common ontology and objective: simulating material transfers among a set of production units (PU's) represented as stocks subject to filling or emptying actions so as to answer stock management questions:

- Who? (which PU should transfer its material?)
- Where? (towards which other PU?)

- When? (with respect to which schedule or event?)
- How? (with which vehicle, workforce, resource?)
- How much? (what quantity?).

How agricultural production systems can be represented using the 'Action-Flow-Stock' (A-F-S) ontology common to all of our models, what do these models look like and what they are used for is presented in this article.

2. ACTION-FLOW-STOCK MODELLING

2.1. Production unit

Every agricultural production system may be represented as a set of more or less elementary PU's linked by material flows of different natures. Each PU is a stock with an inflow, an outflow and, possibly, an overflow (Fig. 1).
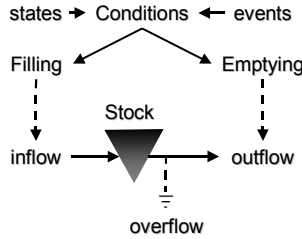
Figure 1: Action-Flow-Stock Representation of an Elementary Production Unit

It is classically represented as a continuous variable by a mass-balance differential equation:

$$\frac{dV(t)}{dt} = Q_{in}(t) - \left[Q_{out}(t) + Q_{over}(t)\right] \qquad (1)$$

Material transfers from one PU to another are realized thanks to filling or emptying actions encoded as binary variables (either they occur or not) subject to time-varying constraints and conditions $C_A(t)$:

$$s_A(t) = \begin{cases} 1 & \text{if} \quad C_A(t) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

where $s_A(t)$ denotes the binary state function of an action A. It thus exhibits a rectangular time evolution in which each pulse, accounting for the realization of action A, is characterized by its start date, end date and duration (Fig. 2).

Encompassing both continuous and discrete variables, the models based on this ontology belong to the family of hybrid dynamical systems (Antsaklis et al., 1998).
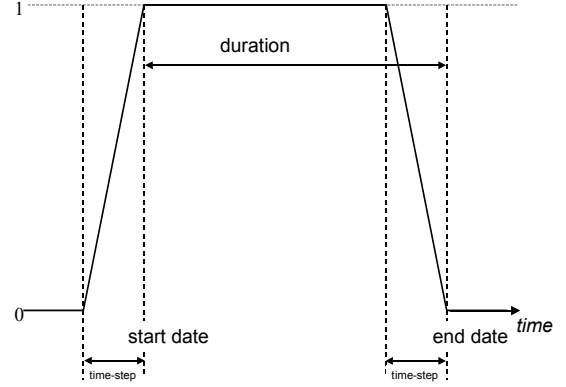
Figure 2: Singular Action Realization

2.2. Action course

As stated by Eq. (2) and sketched in Fig. 2, every action is described as a binary dynamic process subject to a realization condition: its state is 1 whenever $C_A(t)$ is true or 0 when it is false. The action state function defines a partition of the temporal domain between the periods when action is realized and those when it is not. The realization condition of every action is defined by the combination of elementary conditions relative to the states of various processes (clocks, schedules, observations, other actions), continuous or discrete, upon which relevant events for triggering or inhibiting action are detected. These are these events which determine the start and end dates of action realizations by altering the evaluation of their condition. When the condition is no more satisfied while action is ongoing it stops.
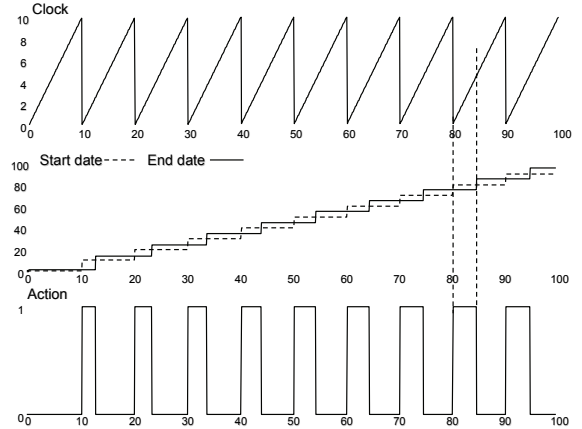
Figure 3: Clock (top), Start and End Dates (middle) and Action Course Evolutions

Figure 3 illustrates how an action course may be simulated using the following concepts:

- Observed process: here a clock, encoded as a piecewise linear function, is taken.
- Events prior to action (EPA): here the time points at which the clock function is zero (clock ticks), though it could be also any state value of any other process.
- Start dates: time points determining each pulse after an EPA, possibly complying with the

conjunction of various constraints such as disturbances (rainfalls,…), resource availabilities (material, workforce, vehicle…), management rules (e.g., delay, postponement).

- End dates: time points derived either from an absolute time-varying duration (like in Fig. 3) or from events detected upon processes (like for the start dates).
- Management rules: specifying the quantities to be transferred, the priorities among actions, logistic constraints (payload, transportation delay,…).
- Action course: encoded as a binary dynamic state function as in Eq. (2).

An action is converted into effects by combining (e.g., by making the product of) its state function with a base flow (e.g., working time available daily, filling rate of a tank). This flow may vary over the action course to represent some progressiveness in the effect or depend upon other processes to represent the modulation of the effect according to the action context.

2.3. Complex activities

Activities are sets of coordinated actions. Figure 4 gives an example where an action A (from the example given in Fig. 3) fills a stock whereas an action B empties it so that the stock level $V(t)$ remains between an upper and a lower threshold.
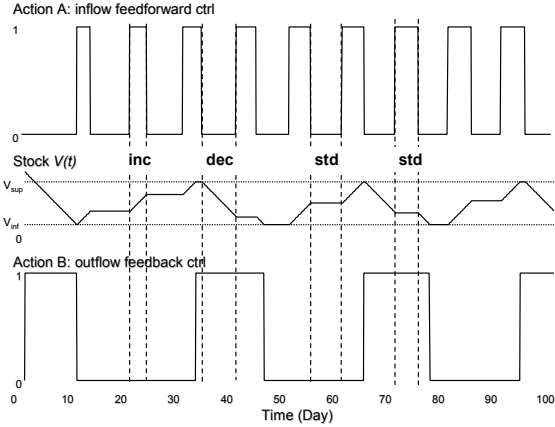
Figure 4: Simulating an Action B Indirectly Derived from an Action A to Control the Stock Level $V(t)$ between Thresholds V_{sup} and V_{inf} (inc: increase; dec: decrease; std: steady)

It is assumed both inflow and outflow are the same and constant. For this reason (see Fig. 4), whenever:

- A is on and B is off, then the stock increases;
- A is off and B is on, then the stock decreases;
- $A = B$, then the stock is steady.

Specifying an action directly as a function of another action so that both satisfy some temporal relation is also possible. This is done by setting the constraints on their start dates and durations derived from the Al-

len's (1984) temporal relation that should hold among them: BEFORE, MEETS, OVERLAPS, DURING, STARTS, FINISHES, EQUAL or their inverses. For example (see Fig. 5), specifying an action E according to an action D so that E is on during D (i.e., DURING(E,D) holds) implies E starts after D and its duration does not exceed that of D. It is thus possible to specify complex activities made of sets of coordinated actions as that of Figure 5.
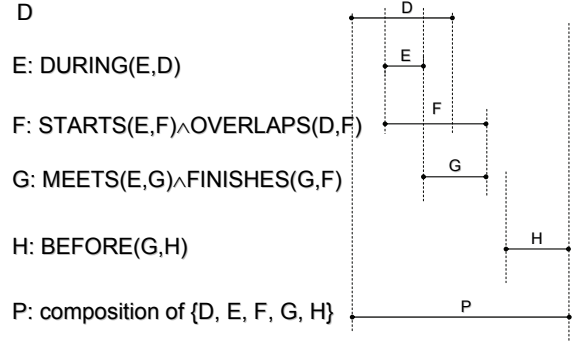
Figure 5: Specification of a Set of Coordinated Actions using Allen's (1984) Temporal Relations

In the case of interruptions (deterministic or random), the realizations of actions can be cancelled, interrupted, resumed, while their total duration is kept by the means of local feedback controllers.

Action courses composing an ongoing activity can be interpreted as external processes by identifying the Allen's relation holding between any two of them. This is done by verifying the temporal order of their start and end dates for each Allen's relation. For example, in Figure 5, the relation DURING(E,D) can be observed if E has started after D, is already achieved, and D is still being pursued. This comes down to evaluate the proposition: $t_D^+ < t_D^- < t_E^- < t_E^+$, where t_i^- and t_i^+ are the start and end dates of action i respectively.

Action realization conditions, action state functions and temporal relations between actions are all represented by dynamic binary functions. These functions may be interpreted as truth values of timed logical propositions: their simulation interval is dynamically partitioned into periods with value 0 (false) or 1 (true). Using the Boolean algebras $\langle \{0,1\}, +, \times \rangle$ or $\langle \{0,1\}, \max, \min \rangle$, it is easy to define operations on action state functions equivalent to the classical logical connectives (NOT, AND, OR, XOR). This allows logical reasoning about action or, more generally, situations they are part of, to be simulated in order to derive new actions and new situations. For example, Figure 6 describes the three possible actions, specified in terms of both actions A and B's state functions, that could be derived to account for the logical equivalences of the three cases enumerated above about the stock evolution in the example given Figure 4:

- stock $V(t)$ increases whenever A and not B;

- stock $V(t)$ decreases whenever not A and B;
- stock $V(t)$ is steady whenever either A and B or not A or B.

$$inc(V) \Leftrightarrow A \wedge \neg B \equiv s_A(1-s_B)$$

$$dec(V) \Leftrightarrow \neg A \wedge B \equiv (1-s_A)s_B$$

$$std(V) \Leftrightarrow (A \wedge B) \vee \neg(A \vee B) \equiv 1-s_A-s_B+2s_As_B$$
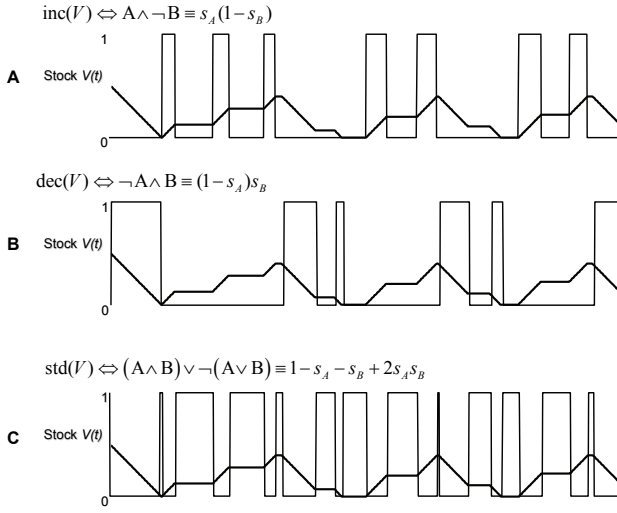
Figure 6: Deriving the Possible Actions According to the Three Logical Combinations of Actions A and B Controlling the Inflow and Outflow of Stock $V(t)$

2.4. Agricultural production system

Actual agricultural production systems may be represented at several organization levels (single field or workshop, whole-farm, group of farms in a territory, agricultural supply chain) as networks of A-F-S units exhibiting various configurations according to the real case to be accounted for (see examples in section 3).

Two types of flows are distinguished according whether they result from man-driven or natural processes: "workable flows", occurring only when man intervenes, and "biophysical flows", occurring even in absence of human intervention. Both types of flows interact by the means of human activity which aims at influencing the biophysical flows leading to the system's products and impacts by the means of the workable flows it generates.

The management of an agricultural production system may thus be seen as controlling a set of stocks by the means of farming activities. These, fundamentally, are exerted within a dynamical environment, evolving independently of any action, making crucial taking account of time irreversibility in management. These activities stem from the confrontation between the actual situations encountered by the farmer and the management strategies he/she has devised to deal with.

By "situation", it is meant the set of constraints of different natures (resource availability, climate or economical conditions,…) exerted at given time and location which restrict the panel of action possibilities.

Among the "resources" necessary to action, information intervenes in the form of events, either endogenous (state transitions) or exogenous (disturbances or climatic, economical, political, social, regulatory constraints), perceived by the farmer directly or indirectly (i.e., by the means of indicators).

By "strategy", it is meant the combination of general principles of acting and making decisions about the workable flows. A strategy reflects, even these are not easy to characterize, the farmer's knowledge, preferences, objectives and management abilities (interpretation, reactivity, decision, anticipation, experience).

Situations and management strategies are interdependent: implementing strategies contributes to create new situations that, in turn, influence the design of new strategies according to the outcome gained from practical implementations.

3. APPLICATIONS

Based on the 'Action-Flow-Stock' ontology described above, we have built in the recent years several simulation models to deal with specific agricultural production systems. All these models illustrate the capabilities of the A-F-S ontology to cope with production unit networks exhibiting various configurations.

3.1. Magma: livestock effluent management at the farm level

The Magma model (Guerrin, 2001) addresses the case of livestock effluent management within a farm. Two types of units are involved in such a "distribution" (i.e., one-to-many) configuration (Fig. 7): livestock enterprises producing animal wastes and consumption units, such as crop plots or waste treatment plants, where effluents may be spread over or supplied.
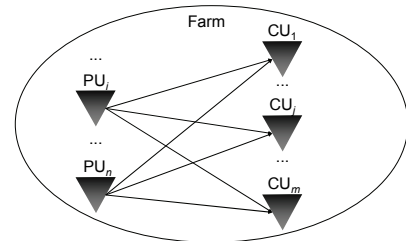
Figure 7: Distribution Configuration of Material Flows at the Farm Level in the Magma simulation Model (PU: livestock enterprise; CU: consumption unit)

Simulating Magma allows management strategies of livestock effluents to be assessed with respect to several indicators: environmental (nitrogen losses due to stock overflowing, fallow land spreading, or crop over fertilization); agronomical (nitrogen applied to crops); economical (working time, vehicle mileages,…) and organizational (frequency and temporal distribution of spreading actions).

3.2. Biomas: livestock effluent management at the territory level

Extending the Magma capabilities over a whole territory encompassing several farms, the Biomas agent-based simulation model (Courdier et al., 2002) addresses the case of livestock effluent transfers between distinct farms (Fig. 8): farms with excess of effluent with respect to the absorption capacity of their own

crops; farms with deficit of organic matter with respect to their own production capacity. Biomas also allows intermediate storage facilities or effluent treatment plants to be added to the network. Biomas puts into play several hundreds of distinct agents (farmers, livestock enterprises, crop plots, vehicles,…) in real-scale applications. These agents, characterized by numeric or symbolic attributes, interact dynamically according to their roles: producer, consumer, organic matter carrier.
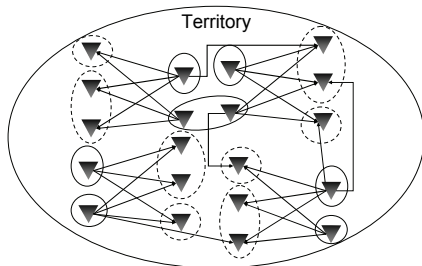
Figure 8: Distribution Configuration of Material Flows in the Biomas Multi-Agents System (full lines: effluent production units in surplus farms; dotted lines: consumption units in deficit farms)

3.3. Approzut: supply of a collective treatment plant by multiple farms

The Approzut model (Guerrin, 2004) deals with the case of simulating the flows in a two-stage production system where the first stage is a set of pig farms producing slurry and the second stage is a collective treatment plant where this slurry is brought in a many-to-one fashion (Fig. 9). The strategy assessment is here mainly done in terms of organization and logistics.
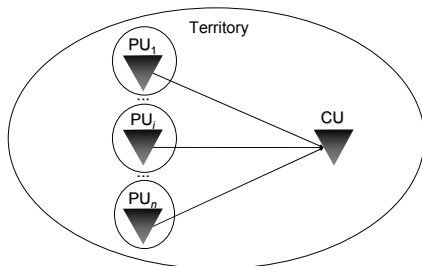
Figure 9: Supply Configuration of Material Flows at the Territory Level in the Approzut Simulation Model (PU: livestock farm; CU: single waste treatment plant)

3.4. MagmAppro: mixed distribution/supply configuration connecting two remote territories

MagmAppro (Lopez-Ridaura et al., 2007) results from coupling together the Magma and Approzut models. It is used to simulate an effluent management collective plan to be implemented in Brittany (Western France) involving dozens of pig farms and crop farmers located about 40kms one from each others (Fig. 10). Coupled with biophysical models, MagmAppro allows namely the impact of such plans to be assessed as nitrogen and methane volatilized at storage and spreading in addition with other indicators as in Magma and Approzut.
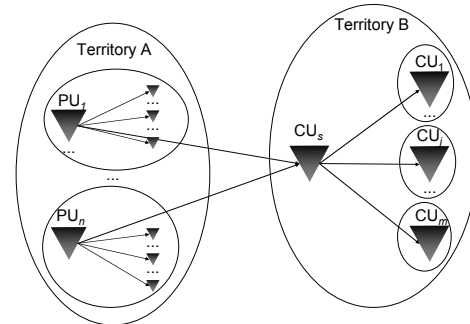
Figure 10: Mixed Distribution/Supply Configuration of Material Flows Connecting Two Remote Territories in the MagmAppro Simulation Model (PU: livestock farms encompassing livestock enterprises and crop plots; CU_s: intermediate stock; CU: crop farms)

3.5. Gamede: global Activity-Flow-Stock model at the whole-farm level

The Gamede model (Vayssières et al., 2007) addresses the case of managing all the material flows (vegetal and animal biomasses and products, nitrogen, water,…) and 19 technical activities (harvesting, fertilising, spreading, feeding, milking…) at the whole-farm level assessed by several sustainability indicators (Fig. 11).
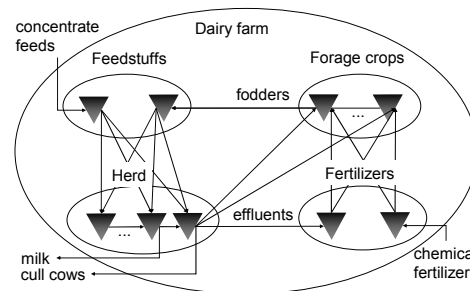
Figure 11: Dairy Farm in the Gamede Simulation Model Viewed as Four Interconnected Sub-systems Encompassing Several A-F-S Production Units

3.6. PigSC: pig industry supply chain in Reunion Is.

Modelling the whole pig industry in the Reunion Island using the A-F-S ontology is being done to represent the pig production chain from the raw feedstuff suppliers to the pig products retailers, passing by the ≈200 rearing enterprises in the whole-Reunion island (Fig. 12).
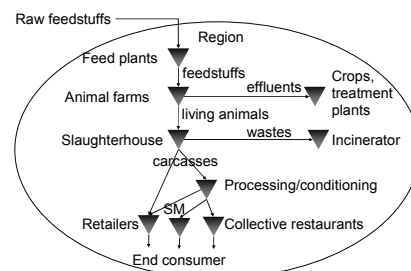
Figure 12: The Pig Industry in the Reunion Island Viewed as a Supply Chain (SM: supermarkets)

11

This model aims at simulating the material flows and coordination among stakeholders. It could be used later to support Life Cycle Analyses of the pig industry.

4. CONCLUSIONS AND FUTURE PROSPECTS

As far as we have experienced it, the A-F-S ontology has proved useful to represent and simulate, at various scales and organization levels, agricultural production systems viewed as sets of interacting biophysical and human-driven processes. In contrast with most decision-support models in agriculture, human activity is explicitly represented. This is deemed crucial as what is wanted is to assess the sustainability of agricultural production systems. In effect, it is by representing at best what is actually done by the farmers, when and how it is done, that one can simulate impacts realistically (e.g., consuming resources, emitting pollutants). In turn, it is possible to assess the reciprocal influence of the context so modified on the course of action itself.

Our tentative action modelling framework (Guerrin, 2005) needs still be generalized to a wider class of production systems and completed with additional features so as to simulate the various means actual actors use to coordinate their activities. Two other issues of major interest for the practical use of our models are still to be further investigated:

- environmental assessment of agricultural production systems at multiple scales and the correlative choice of relevant representational granularities of the model(s) to be used (Lopez-Ridaura et al., 2007);
- participatory use of such models in virtual experiments with agricultural stakeholders involved in actual decision processes (Médoc et al., 2004).

REFERENCES

Antsaklis, P., Koutsoukos, X., Zaytoon, J., 1998. On hybrid control of complex systems: a survey. *APII-JESA*, 32(9-10), 1023-1045.

Allen, J., 1984. Towards a general theory of action and time. *Artificial Intelligence*, 23, 123-154.

Courdier, R., Guerrin, F., Andriamasinoro, F.-H., Paillat, J.-M., 2002. Agent-based simulation of complex systems: Application to collective management of animal wastes. *Journal of Artificial Societies and Social Simulation*, 5(3).

Garcia, F., Guerrin, F., Martin-Clouaire, R., Rellier, J.-P., 2005. The human side of agricultural production management: The missing focus in simulation approaches. *Proceedings of Modsim 2005, International Congress on Modelling and Simulation*, Melbourne, Australia.

Guerrin, F., 2001. Magma: A model to help manage animal wastes at the farm level. *Computers & Electronics in Agriculture*, 33(1), 35-54.

Guerrin, F., 2004. Simulation of stock control policies in a two-stage production system. Application to pig slurry management involving multiple farms.

Computers & Electronics in Agriculture, 45(1-3), 27-50.

Guerrin, F., 2005. Simulation of action in production systems. *Proceedings of Modsim 2005, International Congress on Modelling and Simulation, Advances and application for management and decision-making*, Melbourne, Australia.

Hélias, A., Guerrin, F., Steyer, J.-P., 2008. Using timed automata and model-checking to simulate material flows in agricultural production systems-Application to animal waste management. *Computers & Electronics in Agriculture*, (in press).

Lopez-Ridaura, S., Guerrin, F., Paillat, J.-M., van der Werf, H., Morvan, T., 2007. Agronomic and environmental evaluation of collective manure management for a group of farms. *Proceedings of Farming Systems Design 2007, International Symposium Methodologies for integrated analysis of farm production systems*, Catania, Sicily.

Médoc, J.-M., Guerrin, F., Courdier, R., Paillat, J.-M., 2004. A multi-modelling approach to help agricultural stakeholders design animal wastes management strategies in the Reunion Island. *Proceedings of iEMSs 2004, International Congress Complexity and Integrated Resources Management, International Environmental Modelling and Software Society*, Osnabrück, Germany.

Thornton, P., Herrero, M., 2001. Integrated crop-livestock simulation models for scenario analysis and impact assessment. *Agricultural Systems*, 70, 581-602.

Vayssières, J., Lecomte, P., Guerrin, F., Nidumolu, U.-B., 2007. Modelling farmers' action: decision rules capture methodology and formalisation structure. A case of biomass flow operations in dairy farms of a tropical island. *Animal*, 1, 716-733.

AUTHOR BIOGRAPHY

François Guerrin is doing research in modelling and simulation-aided management of agricultural production systems. He is part of a team devoted to environmental risk management and wastes recycling at Cirad (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) in Saint-Denis (Reunion Island). Prior to his current position (1987-1997) he developed Qualitative Reasoning models with application to environmental systems at Inra (Institut National de la Recherche Agronomique) in Toulouse (F). From 1980 to 1996 he worked successively, as a freshwater ecologist, at Agence de l'Eau Adour-Garonne (Toulouse, F), Cemagref (Institut de recherche pour l'ingénierie de l'agriculture et de l'environnement) in Montpellier (F) and Inra (French Guyana). Within these institutions, he was working on wastewater reclamation and reuse for aquaculture. **F. Guerrin** holds a Ph.D. in hydrobiology from Paul Sabatier University (Toulouse, F) and an "Habilitation à diriger des recherches" (HDR) from Reunion Island University. He is an editorial board member of the Ecological Modelling journal.

STOCHASTIC LEADTIMES IN A ONE-WAREHOUSE, N-RETAILER INVENTORY SYSTEM WITH THE WAREHOUSE CARRYING STOCK

Adriano O. Solis[a], Charles P. Schmidt[b]

[a] Management Science Area, School of Administrative Studies, York University, Toronto, Ontario M3J 1P3, Canada
[b] Department of Information Systems, Statistics and Management Science, The University of Alabama, Tuscaloosa, Alabama 35487-0226 U.S.A.

[a] asolis@yorku.ca, [b] cschmidt@cba.ua.edu

ABSTRACT
This study involves the introduction of stochastic leadtimes between the warehouse and retail sites, in place of the original deterministic leadtimes in an earlier reported multi-echelon inventory model. We had previously investigated analytically the effects of stochastic leadtimes on required retailer base stock levels when the warehouse does *not* carry stock (e.g., serves as a cross-dock point). The current paper examines the effects of stochastic leadtimes in the traditional situation where the warehouse *does* carry stock. The model becomes mathematically intractable, and simulation studies become appropriate. A heuristic involving simulation experiments is devised for selecting a base stock policy, taking off from solutions to the deterministic leadtime model. Resulting average system inventory is found equal or very close to the optimal level. Our study suggests that the optimal solution to the deterministic leadtime case provides an appropriate starting point in searching for a solution in the stochastic leadtime case.

Keywords: inventory management, multi-echelon inventory system, stochastic leadtimes, simulation

1. INTRODUCTION

Graves (1996) developed a multi-echelon inventory model assuming what he refers to as *virtual allocation*. Each of the sites—a central warehouse (CW), transshipment sites, and retail sites—follows a base stock (or "order-up-to") policy. In the two-echelon case involving only a warehouse and N retailers, Graves specifies search procedures for the optimal combination of base stock levels at the CW and the retail sites minimizing the average system on-hand inventory under each of two service level criteria: a probability of no stockout criterion and a fill rate criterion.

The average system inventory under an optimal policy where the CW base stock level is positive (i.e., the CW carries stock) is expected to be lower than under a policy where the CW does not carry stock. This is in view of "statistical economies of scale" made possible by carrying, at least in part, a single-system inventory instead of simply N individual inventories—which Eppen and Schrage (1981) refer to as the *depot effect*. Based on a limited computational study, Graves makes the assessment that the depot effect in his model "seems fairly small except when there are many retailers." Solis (1998) expanded Graves' computational study, and found the depot effect to be felt more the larger the number of retail sites or the smaller the retail site demand rate.

Solis and Schmidt (2007) earlier introduced into Graves' model stochastic leadtimes between the CW and the retail sites in place of the original deterministic leadtimes, but dealing only with the case where the CW does *not* carry stock (e.g., serves as a cross-dock or transshipment point). Effects of stochastic leadtimes on required base stock levels at the retail sites were investigated analytically, taking two alternative treatments of stochastic leadtime distributions into consideration. Results of that earlier study suggest that it may be better to use the deterministic leadtime model with an accurately estimated mean leadtime than a stochastic model with a poorly estimated mean leadtime.

In the current study, the CW, in the traditional sense, *does* carry stock. The mathematics becomes intractable, and simulation studies are undertaken.

2. GRAVES' MODEL

The model involves an arborescent system with M inventory sites, i = 1, 2, ..., M. Each site j has a single internal supplier i = $\rho(j)$, with the exception of site 1, a central warehouse (CW) whose inventory is replenished by an external supplier. Customer demand occurs only at retail sites, which have no successor nodes. All other sites are storage and/or consolidation facilities, called transshipment sites. The unique path linking a retail site to the CW is the *supply chain* for the retail site.

The analysis involves a single item of inventory. The demand at each retail site j is an independent Poisson process with demand rate λ_j. $D_j(s,t)$ represents the demand over the time interval (s,t] for site j. The induced demand rate for any site other than the retail sites is the sum of demand rates at its immediate successors. For transshipment sites as well as the CW, $D_i(s,t)$ is the sum of the demands over the interval (s,t]

at the sites supplied by site i (i.e., at the immediate successors of site i).

A multi-item distribution system is contemplated, with regularly scheduled shipments between sites. Each shipment is a consolidation of orders for various items in inventory, including the item under study. A schedule of preset times $p_j(m)$, $m = 1, 2, ...$, at which site j places its mth replenishment order on its supplier, is followed. Positive leadtimes τ_j for shipments to each site j from its supplier are preset and known. The mth shipment to site j is received at time $r_j(m) = p_j(m) + \tau_j > p_j(m)$. With fixed leadtimes, no order crossing will occur. Hence, $p_j(m) < p_j(m+1)$ and $r_j(m) < r_j(m+1)$. When inventory is in short supply, the supplier will ship less than the quantity ordered and make up for the shortfall on later shipments.

Each site j follows a base stock (or "order-up-to") policy. Initial inventory (at time 0) at site j is B_j, which is the base stock level for site j. Since customer demand is assumed to be fully backordered, site j places an order equal to $D_j[p_j(m-1), p_j(m)]$ at time $p_j(m)$.

$T_j(m)$ represents the *coverage* provided by the supplier to site j, on the occasion of the mth order by site j. A quantity equal to $D_j[T_j(m-1), T_j(m)]$ is shipped by the supplier at time $p_j(m)$, which will later arrive at site j at time $r_j(m)$. If $T_j(m) < p_j(m)$, then the quantity $D_j[T_j(m), p_j(m)]$ remains on backorder. If $T_j(m) = p_j(m)$, then the supplier is able to fill the entire m^{th} order placed by site j. The external supplier is fully reliable and fills every order by the CW (site 1) exactly as scheduled, following a leadtime τ_1. In this case, $T_1(m) = p_1(m)$.

Graves' model differs from other existing models with its assumption of *virtual allocation*. Each site on the supply chain increases its next order quantity by one whenever a unit demand occurs at the retail site. At the same time, each site on the supply chain commits/reserves one unit of its inventory, if available, for shipment to the downstream site on the latter's next order occasion. Priority in the allocation of uncommitted stock is set according to the earliest demand occurrence. While implementing virtual allocation under current information technology is possible, it is not the common practice. It is assumed in the model because it proves to be tractable. It is found by Graves to be near-optimal in many cases.

A random variable requiring attention in this study is $A_j(t)$, which denotes the *available inventory at site j at time t*—the on-hand inventory at site j at time t that has not yet been committed for shipment to another site. $A_j(t) < 0$ indicates outstanding orders (or backorders) at site j at time t. The beginning inventory at site j at time $t = 0$ is the base stock level B_j. Hence, we have $A_j(0) = B_j$. This leads Graves to the following important result:

Theorem If $r_j(m) \leq t < r_j(m+1)$, *then*

$$A_j(t) = B_j - D_j[T_j(m), t].   \tag{1}$$

Suppose that site i is the internal supplier to site j. At time $p_j(m)$, site j places its mth order with site i for a quantity equal to $D_j[p_j(m-1), p_j(m)]$. Since $T_j(m)$ represents the coverage provided by the supplier on the occasion of the mth order by site j, then site i ships to site j at time $p_j(m)$ a quantity equal to $D_j[T_j(m-1), T_j(m)]$, which will arrive at site j at time $r_j(m)$. If site i has sufficient stock and is able to fill the entire m^{th} order placed by site j, then $T_j(m) = p_j(m)$. Otherwise, $T_j(m) < p_j(m)$. In the latter case, $T_j(m)$ equals the time at which site i would run out of available inventory to allocate to site j (i.e., the time when a demand occurrence at some site k, which may or may not be site j, supplied by site i reduces available inventory at site i to zero).

Consider the *relevant* shipment to the supplier i, for which $r_i(n) \leq p_j(m) < r_i(n+1)$. At time $p_j(m)$, site i has received its nth shipment, but not yet its $(n+1)^{th}$ shipment. Define $S_i(n)$ to be the *depletion* or *runout time* for this relevant (n^{th}) shipment to site i. That is, based upon its receipt of the nth shipment, site i is able to cover (or replenish) the demand processes of its successor sites up through the runout time $S_i(n)$. If $S_i(n)$ occurs after $p_j(m)$, then $T_j(m) = p_j(m)$. However, if $S_i(n)$ occurs before $p_j(m)$, then $T_j(m) = S_i(n)$. Thus,

$$T_j(m) = \min \{p_j(m), S_i(n)\}.   \tag{2}$$

The base stock B_i takes the role of a buffer for demand at site i after the coverage time $T_i(n)$ provided by the supplier to site i. This buffer B_i is depleted at the runout time $S_i(n)$. Graves refers to the difference $S_i(n) - T_i(n)$ as the *buffer time* provided by B_i, and establishes that

$$S_i(n) - T_i(n) \sim \text{gamma}(\lambda_i, B_i).   \tag{3}$$

Graves then focuses on a two-echelon system consisting of sites 1 (the CW) and j (the retail sites), reducing the supply chain of interest to just two sites, 1 and j. A *single-cycle* ordering policy is in place: each retail site j orders a fixed number of times for every order placed by the CW. Hence, if θ_1 denotes the length of the CW order cycle and θ_j that of the retail site, then the ratio θ_1/θ_j is a positive integer. The ordering policy is assumed to be *nested*: every time the CW receives a shipment, all retail sites place an order.

Consider an arbitrary (n^{th}) CW order cycle. Graves simplifies the analysis by setting time zero equal to $p_1(n)$, the time at which the nth CW order is placed with the external supplier. Assuming the CW receives delivery of this order at the end of leadtime τ_1, the retail site orders at time τ_1 (which signals the start of the full CW order cycle). Graves draws attention to the last retail site order—i.e., the $(\theta_1/\theta_j)^{th}$ order—within the CW order cycle. The last order within this CW order cycle is placed by retail site j at time $p_j = \tau_1 + \theta_1 - \theta_j$, and is received by the retail site at time $p_j + \tau_j$, where τ_j is the leadtime between the CW and retail site j. The resulting available inventory will be used to cover demand until

the next order (placed at time $p_j + \theta_j$) arrives at the retail site (at time $t_r = p_j + \theta_j + \tau_j$, which is equal to $\tau_1 + \theta_1 + \tau_j$). The instant of time t_r^- *before* this replenishment at $t_r = \tau_1 + \theta_1 + \tau_j$ proves crucial to the analysis. [In this case, the antecedent of the previous Theorem, i.e., $r_j(m) \leq t_r^- < r_j(m+1)$, holds.]

Treating this $(\theta_1/\theta_j)^{th}$ order as the mth order for the retail site j within the nth CW order cycle, the indices m and n are henceforth dropped for notational convenience. Rewriting (2), the coverage provided by the mth shipment to retail site j is given by $T_j = \min\{p_j, S_1\}$, where p_j = the time at which the (m^{th}) order was placed by retail site j, and S_1 is the runout time for the relevant (n^{th}) shipment to the CW. Having set time zero equal to p_1 for convenience, and assuming the external supplier to the CW to be reliable (so that $T_1 = p_1$), we have $S_1 - T_1 = S_1$. Hence, in this case, the runout time S_1 is equal to the buffer time $S_1 - T_1$ provided by the base stock B_1 at the CW. It follows from (3) that $S_1 \sim \text{gamma}(\lambda_1, B_1)$.

Graves establishes the first two moments of T_j, and uses these to specify the mean and variance of the random variable $D_j[T_j, t]$, where t is some specified point in time. Of particular interest is the distribution of $D_j[T_j, t_r]$, where $t_r = \tau_1 + \theta_1 + \tau_j$ is the critical point in time discussed above. Graves refers to $D_j[T_j, t_r]$ as *uncovered demand* (up to time t_r): demand at retail site j not covered by the $(\theta_1/\theta_j)^{th}$ shipment from the CW.

Graves reports to have computationally found the negative binomial distribution, having the same first two moments as $D_j[T_j, t]$, to provide a fairly accurate approximation to the distribution of $D_j[T_j, t]$, while presenting little evidence in support of his assertion. He cites two earlier multi-echelon inventory studies (Graves 1985; Lee and Moinzadeh 1987) in which a negative binomial approximation had also been found effective. Solis, Schmidt, and Conerly (2007) provided a mathematical analysis of the effectiveness of the approximation in the current model, where the maximum absolute deviation between cumulative probabilities of the approximate and exact distributions reach about 0.01 in one of 64 test cases, but much less in most others.

Based upon Graves' specification that the negative binomial distribution used to approximate the distribution of $D_j[T_j, t]$ has the same first two moments as the latter, the discrete density function (d.d.f.) of this approximate distribution may be characterized by $f(x; r, p) = {}_{r+x-1}C_x \, p^r (1-p)^x I_{\{0,1,2,\ldots\}}(x)$, with parameters

$$r = (t - E[T_j])^2 / \text{Var}[T_j] \tag{4}$$

and

$$p = (t - E[T_j]) \, / \, \{(t - E[T_j]) + \lambda_j \text{Var}[T_j]\}. \tag{5}$$

When $\text{Var}[T_j] \to 0$, however, the distribution of $D_j[T_j, t]$ approaches a Poisson distribution with parameter $\lambda_j(t - E[T_j])$.

2.1. Probability of No Stockout as Service Criterion
Each time the retail site places an order with the CW, there is a risk of stockout. This risk exists especially towards the end of each retail order cycle, just before the next shipment to the retail site is received.

Graves notes that, for a nested, single-cycle ordering policy, the probability of the retail site stocking out is greatest for the last order—that is, the $(\theta_1/\theta_j)^{th}$ order—within the CW order cycle. Thus, to set the base stock levels to achieve a given probability α of the retail site *not* stocking out within the CW order cycle, it suffices to consider the probability of stockout at the $(\theta_1/\theta_j)^{th}$ order occasion. Based on the earlier discussion,

$$\Pr\{ A_j(t_r^-) \geq 0 \} \geq \alpha, \tag{6}$$

where $t_r = \tau_1 + \theta_1 + \tau_j$, will need to be assured.

This leads to a computational procedure that searches over possible settings of the base stock level B_1 at the CW. For each B_1, the minimum base stock level B_j at the retail sites that would yield (6) is to be determined. Following (1), the requirement (6) translates into

$$\Pr\{ D_j[T_j, t_r] \leq B_j \} \geq \alpha. \tag{7}$$

The negative binomial distribution, negbin(r,p), with parameters r and p as specified by (4) and (5), is used as approximation to the distribution of $D_j[T_j, t_r]$. The probability $\Pr\{D_j[T_j, t_r] \leq B_j\}$ is approximated by

$$\sum_{x=0}^{B_j} {}_{r+x-1}C_x \, p^r (1-p)^x,$$ starting with $B_j = 1$, and

incrementing B_j by 1 until (7) is satisfied. [Graves did not specify at what level of $B_1 > 0$ the entire search could be allowed to terminate. Solis (1997) established such a stopping condition.]

The base stock level B_1 which yields the lowest average system inventory is finally selected. (In the case of ties, the smallest value of B_1 is preferred—there being no difference assumed in Graves' model between holding costs at the CW and at the retail sites.)

Graves provides the following approximation to expected system on-hand inventory:

$$\text{average inventory} = B_1 + \sum_{1}^{N} B_j - 0.5\lambda_1\theta_1 - \lambda_1\tau_1. \tag{8}$$

He clarifies that (8) should actually be corrected for counting retail backorders at negative inventory—by adding back the time-weighted backorders at the retail sites. However, he points out that, for reasonable service levels, the expected backorder component is very small and insensitive to the inventory policy. Accordingly, this expected backorder component is ignored in using (8) to calculate average system on-hand inventory.

2.2. Fill Rate as Service Criterion

In the case of a fill rate criterion, Graves notes that:

- For a single-cycle ordering policy, the number of backorders over the CW order cycle is equal to the sum of the backorders at the end of each of the θ_1/θ_j retail order cycles, just before receipt of the next shipment to the retail site. (Graves once again points out that this is not completely accurate as it double counts any backorders that may persist for more than one retail cycle.)

- For a nested, single-cycle ordering policy, the expected backorders will be greatest for the last retail order—that is, the $(\theta_1/\theta_j)^{th}$ order—within the CW order cycle.

- Expected backorders over a CW order cycle may be approximated by expected backorders pertaining to the $(\theta_1/\theta_j)^{th}$ retail order within the CW order cycle, since for "realistic" fill rates (> 0.95) effectively all of the backorders occur at this last retail order within the CW order cycle.

We recall that the last order within the CW order cycle is placed by the retail site at time $p_j = \tau_1 + \theta_1 - \theta_j$. This last retail order within the CW order cycle (which arrives at the retail site at time $p_j + \tau_j = \tau_1 + \theta_1 - \theta_j + \tau_j$) may not last until the next order arrives at the retail site (at time $t_r = \tau_1 + \theta_1 + \tau_j$). Backorders would occur if the base stock B_j at the retail site is inadequate.

As already noted, for "realistic" fill rates, we approximate the expected backorders over the CW order cycle by evaluating expected backorders at $t_r = \tau_1 + \theta_1 + \tau_j$.

$E[\{A_j(t_r)\}^-]$, where the symbol y^- stands for $\max\{0,-y\}$, represents expected backorders at time $t_r = \tau_1 + \theta_1 + \tau_j$ (just before the next order arrives). A computational procedure similar to that for the probability of no stockout service criterion arises. For each B_1, we search for the minimum base stock level B_j at the retail sites that would yield

$$E[\{A_j(t_r)\}^-] = \left\{ \sum_{x=0}^{B_j} [(B_j - x) f(x)] \right\}$$
$$- \left\{ B_j - \lambda_j (t_r - E[T_j]) \right\} \leq (1 - \beta) \lambda_j \theta_1, \quad (9)$$

where $\lambda_j \theta_1$ represents mean demand at the retail site over the CW order cycle. The CW base stock level B_1 which yields the lowest average system inventory is chosen.

2.3. Graves' Computational Study

In his computational study, Graves used test scenarios all based on a single system demand rate $\lambda_1 = 36$. Identical retail sites are assumed, with the number N of retail sites being 2, 3, 6, or 18. Hence, the retail site demand rates λ_j are 18, 12, 6, or 2, respectively. The length of the retail site order cycle is fixed at $\theta_j = 1$. Four different parameter combinations $< \theta_1, \tau_1, \tau_j >$, involving the length of the CW order cycle (θ_1), the leadtime (τ_1) from the external supplier to the CW, and the leadtime (τ_j) from the CW to the retail site, are tested. This resulted in *16 test scenarios*, summarized in Table 1.

Table 1: Summary of Graves' Test Scenarios

Scenario	Length of CW Order Cycle θ_1	External Supplier Leadtime τ_1	CW to Retailer Leadtime τ_j	No. of Retailer Sites N	Retail Site Demand Rate λ_j
1	2	1	1	18	2
2				6	6
3				3	12
4				2	18
5	2	1	5	18	2
6				6	6
7				3	12
8				2	18
9	5	4	1	18	2
10				6	6
11				3	12
12				2	18
13	5	4	5	18	2
14				6	6
15				3	12
16				2	18

For the probability of no stockout service criterion, four different levels of α were used: 0.80, 0.90, 0.95, and 0.975. Similarly, four different fill rate levels β were tested: 0.95, 0.98, 0.99, and 0.999. Thus, for each service criterion, a total of 64 test cases were utilized by Graves.

3. STOCHASTIC LEADTIMES

Graves (1996) suggested a number of possible extensions of his model, including stochastic leadtimes τ_j from the CW to the retail sites. The current study investigates effects of stochastic leadtimes τ_j from the CW to the retail sites when the CW carries stock ($B_1 > 0$).

The situation where τ_j is no longer preset and known is a departure from the original deterministic leadtime τ_j in Graves' model, as in many others in the literature on multi-echelon inventory systems—and, particularly more so, from the zero leadtime assumed in some other models (McGavin, Schwarz, and Ward 1993; Nahmias and Smith 1994). A zero leadtime may be plausible where regularly scheduled deliveries from the CW are made overnight while the retail sites are closed. Graves (1996) likewise assumes regularly scheduled shipments. He offers the motivation that, in a multi-item distribution system where each item occupies only a portion of a truckload, a fixed replenishment schedule allows consolidation of item shipments and, accordingly, transportation economies. This scenario may appear to allow a fixed, common positive leadtime for shipments between the CW and the retail site—where one truck services one retail site within the latter's order cycle.

While the order occasion $p_j(m)$ remains fixed, the time $r_j(m)$ at which the order is received by site j may vary. Hence, the leadtime $\tau_j = r_j - p_j$ may be treated as a random variable that varies according to a number of possible factors, for example, truck, road, and weather conditions, or even loading times for the consolidated, multi-item shipments. The realized time of receipt r_j and, as a result, the leadtime τ_j do not depend on the demand process or the quantity ordered of the specific item under consideration. For Graves' original results to continue to apply in the stochastic leadtime case, the condition that orders do not cross—that is, $r_j(m-1) \leq r_j(m)$ for any m—will need to hold.

The treatment here of τ_j is based on $\theta_j = 1$ (i.e., a retail site order cycle length equal to one time unit), as used in all 16 test scenarios in Graves' computational study, but may be extended to any possible value of θ_j. We wish to ensure that the distribution of leadtimes τ_j satisfies the requirement that there is "no order crossing"—specifically, that $r_j(m) < r_j(m+1)$ for any m. This requires the range of leadtimes (maximum leadtime less minimum leadtime) to be at most θ_j. In our treatment, therefore, we require this range of leadtimes to be at most 1. Such a requirement is less restrictive (with respect to the variability of leadtimes) when the mean leadtime is small relative to θ_j, and becomes more restrictive otherwise.

In his computational study, Graves used the fixed leadtimes $\tau_j = 1$ and $\tau_j = 5$. Corresponding to Graves' deterministic leadtime $\tau_j = 1$, we may, for instance, consider leadtimes that vary within the interval between 0.5 and 1.5, or between 0.25 and 1.25, or between 0.9 and 1.9, among other such possibilities for which the range is 1 (equal to θ_j) and $\tau_j = 1$ is included between the minimum and maximum leadtimes. [The range may as well be 0.5 or 0.9, or some other value $< \theta_j$. A range of 1 is, in the case where $\theta_j = 1$, the largest that would be allowed under the no order crossing requirement.]

The lower and upper limits of this interval of leadtimes would depend upon what may actually provide a good approximation to the practice/experience with respect to, among other factors, the consolidation of shipments, the dispatching of delivery vehicles from the CW, and the return of the vehicles to the CW. Since the shipments are made from the CW to the retail sites, an assumption that such shipments are made under a fair degree of control would appear to be a reasonable one to make, in view of the operational requirements of period-to-period consolidated, multi-item shipments—even as the actual leadtimes are subject to some amount of variation due to factors such as loading/ unloading times, weather and road conditions, and unexpected vehicular problems. Hence, we may reasonably expect that leadtime intervals will not be dispersed too widely and, as a result, that orders will not cross.

In our analysis, we shall use a leadtime interval of $(0.5, 1.5)$ arbitrarily, in place of the deterministic leadtime $\tau_j = 1$. The analysis will essentially proceed in the same manner regardless of the final interval that actually applies. Similarly, the interval $(4.5, 5.5)$ is used in place of the deterministic leadtime $\tau_j = 5$. Likewise, regardless of the original deterministic leadtime (be it $\tau_j = 1$ or $\tau_j = 5$), the analysis will proceed in the same manner. Within this interval of $(0.5, 1.5)$ or $(4.5, 5.5)$, we shall look into various stochastic leadtime distributions and their effects on required base stock levels.

Corresponding to Graves' fixed leadtime $\tau_j = 1$, we consider stochastic leadtimes such that $\tau_j - 0.5 \sim$ beta(a, b), where the p.d.f. of τ_j is specified by

$$g(\tau_j) = [1/B(a,b)] (\tau_j - 0.5)^{a-1} (1.5 - \tau_j)^{b-1} I_{(0.5, 1.5)}(\tau_j), \quad (10)$$

with the so-called beta function $B(a,b) = \Gamma(a) \times \Gamma(b) / \Gamma(a+b)$. A random variable with a beta(a, b) distribution varies between 0 and 1, and has a mean of $a/(a+b)$. This distribution is skewed to the right or left depending upon whether $a/(a+b) < 0.5$ or $a/(a+b) > 0.5$. The mean of τ_j, when $\tau_j - 0.5 \sim$ beta(a,b), is equal to $a/(a+b) + 0.5$ and can be anywhere on the interval $(0.5, 1.5)$. In this case, the distribution of leadtimes τ_j would be skewed to the right or left when the mean of τ_j is less than 1 or greater than 1, respectively. [When $a = b = 1$, the distribution of τ_j reduces to a uniform distribution over the interval $(0.5, 1.5)$.]

3.1. The Case Where the CW Actually Carries Stock

When the CW is allowed to carry some of the system inventory, the average system inventory associated with the optimal combination $< B_1, B_j >$ is smaller than the required average system inventory when $B_1 = 0$. Eppen and Schrage (1981) refer to this benefit of the CW carrying stock as the *depot effect*.

Solis and Schmidt (2007) earlier investigated analytically the effects of stochastic leadtime distributions as specified above for the case where the CW does *not* carry stock ($B_1 = 0$). When the CW *does* carry some of the system inventory ($B_1 > 0$), the model does not readily lend itself to the same kind of mathematical analysis.

Accordingly, we conducted simulation studies in order to evaluate the effects of stochastic leadtimes τ_j on the optimal combination $< B_1, B_j >$ that minimizes average system inventory. Baseline simulation models were developed in SIMAN V using the Arena modeling system (see Pegden, Shannon, and Sadowski 1995; Kelton, Sadowski, and Sturrock 2007), for both deterministic and stochastic leadtimes. Preliminary runs indicated that service levels under given $< B_1, B_j >$ combinations may be significantly affected when deterministic leadtimes between the CW and the retail sites (as in Graves' model) are replaced with leadtimes following uniform and various beta distributions.

The simulation experiments were envisioned to study how optimal levels of B_1 and B_j, determined for the fixed leadtime case, would need to be adjusted to compensate for stochastic leadtimes—such that the specified service level is met under either the

probability of no stockout criterion or the fill rate criterion.

For each of the test scenarios, experimental and model frames were developed for both the deterministic leadtime and stochastic leadtime cases. Most of the experiments conducted consisted of 100 replications, with each replication involving 100 CW order cycles. Hence, a typical experiment covered 10,000 CW order cycles.

The Seeds element in SIMAN (Pegden, Shannon and Sadowski 1995) was used to specify separate random number streams for the various sources of randomness—the independent Poisson demand processes at the retail sites, and the stochastic leadtimes τ_j. Thus, experiments with the same test case become comparable, with random demand occurrences and random leadtimes repeating themselves with every run of the experiment, as values of the parameters B_1 and B_j are changed.

Confidence intervals for system service levels (average probability of no stockout and system fill rate)—as reported by Arena's statistical analysis module, which rounds reported figures off to the nearest 0.1—vary in width, and are generally tighter with higher service levels and more retail sites.

3.2. Heuristic for Finding a Policy $< B_1, B_j >$ that Minimizes Average System Inventory

We have devised a heuristic for finding a policy $< B_1, B_j >$ which yields an average system on-hand inventory that appears to closely approximate the smallest possible level of inventory such that the specified service level is satisfied.

From (8), an approximation to expected system inventory when there are N identical retail sites is given by average system inventory $= B_1 + NB_j - 0.5\lambda_1\theta_1 - \lambda_1\tau_1$, which varies only with B_1 and B_j. Hence, to find the minimum average system inventory, we may simply search for the minimum level of echelon base stock $= B_1 + NB_j$.

Let us first introduce additional notation: $e\{< B_1, B_j >\} = $ echelon base stock at $< B_1, B_j > = B_1 + NB_j$; $s\{< B_1, B_j >\} = $ mean service level observed from a simulation experiment involving policy $< B_1, B_j >$; and $< B_1^d, B_j^d > = $ optimal policy for the deterministic leadtime case, found using Graves' search procedure.

Let $< B_1^h, B_j^h >$ represent the current policy as the heuristic searches for an approximation to the optimal policy $< B_1^s, B_j^s >$ for the given stochastic leadtime case—i.e., a policy that minimizes echelon base stock (and, hence, average system inventory). We further let $\delta = B_j^s(0) - B_j^d(0)$—where $B_j^d(0)$ and $B_j^s(0)$ denote the computed optimal retailer base stock levels, if the CW does not carry stock, for the deterministic and stochastic leadtime cases, respectively. The values of $B_j^s(0)$ were earlier determined by Solis and Schmidt (2007) for the same test cases we consider in the current study.

We specify our heuristic as follows:

Heuristic

Step 1. Initialize: $< B_1^h, B_j^h > \leftarrow < B_1^d, B_j^d + \delta >$. If $s\{< B_1^h, B_j^h >\} > \alpha$ or β, then go to step 3.

Step 2. Adjust, via simulation experiments, B_j^h upward until $< B_1^h, B_j^h >$ first satisfies $s\{< B_1^h, B_j^h >\} > \alpha$ or β.

Step 3. Adjust B_1^h downward, via simulation experiments (using, if necessary, some bisection method), until such point that any further reduction in B_1^h will result in $s\{< B_1^h, B_j^h >\} < \alpha$ or β. Select the policy $< B_1^h, B_j^h >$ corresponding to the smallest B_1^h considered such that $s\{< B_1^h, B_j^h >\} > \alpha$ or β.

We express the condition that the specified service level be satisfied in terms of $s\{< B_1^h, B_j^h >\} > \alpha$ or β—instead of using the more familiar $\geq$ inequality—in light of the simulation experiment. This is in recognition of $\Pr[s\{< B_1^h, B_j^h >\} = \alpha$ (or β) $] = 0$.

A motivation for step 1 (the initialization of $< B_1^h, B_j^h >$) is our intuition—which we are unable to establish mathematically—with respect to δ. On the one hand, if shifting from deterministic to stochastic leadtimes τ_j leads to an increase ($\delta > 0$) in the required level of B_j for the case where the CW does not carry any stock ($B_1 = 0$), our expectation is that the required increase in B_j (for given $B_1 > 0$)—when the CW does carry some of the system inventory—must be *at least* equal to δ for the given service level to be satisfied. This would seem to make sense because, when $B_1 > 0$, the retail site appears to be at greater risk of experiencing stockouts/backorders. On the other hand, if the stochastic leadtimes result in a decrease ($\delta < 0$) in the required B_j for the case where $B_1 = 0$, we would expect the magnitude of the corresponding decrease in B_j (for given $B_1 > 0$) to be *at most* equal to $|\delta|$.

We provide here a simple case study to illustrate application of the heuristic.

Case Study

Consider one warehouse supplying three retail stores selling, among others, a certain large screen TV model. Demand at each retail store is independent of demand at every other retail store, averaging 12 units per week and observed to essentially follow a Poisson distribution. The warehouse places replenishment orders with the electronics manufacturer every two weeks, with a guaranteed delivery leadtime of one week. Each retail site places a replenishment order once a week. The average delivery leadtime from the warehouse to each retail site averages one week from the time the order was placed, but has been found to vary between 0.5 and 1.5 weeks, more or less according to a beta(6,2) distribution. The desired probability of stockout is not more than 5%.

The above illustrative case leads to the following computational example:

Example 1

Given N = 3, λ_j = 12, θ_1 = 2, θ_j = 1, τ_1 = 1, and E[τ_j]= 1, we have scenario 3 (refer to Table 1). Graves (1996) reported < B_1^d, B_j^d > = <56, 39> as the optimal base stock policy, with echelon base stock = 56 + (3 × 39) = 173. However, if the warehouse does not carry stock, $B_j^d(0)$ = 60 and echelon base stock = 3 × 60 = 180. With α = 0.95 and τ_j - 0.5 ~ beta(6,2), Solis and Schmidt (2007) have reported $B_j^s(0)$ = 63, leading to an echelon base stock of 3 × 63 = 189. Thus, δ = $B_j^s(0)$ - $B_j^d(0)$ = 63 - 60 = 3.

Step 1. Initialize: < B_1^h, B_j^h > ← < B_1^d, B_j^d + δ > = <56, 42>, with echelon base stock = 173 + 3δ = 182. s{<56, 42>} = 92.8% < α.

Step 2. B_j^h ← B_j^h + 1 = 43. s{<56, 43>} = 94.8% < α. B_j^h ← B_j^h + 1 = 44. s{<56, 44>} = 96.2% > α. Echelon base stock at <56, 44> is 182 + 6 = 188.

Step 3. In adjusting B_1^h = 56 downward, we find that s{<54,44 >} = 95.2% > α, while s{<53, 44>} = 94.7% < α. < B_1^h, B_j^h > ← <54, 44>, with echelon base stock = 188 - 2 = 186.

The average observed probability of no stockout at the selected policy <54, 44>—based on our typical simulation experiment consisting of 100 replications—does not lead to a statistically significant conclusion that the true mean probability is, indeed, no less than 95%. [We created a separate file (using the Files element in SIMAN) in a worksheet file structure, and used Excel to compute statistics (and report them beyond Arena's one digit after the decimal point).] At the policy <54, 44>, we computed a mean of 95.183(%), with a standard deviation of 1.791. The t statistic for a right hand-tailed hypothesis test, given a sample size of 100, is only 1.024 (which is less than the critical value of 1.6604 at a 5% level of significance). When we expanded our experiment to 400 replications, we observed a mean of 95.186 and a standard deviation of 1.645. The computed t statistic of 2.259 exceeds the critical value.

Table 2 presents figures relevant to our illustration, which will allow us to evaluate how well the heuristic performs, in this case, relative to the "true" minimum level of echelon base stock. Since echelon base stock at the selected policy <54, 44> is 186 units, we now focus our attention on the next lower level of echelon base stock (185 units). Table 2 enumerates all possible (integer valued) policies < B_1, B_j > such that e{< B_1, B_j >} = B_1 + 3B_j = 185 or 186, and such that $B_j \geq B_j^{min}$ = 36. B_j^{min} is a lower bound on B_j, determined by way of a minor heuristic developed by Solis (1997). Below B_j^{min}, the specified service level α will definitely *not* be satisfied. The average observed probability of no stockout s{< B_1, B_j >}, based on our typical experiment (consisting of 100 replications involving 100 CW order cycles per replication), is shown for each of the listed policies. Moreover, for some policies of interest, 95% confidence intervals for

the mean probability of no stockout are shown below the values of s{< B_1, B_j >}. [Arena outputs only two-sided confidence intervals. Figures with one digit after the decimal point are as reported using Arena. Those showing two decimal places, in the case of certain policies of interest, are as computed using Excel.]

Table 2: Average Observed Probabilities of No Stockout (%) – Example 1

Echelon Base Stock = 185			Echelon Base Stock = 186		
B_1	B_j	Average Observed Prob of No Stockout (%)	B_1	B_j	Average Observed Prob of No Stockout (%)
			0	62	92.9
2	61	92.4	3	61	93
5	60	92.4	6	60	93.1
8	59	92.5	9	59	93.3
11	58	92.7	12	58	93.4
14	57	92.9	15	57	93.5
17	56	93	18	56	93.7
20	55	93.2	21	55	93.9
23	54	93.3	24	54	94
26	53	93.5	27	53	94.1
29	52	93.6	30	52	94.3
32	51	93.8	33	51	94.3
35	50	93.8	36	50	94.5
38	49	94	39	49	94.69
41	48	94.1	42	48	94.69
		(93.8, 94.5)			
44	47	94.2	45	47	94.82
		(93.9, 94.6)			
47	46	94.4	48	46	94.99
		(94.0, 94.8)			
50	45	94.48	51	45	95.05
		(94.10, 94.86)			
53	44	94.66	54	44	95.18
		(94.28, 95.03)			(94.83, 95.53)
56	43	94.76	57	43	95.31
		(94.40, 95.12)			
59	42	94.77	60	42	95.29
		(94.42, 95.13)			
62	41	94.79	63	41	95.3
		(94.41, 95.17)			
65	40	94.57	66	40	95.04
		(94.16, 94.98)			
68	39	94.2	69	39	94.62
		(93.8, 94.6)			
71	38	93.5	72	38	93.9
		(93.1, 93.9)			
74	37	92.3	75	37	92.5
77	36	90.4	78	36	90.6

In Example 1, we note that the upper limit of the 95% confidence interval exceeds the specified 95(%) service level at a number of policies < B_1, B_j > for which e{< B_1, B_j >} = 185—specifically, at the policies <53, 44>, <56, 43>, <59, 42>, and <62, 41>. Using Excel, we compute for <53, 44> a 94.657 average probability of no stockout, with a standard deviation of 1.911. The resulting t statistic for a left hand-tailed test of hypothesis on the mean is -1.797, which is less than the critical value at a 5% level of significance. Tests are inconclusive for the other three policies. By again expanding our experiment with each of these three policies, we find, at a 5% level of significance, that mean service levels are below α = 0.95. [For instance, for the policy <62, 41>, an expanded experiment of 200 replications showed a mean of 94.763 and a standard deviation of 1.876, yielding a t statistic of -1.784 (which is below the critical value for a left hand-tailed test of hypothesis at a 5% level of significance).] Hence, we are able to infer, at a 5% level of significance, that no policy < B_1, B_j > involving 185 units of echelon base stock satisfies α = 0.95.

Having found, at a 5% level of significance, that the selected policy <54, 44>—with 186 units of echelon base stock—satisfies α = 0.95, and that no policy involving 185 units of echelon base stock is able to satisfy the given service level, we now make important theoretical observations that apply to the general situation.

Lemma 1   For any given B_j, *the probability of no stockout associated with the policy* $< B_1, B_j >$ *increases monotonically with* B_1.
Proof: In section 2, we had noted that $S_1 \sim$ gamma(λ_1, B_1). It follows that $E[S_1] = B_1/\lambda_1$, and $E[S_1]$ increases with B_1. Since $T_j = \min\{p_j, S_1\}$, then $E[T_j]$ also increases with B_1. Accordingly, the expected width of the time interval $(T_j, t_r]$, given τ_j, decreases with B_1. Therefore, for any given level of retailer base stock B_j, Pr$\{ (D_j[T_j, t_r] \mid \tau_j) \le B_j \}$ increases with B_1, as does the probability of no stockout, Pr$\{ D_j[T_j, t_r] \le B_j \}$. *q.e.d.*

Lemma 2   Take any level E *of echelon base stock. Consider the sets* P_E *and* P_{E-1} *of all (integer-valued) policies* $< B_1, B_j >$ *at the echelon base stock levels* E *and* E-1, *respectively. Then the maximum probability of no stockout possible for all policies in* P_{E-1} *will be less than the maximum probability of no stockout possible for all policies in* P_E.
Proof: Let s_{E-1} be the maximum probability of no stockout for all policies in P_{E-1} and s_E be the maximum probability of no stockout for all policies in P_E. Assume that $s_{E-1} \ge s_E$. Let the probability s_{E-1} be at some specific policy $< B_1, B_j >$ in P_{E-1}. Then the policy $< B_1+1, B_j >$ is one of the policies in the set P_E, since e$\{< B_1, B_j >\}$ = $B_1 + NB_j = $ E-1 and e$\{< B_1+1, B_j >\}$ = $(B_1+1) + NB_j = $ E. If s^+ denotes the probability of no stockout at $< B_1+1, B_j >$, then $s^+ \le s_E$—and, therefore, $s^+ \le s_{E-1}$ according to the assumption we made. This runs contrary to Lemma 1, which tells us that s^+ (the probability of no stockout at $< B_1+1, B_j >$) must be greater than s_{E-1} (the probability of no stockout at $< B_1, B_j >$). We conclude that $s_{E-1} < s_E$. *q.e.d.*

Returning to Example 1, it follows from Lemma 2 that the maximum service level possible for any echelon base stock level less than 185 will be less than that for E = 185, at which no "candidate" policy satisfies α = 0.95. Hence, we conclude that the minimum level of echelon base stock that satisfies α = 0.95 is 186 units.

Interpretation of Heuristic Solution for the Case Study
To achieve a 95% probability of not stocking out, the optimal base stock policy with deterministic leadtime τ_j = 1 is $< B_1^d, B_j^d > = <56, 39>$, with echelon base stock = 173 units. Using this deterministic leadtime optimal policy as the starting point, the search heuristic under stochastic leadtimes yielded the policy $< B_1^h, B_j^h > = <54, 44>$, with echelon base stock of 54 + (3 × 44) = 186 units. This base stock policy

satisfies the specified 95% probability of not stocking out and yields the optimal echelon base stock in the given case when the warehouse is made to carry stock. It compares favorably with the optimal retailer base stock level of 63 units, leading to an echelon base stock of 3 × 63 = 189 units, if the warehouse does not carry stock (Solis and Schmidt 2007).

The next example involves the fill rate criterion.

Example 2
Consider scenario 7 (N = 3), with τ_j - 4.5 $\sim$ beta(2,6) and β = 0.99. We know that $< B_1^d, B_j^d > = <59, 91>$, with echelon base stock = 332; $B_j^d(0) = 112$; $B_j^s(0) = 110$; and $\delta = B_j^s(0) - B_j^d(0) = 110 - 112 = -2$.
Step 1. Initialize: $< B_1^h, B_j^h > \leftarrow < B_1^d, B_j^d + \delta > = $ <59, 89>, with echelon base stock = 332 + 3δ = 326. s$\{<59, 89>\}$ = 99.05% > β. In this case, we proceed to step 3.
Step 3. In trying to adjust B_1^h = 59 downward, we find that s$\{<58, 89>\}$ = 98.98% < β. We keep the current policy $< B_1^h, B_j^h > = <59, 89>$, with echelon base stock = 326.

In step 1, at the policy <59, 89>, our typical experiment yielded an average observed fill rate of 99.050(%), with a standard deviation of 0.523. The computed t statistic was only 0.95, which falls below the critical value for the right hand-tailed test of hypothesis at a 5% level of significance. However, in expanding the experiment to 400 replications, we computed a mean fill rate of 99.049 and a standard deviation of 0.535. The resulting t value was 1.818, and we infer that the mean fill rate satisfies β = 0.99 at a 5% level of significance.
A heuristic lower bound on B_j, as developed by Solis (1997), yielded a value of $B_j^{min,h}$ = 83. Table 3 shows the average observed fill rates at policies of interest.
At the echelon base stock level of 325 (one unit below the level at the selected policy <59, 89>), we consider, for instance, the policy <46, 93>. Our typical experiment yielded an average observed fill rate of 99.035(%), with a standard deviation of 0.488. The resulting t statistic of 0.71 is below the critical value at a 5% level of significance for a right hand-tailed test of hypothesis about the mean fill rate. However, when we expanded the experiment to 500 replications, we found a mean of 99.044 and a standard deviation of 0.507, yielding a t value of 1.953, which exceeds the critical value at a 5% level of significance. We have thus found one policy—among a number of policies, in fact— satisfying β = 0.99, and at which the echelon base stock level of 325 is one unit lower than the level of 326 at the policy <59, 89> selected by the heuristic.
At the next lower echelon base stock level of 324, there are eight policies in all (<33, 97>, <36, 96>, <39, 95>, <42, 94>, <45, 93>, <48, 92>, <51, 91>, and <57, 89>) for which left hand-tailed tests of hypothesis concerning the mean fill rate are not conclusive based

on our typical experiments. However, by increasing the number of replications, we are able to infer at a 5% level of significance that all policies with echelon base stock = 324 result in mean fill rates that are lower than β = 0.99. [For example, at the policy <51, 91>, our typical experiment yielded a mean of 98.954(%), with a standard deviation of 0.529, resulting in a t statistic of only -0.862. With 400 replications, the observed mean was 98.947, with a standard deviation of 0.551. The computed t value in the latter case was -1.918, which is to the left of the critical value for the left hand-tailed test of hypothesis at a 5% level of significance.]

Table 3: Average Observed Fill Rates (%) – Example 2

Echelon Base Stock = 324			Echelon Base Stock = 325			Echelon Base Stock = 326		
B_1	B_j	Average Observed System Fill Rate (%)	B_1	B_j	Average Observed System Fill Rate (%)	B_1	B_j	Average Observed System Fill Rate (%)
0	108	98.76	1	108	98.86	2	108	
3	107	98.79	4	107	98.87	5	107	
6	106	98.8	7	106	98.89	8	106	
9	105	98.81	10	105	98.9	11	105	
12	104	98.82	13	104	98.905 (98.80, 99.01)	14	104	
15	103	98.84	16	103	98.928 (98.83, 99.03)	17	103	
18	102	98.87	19	102	98.947 (98.85, 99.04)	20	102	
21	101	98.88	22	101	98.959 (98.86, 99.05)	23	101	
24	100	98.89	25	100	98.975 (98.88, 99.07)	26	100	
27	99	98.9	28	99	98.98 (98.89, 99.07)	29	99	
30	98	98.908 (98.81, 99.01)	31	98	98.991 (98.90, 99.09)	32	98	
33	97	98.919 (98.82, 99.02)	34	97	99.004 (98.91, 99.10)	35	97	
36	96	98.928 (98.83, 99.03)	37	96	99.013 (98.92, 99.11)	38	96	
39	95	98.942 (98.84, 99.04)	40	95	99.024 (98.93, 99.12)	41	95	
42	94	98.953 (98.85, 99.05)	43	94	99.031 (98.94, 99.13)	44	94	
45	93	98.954 (98.85, 99.05)	46	93	99.035 (98.94, 99.13)	47	93	
48	92	98.949 (98.85, 99.05)	49	92	99.026 (98.93, 99.12)	50	92	
51	91	98.954 (98.85, 99.06)	52	91	99.026 (98.93, 99.13)	53	91	
54	90	98.933 (98.83, 99.04)	55	90	99.01 (98.91, 99.11)	56	90	
57	89	98.913 (98.80, 99.02)	58	89	98.985 (98.88, 99.09)	59	89	99.05 (98.95, 99.15)
60	88	98.87	61	88	98.94 (98.83, 99.05)	62	88	
63	87	98.81	64	87	98.88	65	87	
66	86	98.71	67	86	98.77	68	86	
69	85	98.57	70	85	98.63	71	85	
72	84	98.38	73	84	98.43	74	84	
75	83	98.1	76	83	98.14	77	83	

It is difficult to mathematically make theoretical observations for the fill rate criterion as we had done for the probability of no stockout criterion in Lemmas 1 and 2. This difficulty arises because expected backorders do not work out as "neatly" as the probability of not stocking out. Nevertheless, "empirical evidence" from our simulation experiments suggests that, for given B_j, the (observed) fill rate associated with policy < B_1, B_j > is likewise monotone increasing as B_1. We state this observation as well as a statement analogous to Lemma 2 as *conjectures*.

Conjecture 1 For any given B_j, the fill rate associated with the policy < B_1, B_j > increases monotonically with B_1.

Conjecture 2 Take any level E of echelon base stock. Consider the sets P_E and P_{E-1} of all (integer-valued) policies < B_1, B_j > at the echelon base stock levels E and E-1, respectively. The maximum fill rate possible for all policies in P_{E-1} will be less than the maximum fill rate possible for all policies in P_E.

Returning to Example 2, since no policy having echelon base stock = 324 (or less) satisfies β = 0.99, we conclude by virtue of Conjecture 2 that no policy with echelon base stock less than 325 will satisfy the specified fill rate. Hence, the optimal echelon base stock level in this case is 325 units, and the policy <59, 89> selected by the heuristic yields an inventory level that is one unit more than optimal.

The above two examples we have provided illustrate how well our heuristic appears to perform with respect to selecting a policy < B_1, B_j > which yields an echelon base stock level (and, thus, an average system on-hand inventory level) that is close to the smallest possible level at which the specified α or β is satisfied. In these, as well as other test cases we have thus far investigated—albeit limited, in view of the computing effort involved, to only a portion of Graves' 64 test cases under either service criterion—we have found the selected policy to result in inventory levels that are generally either equal to or one unit more than optimal. Considering average inventory levels of roughly between 100 and 400 in our limited test cases, our heuristic seems to perform fairly well.

The choice of the optimal policy < B_1^d, B_j^d > in the deterministic leadtime case as the starting point for the heuristic appears to be an appropriate one. We are unable to provide a guarantee, however, that the heuristic would err by no more than one unit away from the optimal level of average system inventory. We have found that the average observed service levels (based on comparable simulation experiments) at policies having the same echelon base stock are not always unimodal. One, two or more relative maxima may exist within a given echelon base stock level E. We had earlier noted (by way of Lemma 1 or Conjecture 1) that, for given B_j, observed service levels increase as B_1 (and, hence, as the echelon base stock level E = B_1 + NB_j) increases; however, we were unable to observe any uniformity or pattern in the behavior of such increases across varying B_1 values (for a given B_j) or across different B_j values.

Lemma 2 or Conjecture 2 states that the maximum service level (probability of no stockout or fill rate, as the case may be) possible for echelon base stock level E will be larger than the maximum for E-1. We have observed, however, that the maxima at different echelon base stock levels may be located at various levels of B_j. The service level observed for some selected policy < B_1, B_j >, having echelon base stock level E, may satisfy the specified service level α or β fairly closely (in particular, in view of the steps in our heuristic). Yet, the maximum service level at the next lower echelon base stock level E-1 (which could well be at some policy < B_1', B_j' > ≠ < B_1-1, B_j >)—or that at the even lower echelon base stock level E-2, for that matter—may still be able to satisfy α or β.

4. CONCLUSION
In this study, we extended Graves' one-warehouse, N-retailer model by introducing, as Graves suggested, stochastic leadtimes τ_j between the CW and the retail

sites in place of the original deterministic leadtimes. We investigated the effects of stochastic leadtimes—on the optimal base stock policy $< B_1, B_j >$ for the deterministic leadtime case, in which the CW actually carries stock—by way of simulation studies. We have devised a heuristic for selecting, in the stochastic leadtime case, a policy $< B_1, B_j >$ that seeks to minimize average system inventory. The heuristic appears, based on test cases we have investigated, to select a policy at which the average system inventory is equal or very close to the optimal average inventory level.

REFERENCES

Eppen, G. and Schrage, L., 1981. Centralized ordering policies in a multi-warehouse system with lead times and random demand, In: L.B. Schwarz, ed. *Multi-level production/inventory control systems: theory and practice*, TIMS Studies in the Management Sciences, 16, Amsterdam: North Holland, 51-67.

Graves, S.C., 1985. A multi-echelon inventory model for a repairable item with one-for-one replenishment, *Management Science*, 31 (10), 1247-1256.

Graves, S.C., 1996. A multiechelon inventory model with fixed replenishment intervals. *Management Science*, 42 (1), 1-18.

Kelton, W.D., Sadowski, R.P., and Sturrock, D.T., 2007. *Simulation with Arena*, 4th ed., New York, NY: McGraw-Hill.

Lee, H.L. and Moinzadeh, K., 1987. Two-parameter approximations for multi-echelon repairable inventory models with batch ordering policy. *IIE Transactions*, 19, 140-149.

McGavin, E.J., Schwarz, L.B., and Ward, J.E., 1993. Two-interval inventory-allocation policies in a one-warehouse, N-identical-retailer distribution system. *Management Science*, 39 (9), 1092-1107.

Nahmias, S. and Smith, S.A., 1994. Optimizing inventory levels in a two-echelon retailer system with partial lost sales. *Management Science*, 40 (5), 582-596.

Pegden, C.D., Shannon, R.E., and Sadowski, R.P., 1995. *Introduction to simulation using SIMAN*, 2nd ed., New York, NY: McGraw-Hill.

Solis, A.O., 1997. Evaluation of the negative binomial approximation and stochastic leadtimes in a multi-echelon inventory model. Thesis (PhD). The University of Alabama.

Solis, A.O., 1998. On the depot effect in a multi-echelon inventory model with fixed replenishment intervals. *Proceedings of the Twenty-Seventh Annual Meeting of the Western Decision Sciences Institute*, pp. 537-540, April 7-11, Reno (Nevada, USA).

Solis, A.O., and Schmidt, C.P., 2007. Stochastic leadtimes in a one-warehouse, N-retailer inventory system with the warehouse not carrying stock. *European Journal of Operational Research*, 181 (2), 1004-1013.

Solis, A.O., Schmidt, C.P., and Conerly, M.D., 2007. On the effectiveness of the negative binomial approximation in a multi-echelon inventory model: a mathematical analysis. *Proceedings of the International Conference of Numerical Analysis and Applied Mathematics*, American Institute of Physics Conference Proceedings 936, pp. 531-534, September 16-20, Corfu (Greece).

WAREHOUSE INVENTORY MANAGEMENT BASED ON FILL RATE ANALYSIS

Antonio Cimino[a], Duilio Curcio[b], Giovanni Mirabelli[c], Enrico Papoff[d]

[a] [b] [c] [d]Modeling & Simulation Center - Laboratory of Enterprise Solutions (MSC – LES)
M&S Net Center at Department of Mechanical Engineering
University of Calabria
Via Pietro Bucci, Rende, 87036, ITALY

[a] [b] [c] [d]{acimino, dcurcio, g.mirabelli, e.papoff}@unical.it

ABSTRACT

The focus of this paper is the warehouse inventory management considering three inventory control policies under the effect of different demand patterns and lead times. The behavior of the inventory control policies in terms of fill rate is tested and the relationship between the policies and the demand patterns and lead time is evaluated.

Keywords: warehouse inventory management, inventory control policies, fill rate, Modeling & Simulation

1. INTRODUCTION

Different research studies on warehouse systems' design, planning and control have been carried out during the last years. Ratliff and Rosenthal (1983) address the problem of order-picking in a rectangular warehouse with crossovers only at the end of the aisles. Cormier and Gunn (1996) examine simple analytical models to find the unique warehouse size that satisfies some service requirements or minimizes costs. Chen and Samroengraja (2000) carry out a research study on a warehouse, multi-retailer systems for testing the effectiveness of two allocation policies and to define their optimal parameters values. Roodbergen and De Koster (2001) introduce a routing algorithm for finding the shortest order picking tour within a warehouse. Amato et al. (2005) asses the difference between planning and control activities; planning activities include items assignment to shelves levels and warehouse design problems, control activities are related to activities scheduling and deliveries monitoring. Hsieh and Tsai (2006) implement a simulation model for finding the optimum design parameters of a real warehouse system. Concerning the inventory management and optimal items allocation on the shelves, Christofides and Colloff (1973) implement an algorithm for achieving the desired rearrangement of items minimizing items movements and costs; Ahire and Schmidt (1996) analyze the performance of a warehouse, n-retailer system with a continuous review inventory policy. Hoare and Beasley (2001) carry out a research study on

a system development for the effective design of stockroom layouts.

The main goal of this paper is to analyze three different inventory control policies within a warehouse under the effect of different demand patterns and lead times. The performance measure being considered for monitoring policies behavior is the fill rate. The authors develop a simulation model of the warehouse, implement the inventory control policies and perform a number of simulation experiments (supported by statistical techniques) for achieving the objective above mentioned. Before getting into the details of the study, a brief summary of the paper is presented. Section 2 reports a description of the analytical models of the inventory control policies; section 3 describes the warehouse simulation model; section 4 presents the analysis of the inventory control policies under the effect of demand patterns and lead times; finally, conclusions summarize critical issues and results of the paper.

2. THE INVENTORY CONTROL POLICIES

The inventory control policies being considered are the following:

- the periodic-review, order-up-to-level policy *(R, S)*;
- the periodic-review, order-point, order-up-to-level policy *(R, s, S)*;
- the order-point, order-quantity policy *(s, Q)*.

In the sequel a brief description of each inventory control policy is reported. The following notation is used:

- R_i, review period of the item i;
- $S_i(t)$, order-up-to-level at time t of the item i;
- $s_i(t)$, order-point at time t of the item i;
- $I_i(t)$, on-hand inventory at time t of the item i;
- $QO_i(t)$, quantity already on order at time t of the item i;
- $QS_i(t)$, quantity to be shipped at time t of the item i;

- $Q_i(t)$, quantity to be ordered at time t of the item i;
- $D_i(t)$, demand at time t of the item i;
- LT_i, lead time of the item i.

2.1. The periodic-review, order-up-to-level policy (R,S)

As reported in Silver et al. (1998), this policy is the classical replenishment cycle policy, particularly adopted by companies not using computer control techniques. This policy works better when all the items are ordered from the same supplier. At each review period a control on the Inventory Position ($IP_i(t)$) is made and, if necessary, a quantity is ordered to increase it up to the re-order level $S_i(t)$. $IP_i(t)$ is evaluated as the on-hand inventory plus the quantity already on order minus the quantity to be shipped:

$$IP_i(t) = I_i(t) + QO_i(t) - QS_i(t) \tag{1}$$

In particular, the re-order level $S_i(t)$ is evaluated in this way:

$$S_i(t) = LTD_i(t) + SS_i \tag{2}$$

where:

- $LTD_i(t)$ is the demand forecast over the lead time;
- SS_i is the safety stock calculated as a standard deviation of the lead time demand.

The quantity to be ordered $Q_i(t)$ is given from the following equation:

$$Q_i(t) = S_i(t) - IP_i(t) = LTD_i(t) + SS_i - IP_i(t) \tag{3}$$

This inventory control policy is particularly used when the demand pattern changing with time.

2.2. The periodic-review policy (R , s, S)

This policy has to be regarded as a combination of the order-point, order-up-to-level policy (s,S) and of the periodic-review, order-up-to-level policy (R,S). $IP_i(t)$ is checked every review period so two cases can occur:

- $IP_i(t)$ is at or below the re-order point $s_i(t)$;
- $IP_i(t)$ is above $s_i(t)$.

In the first case the quantity to be ordered is enough to raise the $IP_i(t)$ to $S_i(t)$ while in the second nothing is ordered until the next review. According to Silver et al. (1998), it is demonstrated that, under specific assumptions on demand pattern and cost factors, the (R,s,S) policy generates total costs lower than other inventory control policies. More in detail, in this policy the re-order point is evaluated according to equation 4:

$$s_i(t) = LT * \frac{\sum_{t}^{t+R-1} D_i(t)}{R_i} + SS_i \tag{4}$$

while:

$$S_i(t) = \frac{\sum_{t}^{t+R-1} D_i(t)}{R_i} + s_i(t) \tag{5}$$

The quantity to be ordered is:

$$Q_i(t) = S_i(t) - IP_i(t) \tag{6}$$

2.3. The order-point, order-quantity policy (s,Q)

This policy is a continuous review policy. The fixed quantity $Q_i(t)$ is ordered when the $IP_i(t)$ equals or it is lower than the re-order point $s_i(t)$. In particular, the quantity to be ordered is defined according to the economic order quantity (EOQ).

3. THE SIMULATION MODEL

A general warehouse simulation model describes the system under study in details, but it can lacks in some aspects like flexibility: a warehouse simulation model implemented to satisfy specific requests could not be applied to solve problems of a warehouse system different from that considered. From the other side, a generalized flexible simulation model could be applied to analyze general problems of warehouse systems under different operative scenarios.

The simulation model implemented by the authors aims at achieving a certain level of flexibility in order to be adopted easily for studying similar warehouse systems and to provide the user with a tool for investigating the warehouse behavior under different operative conditions. In effect, the simulation model proposed by authors reproduces all the most important processes and operations that take place inside a warehouse. One of the performance measure monitored by the simulation model is the fill rate (the performance measure for monitoring inventory control policies behavior).

3.1. The model architecture

As before mentioned, the simulation model presented in this research work reproduces all the most important processes and operations of a warehouse:

- orders processing and picking;
- trucks arrival and departure;
- data collection and monitoring (number of handled packages, trucks waiting times and service levels, fill rate, etc.).

The software tool adopted for the simulation model implementation is the commercial package Anylogic™ by *XJ Technologies*. In particular, for reproducing each

process and for increasing model flexibility, different classes have been implemented by using the software library. In addition, ad-hoc programmed routines have been written for implementing all the logics and rules governing the system.

4. DEMAND PATTERNS AND LEAD TIME CONSTRAINTS

The behavior of the three inventory control policies has been investigated under the effect of different demand patterns and lead time constraints. The following factors and notation have been considered:

- demand intensity (*IN*) which can assume three different conditions (low, medium, high);
- demand variability (*VAR*) which can vary from low to high conditions;
- lead time (*LT*) which can be changed respectively in one day, three and five days.

The combinations of such parameters' values provide different operative scenarios and affect the fill rate values of the warehouse system. For evaluating the effect of each possible factors level combination on the performance measure (the fill rate), the Full Factorial Experimental Design is adopted. Table 1 shows factors and levels adopted for the design of experiments.

Table 1: Factors and Levels of DOE

Factors	*Level 1*	*Level 2*	*Level 3*
IN	Low (-1)	Medium (0)	High (+1)
VAR	Low (-1)	Medium (0)	High (+1)
LT	1 (-1)	3 (0)	5 (+1)

As shown in Table 1, each factor has three levels: in particular, Level 1 (-1) indicates the lowest value for the factor, Level 2 (0) the medium value while Level 3 (+1) represents the greatest value. In order to test all the possible factors combinations, the total number of the simulation runs is 3^3 (3 factors x 3 levels x 3 values). Each simulation run has been replicated twice, so the total number of replications is 27 (27x2=54).

The analysis of the simulation results is proposed in the following section.

5. SIMULATION RESULTS ANALYSIS

The results of the simulation model have been analyzed by means of ANOVA and of several graphical methods. The ANOVA partitions the total variability of the performance index in different components due to the influence of the factors considered. According to Montgomery and Runger (2003), the total variability in the data, measured by the total corrected sum of squares SQ_T, can be partitioned into a sum of squares of differences between treatment (factor level) means and the grand mean denoted $SQ_{Treatments}$ and a sum of squares of differences of observations within a treatment from the treatment mean denoted SQ_E, as reported in equation 7.

$$SQ_T = SQ_{Treatments} + SQ_E \qquad (7)$$

More in detail, the difference between observed treatment means and the grand mean defines differences between treatments, while observations' differences within a treatment from the treatment mean can be due only to random errors. As a consequence, examiners can understand how each factor impacts on the performance index introducing an analytical relation (called *meta-model* of the simulation model) between the performance index and factors. In particular, the relation for a three-factor-factorial experiment is:

$$Y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i=1}^{3} \sum_{j>i}^{3} \beta_{ij} x_i x_j +$$
$$+ \sum_{i=1}^{3} \sum_{j>i}^{3} \sum_{k>j}^{3} \beta_{ijk} x_i x_j x_k + \varepsilon_{ijkn} \qquad (8)$$

where:

- *Y* is the performance index;
- β_0 is a constant parameter common to all treatments;
- $\sum_{i=1}^{3} \beta_i x_i$ are the three main effects of factors;
- $\sum_{i=1}^{3} \sum_{j>i}^{3} \beta_{ij} x_i x_j$ are the three two-factor interactions;
- $\sum_{i=1}^{3} \sum_{j>i}^{3} \sum_{k>j}^{3} \beta_{ijk} x_i x_j x_k$ represents the three-factor interaction;
- ε_{ijkn} is the error term;
- *n* is the number of total observations.

In this research study, ANOVA is adopted for a twofold reason:

- during the first step, it is used as a screening tool in order to determine which factors are most significant on the fill rate (*sensitivity analysis*);
- subsequently, ANOVA is used for investigating the effect of the most significant factors in order to evaluate the coefficients of the input-output meta-model reported in equation 8.

5.1. Simulation results analysis for the fill rate – (R,S) inventory policy

Table 2 shows all the 27 combinations of the input factors; the first three columns report the experimental design matrix while the last column reports the simulation results in terms of fill rate (one replication) for the *(R, S)* policy

Table 2: Experimental design matrix and simulation results for the Fill Rate – *(R,S)* policy

IN	VAR	LT	(R,S)
-1	-1	-1	0,931
-1	-1	0	0,896
-1	-1	1	0,881
-1	0	-1	0,901
-1	0	0	0,896
-1	0	1	0,892
-1	1	-1	0,895
-1	1	0	0,883
-1	1	1	0,868
0	-1	-1	0,899
0	-1	0	0,895
0	-1	1	0,878
0	0	-1	0,899
0	0	0	0,899
0	0	1	0,886
0	1	-1	0,889
0	1	0	0,87
0	1	1	0,878
1	-1	-1	0,87
1	-1	0	0,877
1	-1	1	0,876
1	0	-1	0,87
1	0	0	0,87
1	0	1	0,863
1	1	-1	0,859
1	1	0	0,869
1	1	1	0,859

Table 3 reports the sensitivity analysis results provided: the non-negligible effects are characterized by a *p-value* $\leq \alpha$ where *p* is the probability to accept the hypothesis that the factor has no impact on the performance index and $\alpha=0.05$ is the confidence level adopted in the analysis of variance. In this table:

- the first column reports the sources of variations;
- the second column is the degree of freedom (*DOF*);
- the third column is the Sum of Squares;
- the 4th column is the Mean Squares;
- the 5th column is the Fisher statistic;
- the 6th column is the p-value.

Table 3: Sensitivity Analysis Results – *(R,S)* policy

Source	DF	Adj SS (10⁻⁴)	Adj MS (10⁻⁴)	F	P
IN	2	32,51	16,25	21,5	0,001
VAR	2	10,98	5,49	7,26	0,016
LT	2	9,72	4,86	6,43	0,022
IN*VAR	4	1,88	0,47	0,62	0,658
IN*LT	4	6,83	1,70	2,26	0,152
VAR*LT	4	1,33	0,33	0,44	0,777
Error	8	6,05	0,75		
Total	26				

In this case the most significant effects are the first order effects because their p-value is lower than the confidence level. Table 4 shows the coefficients for the input-output meta-model (equation 8). In particular, for each factor the first coefficient value represents the slope of the straight line between its low and medium levels while the value in the column (0) is the slope of the straight line for medium and high factor levels.

Table 4: ANOVA Coefficients – *(R,S)* policy

Term	Coefficient	
	(-1)	(0)
Constant	0,883296	
IN	0,010370	0,004815
VAR	0,005926	0,002926
LT	0,007037	0,000593

Equation (9) reports the input-output meta-model for the performance parameter (the warehouse fill rate) when input parameters change between low and medium levels:

$$Y_{ijkn} = 0,883296 + 0,010370 * IN + \\ + 0,005926 * VAR + 0,007037 * LT \tag{9}$$

Figure 1 shows the fill rate trend in function of the three main effects: the fill rate decreases when demand intensity and variability and lead time increase; in particular, the performance index decreases more rapidly when demand intensity and variability changes from medium to high values.
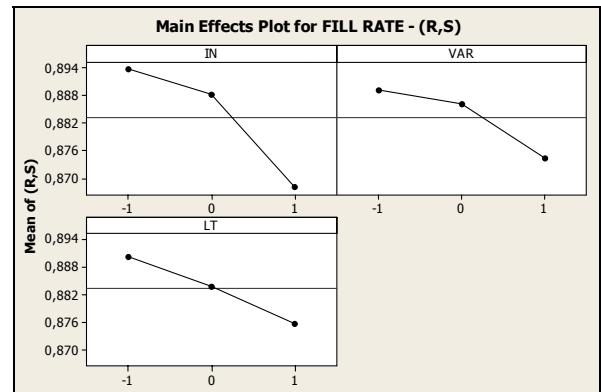
Figure 1: Fill Rate versus Main Effects – *(R,S)* policy

Figure 2 shows the plots concerning the interaction between the main effects. For example, the interaction between demand intensity and variability becomes greater when variability changes from low to medium values while for medium and low variability values the two lines (the black and the red one) are flush. From the other side, the interaction between variability and lead time is more marked when lead time changes from medium to high values.
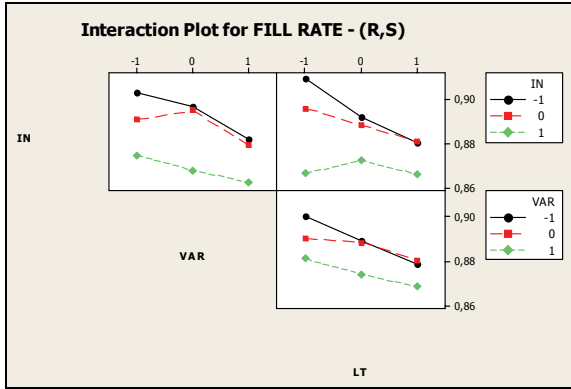
Figure 2: Interaction Plots for Fill Rate

5.2. Simulation results analysis for the fill rate – (R, s, S) inventory policy

Table 5 shows all the 27 combinations of the input factors; the first three columns report the experimental design matrix while the last column reports the simulation results in terms of fill rate (one replication) for the *(R,s,S)* policy. The sensitivity analysis is reported in Table 6: considering a *p-value* ≤ 0.05, the most significant effects are the first order effects. The ANOVA coefficients for the input-output meta-model are displayed in Table 7.

Table 5: Experimental design matrix and simulation results for the Fill Rate – (R,s,S) policy

IN	VAR	LT	(R,s,S)
-1	-1	-1	0,99
-1	-1	0	0,944
-1	-1	1	0,93
-1	0	-1	0,967
-1	0	0	0,941
-1	0	1	0,925
-1	1	-1	0,956
-1	1	0	0,942
-1	1	1	0,937
0	-1	-1	0,973
0	-1	0	0,957
0	-1	1	0,938
0	0	-1	0,952
0	0	0	0,932
0	0	1	0,923
0	1	-1	0,957
0	1	0	0,935
0	1	1	0,907
1	-1	-1	0,961
1	-1	0	0,954
1	-1	1	0,932
1	0	-1	0,936
1	0	0	0,927
1	0	1	0,908
1	1	-1	0,934
1	1	0	0,914
1	1	1	0,917

Table 6: Sensitivity Analysis Results – (R,s,S) policy

Source	DF	Adj SS (10^{-4})	Adj MS (10^{-4})	F	P
IN	2	12,53	6,26	7,39	0,015
VAR	2	22,50	11,25	13,2	0,003
LT	2	53,52	26,76	31,6	0,000
IN*VAR	4	2,92	0,73	0,86	0,526
IN*LT	4	3,11	0,77	0,92	0,499
VAR*LT	4	1,31	0,32	0,39	0,812
Error	8	6,78	0,84		
Total	26				

Table 7: ANOVA Coefficients – (R,s,S) policy

Term	Coefficient	
	(-1)	(0)
Constant	0,940	
IN	0,007667	0,001222
VAR	0,012889	-0,005778
LT	0,018111	-0,001889

The input-output meta-model for the *(R,s,S)* policy is:

$$Y_{ijkn} = 0,94 + 0,007667 * IN + \\ + 0,012889 * VAR + 0,018111 * LT \quad (10)$$

As showed in Figure 3, the fill rate decreases when demand intensity and variability and lead time change from low to medium levels while it remains approximately constant when variability goes from medium to high levels.
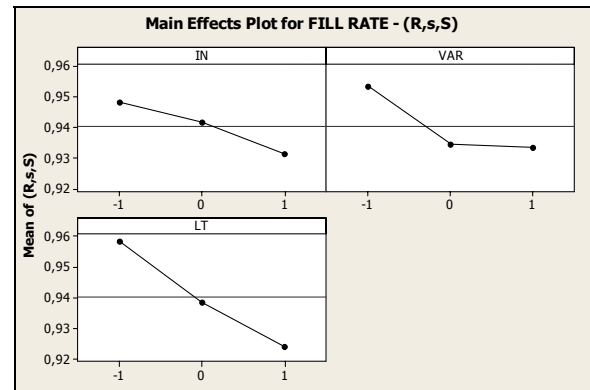
Figure 3: Fill Rate versus Main Effects

5.3. Simulation results analysis for the fill rate – (s,Q) inventory policy

The sensitivity analysis results for the *(s,Q)* policy are reported in Table 8.
The most significant effects are the first order effects and the interaction between the demand intensity and the lead time. Introducing coefficients provided by means of ANOVA, it is possible to evaluate the input-output meta-model (equation 11) for the *(s,Q)* policy (considering the slope of the straight line when each factor changes between its medium and high levels;

Table 8: Sensitivity Analysis Results – (S, Q) policy

Source	DF	Adj SS (10^{-4})	Adj MS (10^{-4})	F	P
IN	2	6,99	3,49	10,3	0,006
VAR	2	7,21	3,60	10,6	0,006
LT	2	4,39	2,19	6,48	0,021
IN*VAR	4	4,27	1,06	3,15	0,078
IN*LT	4	6,25	1,56	4,61	0,032
VAR*LT	4	1,65	0,41	1,22	0,373
Error	8	2,71	0,33		
Total	26				

similar coefficients have been found between low and medium levels).

$$Y_{ijkn} = 0{,}876 + 0{,}002852 * IN +$$
$$+ 0{,}006741 * VAR - 0{,}00081 * LT + \qquad (11)$$
$$- 0{,}003519 * (IN * LT)$$

The validity of results obtained from the ANOVA analysis is still confirmed by the Residuals Analysis.
More in detail:

- the assumption of normality has besn tested using a Normal Probability Plot. The residuals approximately fall along a straight line, consequently the deviation from normality is not severe;
- the assumption of independence has been tested by using the Residuals versus the Order of the Data plot: no positive or negative residuals sequences have been identified;
- the hypothesis of equal variance has been tested by using the Residuals versus the Fitted Values plot: residuals variability doesn't depend on the fitted value;
- the hypothesis of residuals normally distributed has been tested by using the Histogram of the Residuals: the residuals appear as normally distributed.

6. CONCLUSIONS

The main goal of this paper is to analyze three different inventory control policies within a warehouse under the effect of different demand patterns and lead times. The performance measure being considered for monitoring policies behavior is the fill rate. The authors develop a flexible simulation model of a warehouse, implement the inventory control policies and perform a number of simulation experiments (supported by statistical techniques).
The experimental results analyzed by means of ANOVA show that the inventory control policies perform differently in relation to demand intensity and variability levels and lead time values. In addition for each scenario and for each inventory control policy analytical relationships that express the fill rate as function of the demand intensity, demand variability and lead time have been evaluated.

REFERENCES

Ahire, S. L., Schmidt, C. P., 1996. A model for a Mixed Continuous-Periodic Review One-Warehouse, N-Retailer inventory system. *Operational Research*, 92, 69 – 82.

Amato, F., Basile, F., Carbone, C., Chiacchio, P., 2005. An Approach to control automated warehouse systems. *Control Engineering Practice*, 13, 1223–1241.

Axsäter, S. 2003. Optimal Policies for Serial Inventory Systems Under Fill Rate Constraints. *Management Science*, 49(2), 247-253.

Baganha, M. P., Cohen, M.,1998. The stabilizing effect of inventory in supply chains. *Operations Research* 46 (3), S72-S73.

Balci, O., 1998. Verification, Validation and Testing, in *Handbook of Simulation*, edited by J. Banks, pp. 335-393, Wiley Interscience: New York.

Ballou R., 2003. *Business Logistics/Supply Chain Management*. Prentice Hall, Readings.

Beamon, B. M. 1998. Supply chain design and analysis: models and methods, *International Journal of Production Economics* 55 (3), 281-294.

Beamon, B. M. 1999. Measuring supply chain performance.*International Journal of Operations and Production Management* 19 (3), 275-292.

Bertsimas, D., Thiele, A., 2006. A robust Optimization approach to Inventory Theory, *Oper. Res.*, 54(1), 150 – 168.

Birkin, M., Clarke, G., Clarke, M.P, 2002. *Retail Intelligence and Network Planning*. New York: J.Wiley.

Blackwell, R.D., 1997. *From Mind to Market*. New York: Harper-Collins Publisher.

Brunetta, L., Romanin-Jacur, G., 1999. Passenger and baggage flow in an airport terminal: a flexible simulation model. *Air Traffic Management*, 6, 361 – 363.

Bruzzone A.G., Simone S., Bocca E., 2004. Intelligent Management of a Logistics Platform for Fresh Goods. *Proceedings of SCI2004, Orlando, July*.

Bruzzone, A.G., Longo, F., Massei, M., Saetta, S., 2006. The vulnerability of supply chain as key factor in supply chain management. *Proceedings of Summer Computer Simulation Conference* pp. 181-186, Calgary, Canada.

Chan, F.T.S., Chan H.K., 2005. Simulation modeling for comparative evaluation of supply chain management strategies. *Journal of Advanced Manufacturing Technology* 25, 998–1006.

Chang, Y., H. Makatosoris, 2001. Supply chain modelling using simulation, *Int. J. Simulation*, 2001, 2(1), 24 – 30.

Chen, F., Samroengraja, R., 2000. A Staggered Ordering Policy for One-Warehouse, Multiretailer Systems. *Operations Research*, 48 (2), 281 – 293.

Chopra, S., Meindl, P., 2001. *Supply chain management – Strategy, planning, and Operation*, 2001, Prentice-Hall: New Jersey.

Christofides, N., Colloff, I., 1973. The Rearrangement of Items in a Warehouse. *Operations Research*, 21 (2), 577 – 589.

Cohen, M., Lee, H., 1988. Strategic Analysis of integrated production-distribution systems: models and methods. *Operational Research*, 36, 216 – 228.

Cormier, G., Gunn, E.A., 1996. Simple Models and Insights for Warehouse Sizing. *Operational Research*, 47, 690 – 696.

Curcio, D., Longo, F., Mirabelli, G., 2007. Inventory policies comparison in a manufacturing process using modeling & simulation. *Proceedings of the International Mediterranean Modelling Multiconference,* pp. 237 – 242. October 04 – 06, Bergeggi (Italy).

D'Agostino R.B., Stephens, M.A., 1986. *Goodness of fit techniques*, 1986, Marcel Dekker: New York.

Daganzo, C. F., On the stability of the supply chains, *Op. Res.*, 2004, 52(6), 909 – 921.

De Sensi, G., Longo, F., Mirabelli, G., Papoff, E., 2006. Ants colony system for supply chain routes optimization, in the *Harbour, Maritime & Multimodal Logistics Modelling and Simulation Conference*, 763 – 768.

De Sensi, G. , Longo, F., Mirabelli, G., 2007. Inventory policies analysis under demand patterns and lead times constraints in a real supply chain. *International Journal of Production Research*, doi: 10.1080/00207540701528776

Department of Defense, Deputy under Secretary of Defence, DoD modelling and simulation (M&S) management, *DoD Directive 5000.59*, 1994

Dunn R.H., 1987. The quest for software reliability, in *Handbook of software quality assurance*, edited by G.G. Schulmeyer and J.I. McManus, pp. 342 – 384, Van Nostrand Reynold: New york.

Eben-Chaime, M., Pliskin, N., Sosna, D., 1997. Operations management of multiple machine automatic warehousing systems. *Production Economics*, 51, 83–98.

Eben-Chaime, M., Pliskin, N., Sosna, D., 2004. An integrated architecture for simulation. *Computer and Industrial Engineering*, 46, 159–170.

Fleisch, E., Tellkamp, C., 2005. Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain. *International Journal of Production Economics* 95, 373–385.

Ganeshan, R., Boone, T., Stenger, A. J., 2001. The impact of inventory and how planning parameters on supply chain performance: An exploratory study. *International Journal of Production Economics* 71, 111–118.

Giannoccaro, I., Pontrandolfo, P., 2002. Inventory management in supply chains: a reinforcement learning approach, *Int. J. Prod. Econ.*, 2002, 78, 153 – 161.

Graves, S. C., 1991. A single inventory model for a non-stationary demand process, *Manuf. Serv. Op. Manag.*, 1999, 1, 50 – 61.

Graves, S.C., Willems, S. P., 2005. Optimizing the Supply Chain Configuration for New Product. *Management Science*, 51(8), 1165–1180.

Hill, R.M., 1989. Allocating Warehouse Stock in a Retail Chain. *Operational Research*, 40 (11), 983 – 992.

Hoare, N. P., Beasley, J. E., 2001. Placing Boxes on Shelves: A Case Study. *Operational Research*, 52 (6), 605 – 614.

Ingalls, R. G., 1998. The value of simulation in modeling supply chain. *Proceedings of the 1998 Winter Simulation Conference,* pp. 1371-1375 Washington D.C.

Johonson N.L., Kotz, S., Balakrishnan, N., 1994. *Continuous Univariate Distributions*, Vol. 1, 2nd ed., Houghton Mifflin: Boston.

Johonson N.L., Kotz, S., Balakrishnan, N., 1995. *Continuous Univariate Distributions*, Vol. 2, 2nd ed., Houghton Mifflin: Boston.

Johonson N.L., Kotz, S., Kemp, A.W., 1992. *Univariate Discrete Distributions*, 2nd ed.*,* Houghton Mifflin: Boston.

Kelle, P., Milne, A. The effect of (s, S) ordering policy on the supply chain, *Int. J. Prod. Econ.*, 59, 113 – 122.

Kreipl, S., Pinedo, M., 2004. Planning and scheduling in supply chains: an overview of Issues in Practise, *Prod. Op. Manag.*, 2004, 13(1), 77 – 92.

Lee, C., Çetinkaya, S., Jaruphongsa, W., 2003. A Dynamic Model for Inventory Lot Sizing and Outbound Shipment Scheduling at a Third-Party Warehouse. *Operations Research*, 51 (5), 735 – 747.

Lee, H.L., Billington, C., 1993. Material Management in Decentralized Supply Chains. *Operations Research*, 41(5), 835-847.

Lee, Y. H., Cho, M. K., Kim, S. J., Kim, Y. B., 2002. Supply chain simulation with discrete continuous combined modeling. *Computer & Industrial Engineering*, 43, 375-392.

Li, J., Shaw, M., Sikora, R., Tan, G. W., Yang, R., 2001. *The effects of Information sharing strategies on supply chain performance*, Working paper, University of Illinois Urbana-Champaign.

Liebeskind, A., 2005. *How to Optimize Your Warehouse Operations*. 1st ed. USA: Industrial Data & Information Inc.

Longo, F., Mirabelli, G., Papoff, E., 2005. Modeling Analysis and Simulation of a supply chain devoted to support pharmaceutical business retail, *Proceedings of the 18th International Conference on Production Research, July 31 – August 4 , Salerno, Italy.*

Longo, F., Mirabelli, G., Papoff, E., 2006. Material Flow Analysis and Plant Lay-Out Optimization of a Manufacturing System. *International Journal of Computing*, 5(1), 107-116.

Longo, F., Mirabelli, G., 2008. An Advanced Supply Chain Management Tool Based on Modeling & Simulation, *Computer and Industrial Engineering*, Vol 54(3), 570-588.

Moinzadeh, K., 2002. A Multi-Echelon Inventory System with Information Exchange. *Management Science*, 48(3), 414-426.

Monks, J.G., 1996. *Operations Management*, New york: McGraw Hill.

Montgomery, C.D. and Runger, G.C., 2003. *Applied Statistics and Probability for Engineers*. 3rd ed. USA: John Wiley & Sons.

Mulcahy, D.E., 1993. *Warehouse Distribution and Operations Handbook*. 1st ed. USA: McGraw-Hill.

Nahmias, S., 1982. Perishable Inventory Theory: A Review. *Operations Research*, 30 (4), 680 – 708.

Napolitano, M., Gross, J., 2003. *The Time, Space & Cost Guide to Better Warehouse Design*. 2nd ed. USA: Distribution Group.

O'Leary, D.E., 2000. *Enterprise Resource Planning Systems: Systems, Life Cycle, Electronic Commerce, and Risk*. Cambridge University Press.

Parker, R.P., Kapuscinski, R., 2004. Optimal Policies for a Capacitated Two-Echelon Inventory system, *Op. Res.*, 2004, 52(5), 739 – 755.

Persson, F. , Olhager, J., 2002. Performance simulation of supply chain designs, *International Journal of Production Economics*, 77, 231-245.

Ratliff, H.D., Rosenthal, A.S., 1983. Orderpicking in a rectangular warehouse: a solvable case of traveling salesman problem. *Operations Research*, 31, 507 – 521.

Roberson, J.F, 1994. *Logistics Handbook*. Free Press.

Roundy, R. O., Muckstadt, J. A., 2000. Heuristic computation of periodic-review base stock inventory policies, *Manag. Sc.,* 46(1), 104 – 109.

Sanjay, J., 2004. Supply chain management tradeoffs analysis, *Proceedings of the Winter Simulation Conference*, pp. 1358-1364. Washington (USA).

Seifert, D., Marzian, R., McLaughlin, J., 2003. *Collaborative Planning, Forecasting, and Replenishment: How to Create a Supply Chain Advantage*. Amacom

Silver, E.A., Pyke, D.F., Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*. 3rd ed. USA: John Wiley & Sons.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E. *Designig and Managing the Supply Chain*. McGraw-Hill.

Stadtler, H., Kilger, C., 2000. *Supply chain management and advanced planning*, Springer: Berlin.

Suwanruji P., Enns, S.T., 2006. Evaluating the effects of capacity constraints and demand patterns on supply chain replenishment strategies, *Int. J. of Prod. Res.*, 44(21), pp. 4607 – 4629.

Thor, C.G., 1994. *Measures of success–creating a high performing organization*. Essex Junction, VT: Oliver Wight.

Towill, D. R., 1991. Supply chain dynamics, *Int. J. Comp. Integr. Manuf.*, 4(4), 197 – 208.

Wikner, J., Towill, D. R., Naim, M., 1991. Smoothening supply chain dynamics, *Int. J. Prod. Econ.*, 3(3), 3 – 13.

Zhang, C., Zhang, C., 2007. Design and Simulation of demand information sharing in a supply chain. *Simulation Modeling Practice and Theory*, 15, 32-46.

Zipkin, P., 2000. *Foundations of Inventory Management*, McGraw-Hill Higher Education: Boston.

AUTHORS BIOGRAPHIES

ANTONIO CIMINO was born in Catanzaro (Italy) in October the 1th, 1983. He took his degree in Management Engineering, summa cum Laude, in September 2007 from the University of Calabria. He is currently PhD student at the Mechanical Department of University of Calabria. His research activities concern the integration of ergonomic standards, work measurement techniques, artificial intelligence techniques and Modeling & Simulation tools for the effective workplace design. He collaborates with the Industrial Engineering Section of the University of Calabria to research projects for supporting innovation technology in SMEs. His e-mail address is <acimino@unical.it>.

DUILIO CURCIO was born in Vibo Valentia (Italy), on December the 15th, 1981. He took the degree in Mechanical Engineering from University of Calabria (2006). He is currently PhD student at the Mechanical Department of University of Calabria. His research activities include Modeling & Simulation and Inventory Management theory for production systems and Supply Chain design and management. He collaborates with the Industrial Engineering Section of the University of Calabria to research projects for supporting Research and Development in SMEs. His e-mail address is <dcurcio@unical.it>.

GIOVANNI MIRABELLI was born in Rende (Italy), on January the 24th, 1963. He took the degree in Industrial Technology Engineering from University of Calabria. He is currently researcher at the Mechanical Department (Industrial Engineering Section) of University of Calabria. His research interests regard work measurement and human reliability. His e-mail address is <g.mirabelli@unical.it>.

ENRICO PAPOFF was born in Naples (Italy) on February the 03rd, 1948. He took the degree in Mechanical Engineering from University of Napoli Federico II, in 1973. He is currently Associate Professor at the Mechanical Department (Industrial Engineering Section) of the University of Calabria. His research interests regard project management and business plans.

EFFECT OF NEGATIVE PARAMETERS IN INVENTORY MODELS PERFORMANCE

Miguel Cezar Santoro[a], Gilberto Freire[b]

[a] [b] Department of Industrial Engineering, EPUSP, University of São Paulo, Brazil

[a]santoro@usp.br [b]gfreire@usp.br

ABSTRACT

This paper presents a comparative study of four inventory models. Three of them are reactive ones, the Order Up to Maximum Inventory Level, Base Stock and Fixed Lot Size models which don't use demand forecasting to quantify the acquisition decisions. One of them, referred as Requirement Planning, operates using demand forecast. The Requirement Planning and Base Stock models operate with minimal purchase order quantity parameter to improve their competitiveness. Reorder Level and Safety Stock parameters are permitted to be negative. Average operation cost per period is the performance measurement considered, restricted to a minimum service level. The analysis uses a search model jointly a simulator to optimize the parameters. The paper analyses the effects, in the relative models performance, of the negative parameters permission and the value of the Purchase Cost and effectuates a sensibility analysis by testing four different Service Levels.

Keywords: inventory, simulation, search methods, forecast

1. INTRODUCTION

The importance of customer satisfaction for business was demonstrated by Japanese success in the last decades. Methods to improve customer service levels with good cost performance were developed by them and now are used worldwide for the same objective.

Just-in time (JIT) is one of these methods. Usually understood like a production system, it is also a concept of ideal materials flow: the production or supply of items in the exact demand moment and quantities, neither before nor after. This ideal flow condition is the most efficient when low inventory quantities and demand satisfaction are the goals of the production system. The system is managed to low inventory quantities and no shortage. But this is not enough to assure cost efficiency. JIT concept often implies a large number of acquisition/production orders, which can result in large operation costs when the purchase/production fixed costs are high. To specify a minimum service level and search a model with the lowest operation cost is a common way of combine these goals.

To attend the JIT concept and keep the purchase cost low are vital to an inventory model achieve good performance in this environment.

Traditional inventory models decide the moment and the quantity to replenish inventory using parameters and inventory level. They "react" to a system condition, justifying the adjective "reactive" that is used here to identify them. Reactive models need safety stocks to face the total demand variability during the time they need to react, and only when demand variability and reaction time are close to zero they attain the JIT condition.

Another class of inventory models can be seen in Material Requirements Planning systems (MRP). They use forecasts to decide when and how much to buy or to produce. Named here Requirements Planning (RP), this "active" model can anticipate the demand, adjusting the quantities and the times close to the moment of use, the JIT condition, reducing the necessity of safety stock needed to face only forecast errors, instead of all demand variability, which the reactive models need to face. This characteristic, combined with good available forecasts, permits, in theory, the model to be close with JIT concept in demand variation environments. In the other hand, it tends to purchase more times than reactive models (Stock Base is an exception), incurring in high operation costs when the inventory system has high purchase cost.

The comparison of inventory models performances in different conditions becomes interesting since each model has weak and strong points.

A comparative cost performance study was performed, using 4 inventory models, 3 reactive and the RP inventory model. Sixteen different combinations of 2 purchase costs, 4 service levels and permission for models parameters assume negative values (important for improve Up to Order Level and Requirements Planning performances) were tested in 160 items with 36 historical demand real data of an mobility products company. Simulation and search methods were used to optimize models parameters to get, for each model in each demand environment, the minimum mean operation cost condition.

2. LITERATURE REVISION

Inventory uses mathematical models since 1913, when F. W. Harris published his famous article about the

economic lot size calculation. During the II World War, mathematic formulation of models and parameters calculations were focused, particularly for the reorder point and lot size under particular situations of demand, supplier and costs environments. Several works were important references in the dissemination of this approach, like the articles of Dvoretzky and Kiefer (1952) and Arrow *et al.* (1958), and the books of Hadley and Whitin (1963), Naddor (1966) and Brown (1967).

The inventory models initially presented by this approach, today known as traditional models, allow decide when and how much to reorder without demand forecast. The demand is considered continuous with constant average and explained by a fixed and known probability distribution. Studies about inventory models during the 60s and 70s adopted demand behavior based on static Normal and Poisson distributions, which allowed convenient mathematical treatment. Browne and Zipkin (1991) studied the limits of this traditional approach. Agarwall (1974) is an exception in this period, showing that the environment changes are important and modify the best model for inventory management.

The first reference to forecasting applied to inventory is Brown (1967). This author recommends the use of forecasting to calculate reactive inventory models parameters to achieve better efficiency. In the 70's, Eilon and Elmaleh (1970) continued this approach. At the same time, Just-in-time is discussed in the academy, due to the Japanese success in occidental markets, and the idea of reorder only the necessary material instead of lots starts to be a must.

The Material Requirements Planning (MRP) system appeared in sequence. Initially, its focus was the decisions of supply/ production of items with dependent demand which appears in production and assembly structures of items with independent demand. But the need to include trend, seasonality and jumps in demand leaded research to decisions based on forecast. The chronology and relevant aspects of this can be seen in Sipper and Bolfin (1998). Unless the original MRP idea was to buy/produce only the necessary quantities, the high purchase cost in several items leads to several lot forming models applied to MRP systems. Hax (1984) and Orlick (1995) present these models.

Lee and Adam (1986) continued the research of forecasting in inventory. They studied the forecast errors impact in MRP systems, comparing several reactive inventory models performances applied to the dependent items.

Gardner (1990), using real demand historical data of independent items, studied the performance of the inventory system based on the classical Economic Lot Size model.

Defining the lots acquisitions based on forecast demands, generated by four different forecasting models, he calculated "trade off" curves between inventory investment and customer lead-time. Simulation of the system operation during the period of

historical demand data was used in the calculations. His main conclusion was the importance of forecasting errors to the amount of inventory investment needed to achieve a specific customer lead-time. The smaller the error, the smaller the investment needed. The study is important, too, due to the adoption of the joint performance measurement of the forecasting and inventory models.

Fildes and Beard (1992) studied the forecasting use in the production and inventory control. They analyzed the typical characteristics of inventory data and several forecasting models, and concluded that researchers neglected the production and inventory control area, the commercial systems had inadequate forecasting models and users experimented unnecessary large errors, inventories and poor demand satisfaction.

The improvement in problems complexity by the use of forecasting in inventory systems asked the use of simulation and search methods for the definition of near- optimal decisions. Examples of it can be seen in Fu and Healy (1997) and Lopez-Garcia and Posada-Bolivar (1999).

2.1. Forecasting Models

All forecasting models selected use the smoothing logic. Table 1 presents the 4 types and the projection equations for forecast calculation.

Table 1: Forecast models and corresponding formulas

Model Curve	Formula
Constant	$D_t = Q + \xi\sigma$
Trend	$D_t = Q_{to} + I.(t-t_o) + \xi\sigma$
Cyclical with constant	$D_t = Q.S_t + \xi\sigma$
Cyclical with trend	$D_t = [Q_{to} + I.(t-t_o)].S_t + \xi\sigma$
to = initial date St = seasonality coefficient of period t Q = constant Dt = demand forecast of period t I = slope t = period	

These models are analyzed in details in Makridakis and Wheelright (1998) and Hanke and Reitsch (1998). Based on them, smoothing methods are used to quantify the Q, I and F parameters, balancing the importance of old and new demand data in the forecasts.

Table 2: Forecast models and corresponding smoothing parameters and projection formulas

Model Curve	Smoothing Method	Projection
Constant	Simple with α	$PD_{t+k} = Q_t$
Trend	Simple with α and β *(Holt)*	$PD_{t+k} = Q_t + k.I_t$
Cyclical with constant	Simple with α and γ *(Winter)*	$PD_{t+k} = Q_t.S_{t+k-nL}$
Cyclical with trend	Simple with α, β and γ *(Winter)*	$(Q_t + k.I_t).S_{t+k-L}$

Table 3: Formulas used in the period-to-period calculations of Q, I and S parameters

Curve	Formula
Constant	$Q_t = \alpha V_t + (1-\alpha)Q_{t-1} = Q_{t-1} + \alpha(V_t - Q_{t-1})$
Trend	$Q_t = \alpha V_t + (1-\alpha)(Q_{t-1} + I_{t-1})$ $I_t = \beta(Q_t - Q_{t-1}) + (1-\beta)I_{t-1}$
Cyclical with constant	$Q_t = \alpha(V_t/S_{t-L}) + (1-\alpha)Q_{t-1}$ (L=cycle in periods) $S_t = \gamma(V_t/Q_t) + (1-\gamma)S_{t-L}$
Cyclical with trend	$Q_t = \alpha(V_t/S_{t-L}) + (1-\alpha)(Q_{t-1} + I_{t-1})$ (L=cycle in periods) $I_t = \beta(Q_t - Q_{t-1}) + (1-\beta)I_{t-1}$ $S_t = \gamma(V_t/Q_t) + (1-\gamma)S_{t-L}$

Table 2 presents the smoothing parameters and projection formula for the models. At each period, Q, I and S are recalculated, using the real demand occurred (V) and Table 3 formulas. The models adjusted to the item demand series allow the selection of the corresponding α, β and γ for each of them, using the search and simulation routine. Again, Makridakis and Wheelright (1998) and Hanke and Reitsch (1998) provide the optimization method, using 12 initial periods to calculate the initial Q, I and S values and the last 24 periods to adjust parameters to the minimum error, measured by mean absolute deviation - MAD.

2.2. Inventory Models
Following, the inventory models selected for this study are presented.

2.2.1. Periodic Up to Order Level model (UO)
In this model, the lot size is calculated to ensure that the total stock quantity never exceeds a defined quantity, the Order Level (ol) parameter. The decision rule is:

$$poq = (ol - st) \quad \text{if} \quad st \leq rl \tag{1}$$

$$poq = 0 \ (\text{zero}) \quad \text{if} \quad st > rl \tag{2}$$

where
poq is the Purchase Order Quantity, the lot size to replenish the inventory
ol is the Order Level parameter
st is the current stock at the decision moment, including the purchased lots still arriving
rl is the Reorder Level, the quantity parameter that starts the purchase activity.
The periodic review occurs at the end of each period.

2.2.2. Periodic Base Stock model (BS)
The decision rule for this model is
$$poq = (ol - st) \tag{3}$$

This simple rule implies stock replenishment every time a withdraw occurs. It is a particular case of the precedent model, where the Reorder Level is equal to the Order Level parameter.

Like the other, it is a model with a one-period review.

2.2.3. Periodic Economic Order Quantity model (EQ)
This traditional model has the following decision rule

$$poq = n.ls \quad \text{if} \quad st \leq rl \tag{4}$$

$$poq = 0 \ (\text{zero}) \quad \text{if} \quad st > rl \tag{5}$$

where
ls is the fixed lot size, usually calculated using some economic criteria
n is a number of lots (LS sized) that assure the st>=rl after this purchase
Again, the replenish decision occurs at every 1 period.

2.2.4. Requirements Planning Model (RP)
This active model uses the following decision rule:

$$poq = nr \quad \text{if} \quad nr > 0 \ (\text{zero}) \tag{6}$$

$$poq = 0 \ (\text{zero}) \quad \text{if} \quad nr \leq 0 \ (\text{zero}) \tag{7}$$

where
nr is the Net Requirements needed to satisfy the forecast demand of the time defined by the Lead Time plus Reviewing Period and also maintains a safety stock.
The general NR formula is:

$$nr_{t,t+lt} = \sum_{i=1}^{lt+rp} d_{t,t+i} - \sum_{j=1}^{lt-1} poq_{t+i-lt,t+i} - s_t + ss \tag{8}$$

where
nr is the Net Requirements need to satisfy the demand during the next LT + RP periods
t is the system date when the decision is taken
lt is Lead Time of the item
rp is the Reviewing Period
d is the Demand forecast
poq is the Purchase Order Quantity
s is the stock of items
ss is the Safety Stock
The first key in the parameters is the date when the system takes the decision, which generates the value parameter. The second key refers to the deadline to that decision.

The following example in Figure 1 helps to understand the model operation (suppose *ss* = 10).
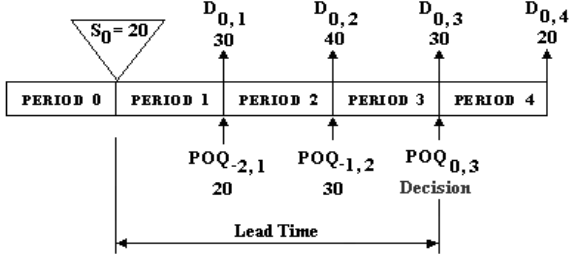
The Net Requirements value in this example is

$$nr = (30 + 40 + 30 + 20) - (20 + 30) - 20 + 10 = 60 \qquad (9)$$

The Order decided in the current t = 0 needs to satisfy the next 4 periods, because the lot will only arrive at the end of period 3, to be used in period 4. To order only the necessary quantity, the *poq* decided in the past (and not received) and the current stock is subtracted from the initial sum. Finally, a safety stock is added to prevent shortages due to forecasting errors. Therefore, this model always tries to end the last period with only the Safety Stock in the inventory. Safety Stock is the single parameter for this pure form model.

2.2.5. Minimum Net Requirements (MNR)
The Base Stock models and the Requirements Planning Model tend to order with low quantities due to their decision rules. This is a great difficulty in environments where the purchase cost is high.

To improve the competitiveness of these 2 models, a second parameter is introduced, the Minimum Net Requirements. It is the smallest quantity the system must order when a quantity different from zero is needed. If a quantity greater than MNR is needed, the model orders this quantity instead of MNR. The new decision rules for Periodic Base Stock and Requirements Planning Model are, respectively:

$$poq = (ol - st) \quad \text{if} \ (ol - st) \geq mnr \qquad (10)$$

$$poq = 0 \qquad \text{if} \ (ol - st) < mnr \qquad (11)$$

$$poq = nr \qquad \text{if} \ nr \geq mnr \qquad (12)$$

$$poq = mnr \qquad \text{if} \ 0 < nr < mnr \qquad (13)$$

$$poq = 0 \ (\text{zero}) \quad \text{if} \ nr \leq 0 \, (\text{zero}) \qquad (14)$$

These modified rules were used in all simulations of this study.

On a mathematical approach, the parameter reorder level (*ro*) used in UO and RP models would be negative in cases of low minimal service level allowed. In the real word we do not see negative parameters probably by the difficulty of operation with the models.

This work analyses the different performance of the UO and RP models in case of negative parameters allowed.

3. MODELLING
Figure 2 presents the experimental design for models comparison. Actual historical data series of 160 items, each one with 36 periods, are used to compare the inventory models. First, this demand data set was used to select the best between the 4 forecasting models, which provided the demand forecast in the active inventory model. The decision variable was the forecasting error, measured by the Mean Absolute Deviation (MAD), calculated for the last 24 periods of each item. The initial 12 periods were used to calculate the forecast parameters seeds required in the simulation and search routine. Then, the 4 inventory model adjusted to revision period equal to 1 were analysed using the same demand data set.
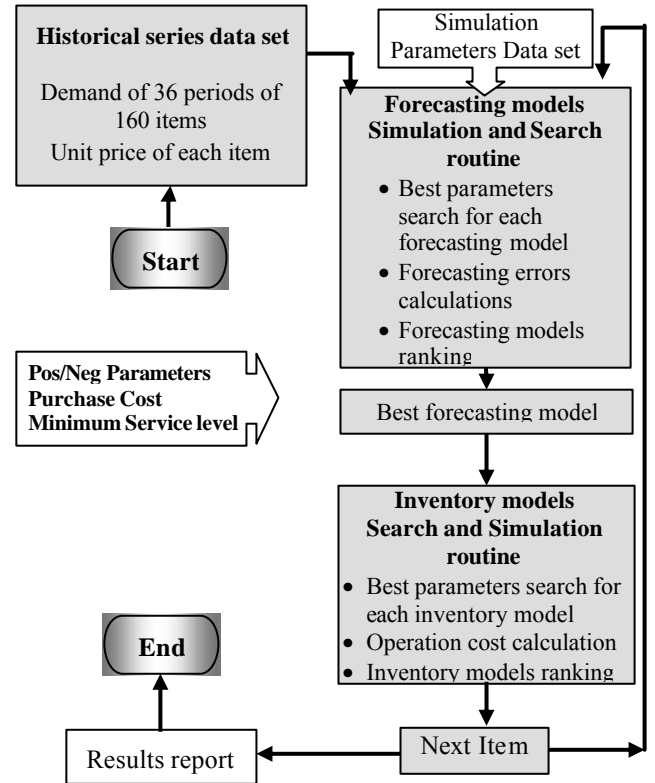
Figure 2: Experimental design for the study

The modeling of the inventory systems uses a multistart neighborhood search. Each search corresponds to a pair of parameters of the inventory model tested. A simulator is used as a function to calculate de search function which, in this case, is a cost function, and a service level restriction is considered. Each simulation runs considering all the 36 months of the planning horizon. The simulator and search routine is fully parameterized with respect to number of restart (5 in this work), search resolution values (3 values of 0.10, 0.05 and 0.02) and termination criteria (3).

The initial seed values of the parameters is obtained by the traditional calculation present in the inventory theory and a routine modify these values founding new feasible seeds with best performance than the originally ones.

Since the cost function in this routine is the average operation cost of the planning horizon, a set of unit costs was supplied to the software:

- Purchase cost – cost incurred each time a purchase order is decided - 2 values were used in the experiment: 9 (high level), and 3 (low level) for all items;

- Holding cost – cost incurred in holding one unit of the item during 1 period – this cost was equal to 1% of the item acquisition price, informed to the system in the simulation parameters data set. The current items prices are used in this study.

Once a minimum service level is considered, the shortage cost can be disconsidered. The average operation cost was the sum of average purchase and holding costs, considering only the periods the model was running on phase.

Table 4 presents the 16 combinations of range of the parameters: positive/negative parameters allowed, purchase costs, and service levels, resulting in 16 comparative simulation runs.

Table 4: Parameters values used in the simulation runs

| Run | Positive | | Purchase | | Service Level | | | |
	Y	N	9	3	78	84	90	96
1	o		o		o			
2	o		o			o		
3	o		o				o	
4	o		o					o
5	o			o	o			
6	o			o		o		
7	o			o			o	
8	o			o				o
9		o	o		o			
10		o	o			o		
11		o	o				o	
12		o	o					o
13		o		o	o			
14		o		o		o		
15		o		o			o	
16		o		o				o

A sensibility analysis was made considering the 4 different Service Levels: 78 %, 84%, 90% and 96%. Combination of all tested parameters conditions permits a good analysis of the system efficiency and stability.

Visual Basic Applications software and Excel worksheets were the basic tools used for all routines.

4. EXPERIMENTAL RESULTS

The 16 experimental runs were conducted in an Intel Core 2 Duo 1.8 GHz, 2 Gb RAM computer. Each run took approximately 3 minutes. The main results were the inventory models ranking for each item in each run. The model with the lowest Average Operation Cost was ranked in the first place, the model with the next lowest

Average Operation Cost was ranked in the second place, and so on.

At the end of each run, 160 first places were obtained. Then, a percentage of first places for each inventory model at each run were calculated. These percentages are presented in tables 5 to 9, grouped by parameters values to permit better analysis. The greatest percentage in each run is shaded to identify the best performance model.

Table 5 shows the results of runs where only positive values are permitted for parameters *ro* and *ss*. Clearly, the EQ model has the best performance for lower service levels, losing its advantage when the service level grows. RP model takes the first position at the highest service level value, and this happens in both Purchase Cost levels tested, but with more emphasis in the lower PC level. The other 2 simulated models show similar EQ behaviour when service level grows.

Table 5: Percentage of First Places with Positive Parameters Only

Positive Parameters Only						
Run	PC	SL	UO	BS	EQ	RP
1	9	78	18.8%	23.8%	41.3%	16.3%
2	9	84	20.0%	18.1%	42.5%	19.4%
3	9	90	20.0%	11.3%	41.9%	26.9%
4	9	96	13.8%	17.5%	33.1%	35.6%
5	3	78	20.0%	20.6%	36.3%	23.1%
6	3	84	15.6%	13.1%	38.8%	32.5%
7	3	90	10.0%	10.0%	40.6%	39.4%
8	3	96	11.3%	10.6%	28.1%	50.0%

Table 6: Percentage of First Places with Positive and Negative Parameters

Positive and Negative Parameters						
Run	PC	SL	UO	BS	EQ	RP
9	9	78	16.9%	18.1%	26.9%	38.1%
10	9	84	13.8%	12.5%	32.5%	41.3%
11	9	90	15.0%	9.4%	30.0%	45.6%
12	9	96	11.3%	12.5%	24.4%	51.9%
13	3	78	16.9%	16.3%	24.4%	42.5%
14	3	84	14.4%	10.6%	25.0%	50.0%
15	3	90	10.6%	8.1%	27.5%	53.8%
16	3	96	8.8%	8.8%	25.0%	57.5%

The permission for negative *ro* and *ss* values generates the runs presented in Table 6. RP model has the best performance in all runs, improving the first places percentage with the SL growth. The negative values empowered the RP, but not UO and EQ models. Analysis of differences between 8 and 16 runs in RP model, for example, show 5,2% operation cost reduction in 54 items with negative *ss*.

Tables 7 and 8 present the same data of the precedent tables, arranged by Purchase Cost level. RP and EQ models present similar performance in both levels, with the expected better RP performance in the low Purchase Cost level. Again, the permission for negative values and the growth in the minimum Service Level give big advantage to the RP model.

Table 7: First Places Percentages - Purchase Cost = 9

Purchase Cost = 9					
Run	PPO	SL	UO	BS	EQ	RP
1	Y	78	18.8%	23.8%	41.3%	16.3%
2	Y	84	20.0%	18.1%	42.5%	19.4%
3	Y	90	20.0%	11.3%	41.9%	26.9%
4	Y	96	13.8%	17.5%	33.1%	35.6%
9	N	78	16.9%	18.1%	26.9%	38.1%
10	N	84	13.8%	12.5%	32.5%	41.3%
11	N	90	15.0%	9.4%	30.0%	45.6%
12	N	96	11.3%	12.5%	24.4%	51.9%

Table 8: First Places Percentages - Purchase Cost = 3

Purchase Cost = 3					
Run	PPO	SL	UO	BS	EQ	RP
5	Y	78	20.0%	20.6%	36.3%	23.1%
6	Y	84	15.6%	13.1%	38.8%	32.5%
7	Y	90	10.0%	10.0%	40.6%	39.4%
8	Y	96	11.3%	10.6%	28.1%	50.0%
13	N	78	16.9%	16.3%	24.4%	42.5%
14	N	84	14.4%	10.6%	25.0%	50.0%
15	N	90	10.6%	9.4%	27.5%	53.8%
16	N	96	9.4%	9.4%	25.0%	57.5%

Table 9: Percentage of First Places with different Service Levels

Run	PPO	PC	SL	UO	BS	EQ	RP
1	Y	9	78	18.8%	23.8%	41.3%	16.3%
5	Y	3	78	20.0%	20.6%	36.3%	23.1%
9	N	9	78	16.9%	18.1%	26.9%	38.1%
13	N	3	78	16.9%	16.3%	24.4%	42.5%
2	Y	9	84	20.0%	18.1%	42.5%	19.4%
6	Y	3	84	15.6%	13.1%	38.8%	32.5%
10	N	9	84	13.8%	12.5%	32.5%	41.3%
14	N	3	84	14.4%	10.6%	25.0%	50.0%
3	Y	9	90	20.0%	11.3%	41.9%	26.9%
7	Y	3	90	10.0%	10.0%	40.6%	39.4%
11	N	9	90	15.0%	9.4%	30.0%	45.6%
15	N	3	90	10.6%	9.4%	27.5%	53.8%
4	Y	9	96	13.8%	17.5%	33.1%	35.6%
8	Y	3	96	11.3%	10.6%	28.1%	50.0%
12	N	9	96	11.3%	12.5%	24.4%	51.9%
16	N	3	96	9.4%	9.4%	25.0%	57.5%

Finally, table 9 shows the data arranged by Service Level. The same pattern (EQ better when negative values aren't permitted and RP when it is possible) is present in all levels except in the highest 96% SL value, where RP model wins in all combinations.

5. CONCLUSION

The results indicate the best overall performance of the Periodic Economic Order Quantity (EQ) and Requirements Planning (RP) models. The EQ prevails when low minimum Service Levels (SL) and positive values are used. Its advantage decreases with the minimum Service Level growth and disappears when the RP Safety Stock parameter (*ss*) is free to receive negative values. The same flexibility in the UO and EQ Reorder Point parameter (*ro*) results in no performance improvements.

REFERENCES

Agarwall, S.C., 1974. A review of Current Inventory Theory and its Applications. *International Journal of Operations Research* v 12: 443-482.

Armstrong, J. S., Jan. Feb. 1986. Research on forecasting: a quarter-century review, 1960-1984. *Interfaces* v 16: 52 – 66.

Arrow, K.A., Harris, T.E.,. Marschak, J., 1958. Optimal inventory policy. *Econometrica* v 19: 250-272.

Brown, R.G., 1959. *Statistical Forecasting for Inventory Control.* New York: McGraw-Hill.

Brown, R.G., 1967. *Decision rules for inventory management.* New York: Holt, Rinehart and Winston.

Browne, S., Zipkin, P., 1991. Inventory models with continuous, stochastic demands. *The Annals of Applied Probability* v. 1: 419-435.

Dvoretzky, A., Kiefer, J., 1952. The inventory problems: I. Case of known distributions of demand; II. Case of unknown distributions of demand. *Econometrica* v. 20: 187-222 and 450-466.

Eilon, S., Elmaleh, J., 1970. Adaptive Limits in Inventory Control. *Management Science* v. 16-8: 533-548.

Fildes, R., Beard, C., 1992. Forecasting Systems for Production and Inventory Control. *International Journal of Operation & Production Management* v. 12-4: 4-27.

Fu, M. C., Healy, K. J., 1997. Techniques for optimization via simulation: an experimental study on an (s,S) inventory system. *IIIE Transaction* v. 29: 191-199.

Gardner Jr., E. S., April 1990. Evaluating Forecast Performance in an Inventory Control System. *Management Science* v. 36-4.

Goddard, W. E., Sept. 1989. Let's scrap forecasting. *Modern Materials Handling* v.39.

Hadley, G., Whitin, T. M., 1963. *Analysis of inventory systems.* New Jersey: Prentice Hall.

Hanke, J. E., Reitsch, A. G., 1998. *Business Forecasting*. New Jersey: Prentice Hall.

Hax, A., Candea, D., 1984. *Production and inventory management.* New Jersey: Prentice Hall.

Lee, T. S., Adam Jr., E. E., 1986. Forecasting error evaluation in material requirements planning (MRP) production-inventory systems. *Management Science* v. 2: 1186–1205.

Lopez-Garcia, L., Posada-Bolivar, A., 1999. A simulator that uses Tabu search to approach the optimal solution to stochastic inventory models. *Computers & Industrial Engineering* v. 37 Issues 1-2: 215-218.

Makridakis, S., Wheelright, S. C., 1998. *Forecasting-Methods and Applications.* New York: Wiley.

Naddor, E., 1966. *Inventory systems.* New York: Wiley.

Orlicky, J., 1995. *Material Requirements Planning*. New York: McGraw-Hill.

Sipper, D., Bulfin, R., 1998. *Production planning, control, integration*. New York: McGraw-Hill.

THE MODELLING AND CONTROL OF THE AGRICULTURAL SET DRIVER

Jacek KROMULSKI[a], Tadeusz PAWŁOWSKI[b]

[a]Industrial Institute of Agricultural Engineering, Poznan, Poland
[b]Industrial Institute of Agricultural Engineering, Poznan, Poland

[a]kromulsk@pimr.poznan.pl, [b]tadek@pimr.poznan.pl

ABSTRACT

This paper presents a driver model based upon the control model of the human operator. Vehicle driven by operator may be considered as control system with driver as a controller and vehicle as controlled system. Driver efficiency depends on situation on the road, vehicle characteristic and psychical and physical driver features. The state space model of human operator formulation uses Pade approximations to model the time delays inherent in the human operator. Driver is included in road-vehicle driver system as feedback element. Parameters describing drives behaviour could be identified based on system model in frequency domain (from transmissibility) or in time domain (from correlation). The paper presents initial conception taking into consideration driver working in the tractor – agricultural machine set model.

Keywords: control theory, vehicle dynamics, vehicle steerability control, prognostic simulation

1. INTRODUCTION

Vehicle dynamics is one of the most important fields in research of motorization engineering. Running, acceleration, braking, changing of motion direction are typical dynamic processes. Quality of their execution depends on a driver.

Vehicle driven by driver may be considered as control system with driver as a controller and vehicle as controlled system. Driver efficiency depends on situation on the road, vehicle characteristic and psychical and physical driver features. From a linear input-output system viewpoint, the vehicle may be considered a generalized plant with the driver acting as the feedback control element. The driver regulates the system outputs (lateral velocity and yaw rate) in order to follow a desired roadway path. A main objective of the control of vehicle dynamics is to improve the handling performance, or maneuverability, in order to obtain safer and more joyful driving. Model is applied to tractor – potato planter set.

2. STATE SPACE MODEL OF THE HUMAN OPERATOR

Optimal control models of the human operator assume that the operator behaves in an "optimal" manner subject to human limitations.

The general vector-matrix form of the state space model for Multi-Input Multi-Output (MIMO) operator models with a mild restriction on the form of the dynamical model of the neuromuscular system is given by:

$$\dot{x} = Ax + B\delta + Ew$$

$$y = Cx + D\delta \qquad (1)$$

where:

x(t) is a state vector composed of both plant and system disturbance states,

δ (t) is a plant input (is a vector of the operator inputs to the controlled element),

w(t) is a disturbance vector modeled as a zero mean Gaussian white noise process with intensity W,

y(t) is a output vector containing the parameters that the operator is trying to control (e.g. tracking error, error rate).

The modified optimal control models formulation uses Pade approximations to model the time delays inherent in the human operator.

There are as many Pade delays as there are operator inputs to the controlled element. The delays are placed at the operator's outputs and are considered a part of the plant dynamics for the purpose of synthesis. A second-order Pade approximation is chosen because it provides a very good approximation to a pure delay over the operator's frequency range of interest (approximately 0.1 to 10 rad/sec). Use of at least a second-order Pade approximation is assumed to be necessary to accurately model operator magnitude and phase compensation at the high end of the operator's bandwidth, such as the operator high frequency neuromotor resonant peak .

A second-order Pade approximation is given by (Doman 1998).:

$$\frac{u_d}{u_p} = \frac{1 - \frac{1}{2}(\tau s) + \frac{1}{8}(\tau s)^2}{1 + \frac{1}{2}(\tau s) + \frac{1}{8}(\tau s)^2} \qquad (2)$$

where:

τ is the delay interval (in seconds),

u_p and u_d are the undelayed and delayed operator inputs to the controlled element, respectively.

In state space form, this can be expressed by:

$$\dot{x}_d = A_d x_d + B_d u_p$$

$$\delta = u_d = C_d x_d + u_p \qquad (3)$$

where

x_d is a two-element vector of Pade delay states.

The undelayed operator inputs do not physically exist as measurable quantities but they are used as an intermediate step in synthesizing the operator model. Augmenting the controlled element and disturbance with the delay dynamics yields:

$$\begin{bmatrix} \dot{x} \\ \dot{x}_d \end{bmatrix} = \begin{bmatrix} A & BC_d \\ 0 & A_d \end{bmatrix} \begin{bmatrix} x \\ x_d \end{bmatrix} + \begin{bmatrix} B \\ B_d \end{bmatrix} u_p + \begin{bmatrix} E \\ 0 \end{bmatrix} w$$

$$y = \begin{bmatrix} C & DC_d \end{bmatrix} \begin{bmatrix} x \\ x_d \end{bmatrix} + D u_p \qquad (4)$$

or, with appropriate definitions,

$$\dot{x}_s = A_s x_s + B_s u_s + E_s w$$

$$y = C_s x_s + D_s u_p \qquad (5)$$

You can easily convert any state-space model to transfer function form (Ljung and Glad 1994).

$$y(s) = C(sI - A)^{-1} B u(s) \qquad (6)$$

3. CONTROL MODEL OF THE HUMAN OPERATOR

The mathematical model of vehicle takes into consideration driver activity consists of two main parts: driver model and vehicle model (Doman 1998). The connections between road, driver and driven vehicle are shown on fig, 1. A set of mathematical equations in the analytical model is used to predict the important

operating parameters of the vehicle's steerability (Schafer 2004). Identification of driver model parameters is achieved through an optimal control approach. (Zhaoheng 2007). Parameters describing drives behaviour could be identified based on system model in frequency domain (from power spectral density) or in time domain (from correlation).
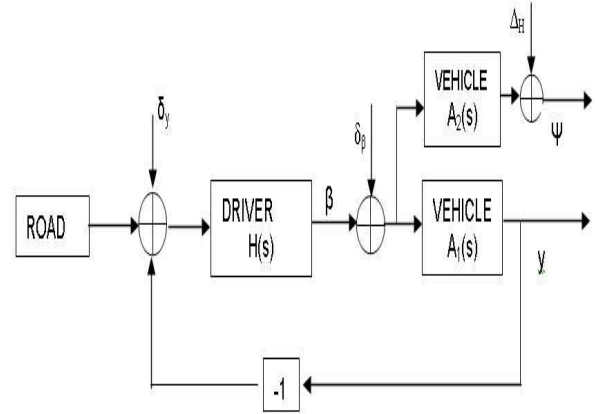
Fig. 1. Simplified diagram of road – driver -vehicle system with feedback

Steering angle $\beta(t)$, vehicle yaw angle $\Psi(t)$ and its lateral displacement y are specified by system of equation (7) (with initial condition $y=0$).

$$\begin{cases} \beta = H * \delta_y - H * y \\ y = A_1 * \delta_\beta + A_1 * \beta \\ \Psi = A_2 * A_1^{-1} * + \Delta_H \end{cases} \qquad (7)$$

It is possible to obtain dependences characterized transmittance H of driver model on the basis of above system of equation and spectral density and reciprocal spectral density of measured signals: $\delta_y * \delta_y^*$, $\delta_\beta * \delta_\beta^*$, $\Delta_H * \Delta_H^*$, $\delta_\beta * \delta_y^*$, $\Delta_H * \delta_y^*$.

The transmittance A_1 and A_2 are depending of physical parameter of the vehicle only. It is possible to estimate the transmissibility function or autocorrelation function of a driver depending on driver response time delay (τ) and constant parameters depending on driver's reaction time (Pawłowski 2008).

4. IDENTIFICATION OF DRIVER MODEL FROM EXPERIMENTAL DATA

The data used for system identification was collected from a series of in a typical operational condition driving scenario. The transmittances H and autocorrelations k of driver for different driver and vehicle way have been investigated.

The Power Spectral Density for steering angle $\beta(t)$ can also be expressed as equation (8):

$$G_{\beta\beta} = G_{\delta_\beta}(1 + b^2 T_\beta^2 \omega^2)/(1 + T_\beta^2 \omega^2)^3 \qquad (8)$$

where:

$$b = T_\beta^{-1}(\tau + 3T_\beta + \sqrt{2}T_\delta)$$

The Wiener-Khintchine theorem states that the autocorrelation function is given by the Inverse Fourier Transform of the Power Spectral Density (PSD).
The autocorrelation function can be defined by:

$$k_{\beta\beta}(\theta) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G_{\beta\beta}(\omega) e^{i\omega\theta} d\omega \qquad (9)$$

An expression for the normalized autocorrelation function $\rho_{\beta\beta}(\theta) = k_{\beta\beta}(\theta)/k_{\beta\beta}(0)$ for steering angle β(t) is derived below (equation 10):

$$\rho_{\beta\beta}(\theta) = e^{-\left|\frac{\theta}{T_\beta}\right|}\left[1 + \left|\frac{\theta}{T_\beta}\right| + \frac{1-b^2}{3+b^2}\left(\left|\frac{\theta}{T_\beta}\right|\right)^2\right] \qquad (10)$$

Analysis of the control system was carried out using MATLAB analysis software. The optimization method was carried out for identification time parameters and reaction time delay τ from autocorrelation using Nelder-Mead simplex algorithm.

Figure 2 show the plots of normalized autocorrelation functions of steering angle β during field works for three different drivers.
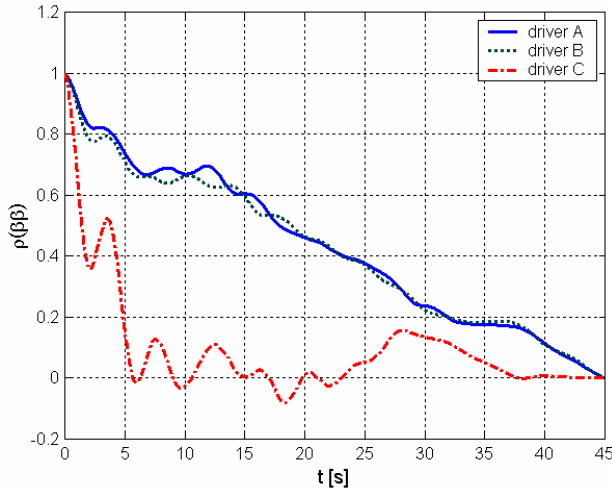
Fig. 2. Correlogram comparison for different drivers: A, B and C

Calculated for individual drivers reaction time τ using Nelder-Mead simplex algorithm was different. The driver reaction time for individual drivers ranged from 0.12 (driver A) to 0.43 (driver C) seconds.

5. PROGNOSTIC SIMULATION

After parametric identification we approach to research of system sensitivity influence on change of parameters with application of prognostic simulation methods.

An important element for project engineer is examining the influence of many design parameters like circles base, rigidity of tires and - the most important - driver perception of dynamic vehicle condition reception (e.g. gradient of acceleration of seat) and proper reaction of driver related with this.

6. CONCLUSIONS

The paper presents initial conception taking into consideration driver working in the tractor – agricultural machine set model. Vehicle driven by driver may be considered as control system with driver as a controller and vehicle as controlled system.

Parameters describing drives reaction time could be identified based on control system model in frequency domain (from trans-missibility) or in time domain (from correlation). Conducted tests proved that proposed model describes well tractor – potato planter combination steerability. The author is of the opinion that the model may be base for next research and specifying.

REFERENCES

Doman, D., 1998. *Projection Methods for Order Reduction of Optimal Human Operator Models*, PhD Dissertation, Virginia Polytechnic Institute and State University..

Ljung, L., Glad, T., 1994. *Modeling Of Dynamic Systems*, Englewood Cliffs, New Jersey, Prentice Hall.

Pawłowski, T., Kromulski, J., Szczepaniak, J., 2008. Influence of driver perception on safety of movement of agricultural machines, *XXI Conference "Problemy Rozwoju Maszyn Roboczych"*, 449-455, 21-24.01.2008, Zakopane.

Schafer, F.R., 2004. *Generalized Feedback Control and Application to Vehicle Path Following Control*, PhD Dissertation, TU Berlin.

Zhaoheng, L., 2007. Characterisation of optimal human driver model and stability of a tractor-semitrailer vehicle system with time delay, *Mechanical Systems and Signal Processing*, 21: 2080–2098.

AUTHORS BIOGRAPHY

Tadeusz PAWŁOWSKI is the Director of the Industrial Institute of Agricultural Engineering in Poznan. He received his MS (1979) and the Ph.D. (1999) in Engineering from Poznan University of Technology (Faculty of Working Machines and Transportation).

His main research areas include agricultural engineering, modelling, simulation, signal processing and finite element modeling. During his career, Dr. Pawłowski has authored or co-authored over 120 technical papers.

Dr. Pawłowski is also a designer of agricultural machines awarded in international fairs. For example, for device for opening the pair of side arms of the multi-

segment land of roller he achived a gold medal of The Belgian and International Trade Fair for Technological Innovation in 2007 and a silver medal of The 107th International Invention Salon "Concours – Lepine" in 2008.

Jacek KROMULSKI is an assistant professor at the Industrial Institute of Agricultural Engineering in Poznan. He received his MS (1979) in physics from the Adam Mickiewicz University Poznań. In 1998 he received the Ph.D. in physics (Adam Mickiewicz University, Institute of Acoustics) with a thesis titled "Investigation of dynamic characteristics of vibration of acoustic sources with complex mechanical structure by means of modal analysis". His main research areas include analytical and experimental modal analysis, structural dynamics, modelling, simulation and signal processing. During his career, Dr. Kromulski has authored or co-authored over 80 technical papers.

WAREHOUSE AND INTERNAL LOGISTICS MANAGEMENT BASED ON MODELING & SIMULATION

Enrico Bocca [a], Duilio Curcio[b], Francesco Longo[c], Alberto Tremori[d]

[a] [d] Liophant Simulation
University of Genoa
Via Molinero, Savona, 17100, ITALY

[b] [c] Modeling & Simulation Center - Laboratory of Enterprise Solutions (MSC – LES)
M&S Net Center at Department of Mechanical Engineering
University of Calabria
Via Pietro Bucci, Rende, 87036, ITALY

[a]enrico.bocca@liophant.org, [b] [c] {dcurcio, f.longo}@unical.it, [d]alberto.tremori@liophant.org

ABSTRACT
The paper focuses on warehouse and internal logistics management. The objective of the paper is to investigate the effect of some critical parameters (i.e. the number of incoming trucks from suppliers, the number of outgoing trucks to retailers, the number of forklifts and lift trucks, etc.) on the warehouse internal logistic costs. To this end the authors develop a parametric simulator equipped with a graphic user interface (for changing the most important warehouse parameters) that the user can use for performing what-if analysis and scenarios investigation. An application example is finally proposed: after a sensitivity analysis, the analytical relationship between the internal logistic costs and the warehouse critical parameters is evaluated.

Keywords: Warehouse Management, Internal Logistics, Simulation, DOE, ANOVA

1. INTRODUCTION
During the last years number of research works has been developed in the sector of warehouse management and internal logistics planning and control. In particular, researches in this field are favored by the continuous development of computer technology and new materials handling equipment. Planning activities within a warehouse concern with goods assignment in each storage location and warehouse system design. Control problems are related to storage sequencing and scheduling optimization, retrieval requests in the dispatching control. Recent research studies also regard data/information management in warehouse systems.

Mason et al. (2003) investigate the integration between the warehouse management system (*WMS*) and the transportation management system (*TMS*) of an industrial company. In particular, the WMS contains information on supplier/customer warehouse inventory level and about customers' orders while the TMS stores all the information about retailers' locations, items to deliver and vehicles to adopt.

Eben-Chaime and Pliskin (1997) investigate the effect of operations management tactics on performance measures of automatic warehousing systems with multiple machines.

Moreover, several researches focalize on warehouse system design (i.e. Hsieh and Tsai, 2006 implement a simulation model for finding the optimum design parameters of a real warehouse system).

The paper investigates how some warehouse critical parameters affect the logistic internal costs. To this end, the authors use as support tool a simulation model capable of recreating the high complexity of a real warehouse.

Before getting into details of the study, a brief overview of the paper is proposed. Section 2 reports the description of the warehouse. Section 3 describes the implementation of the simulation model. Section 4 investigates the relationship between the internal logistic costs and the warehouse management parameters. The last section reports the conclusions that summarize critical issues and results of the paper.

2. WAREHOUSE DESCRIPTION
This research work proposes a study on the relationship between the warehouse management parameters and the effect of such parameters on the internal logistics costs.

The warehouse being considered in this paper is a real warehouse (supporting the retail sector) with the following characteristics:

- a surface of 13000 m^2;
- a shelves' surface of 5000 m^2;
- 3 levels of shelves;
- a capacity in terms of pallets of 28400 pallets;
- a capacity in terms of pallets for each product of 7100 pallets;

- a capacity in terms of packages of about one million packages.

3. THE WAREHOUSE SIMULATION MODEL

According to Eben-Chaime et al. (2004), simulation is the most effective tools for designing and analyzing manufacturing systems and, more in detail, for making warehousing context analysis. In fact, one of the most important advantages of Simulation is to explore and experiment possibilities for evaluating system behavior under internal/external changes.

In particular, the warehouse simulation model implemented by the authors has the objective to investigate the relationship between input (warehouse management parameters) and output (warehouse performance measures) parameters. To this end the simulator is characterized by high flexibility in terms of parameters variation and scenarios definition.

3.1. The simulation model architecture

The simulation model presented in this research work reproduces all the most important processes and operations of the real warehouse related to:

- trucks arrival and departure for items deliveries (from suppliers to the warehouse and from the warehouse to retailers);
- forklift and lift trucks for materials handling operations
- performance measures control and monitoring (number if items handled, waiting times for trucks, suppliers' and retailers' service levels).

The software tool adopted for the simulation model implementation is the commercial package Anylogic™ by *XJ Technologies*.

In particular, for reproducing each process and for increasing model flexibility, different classes have been implemented by using software libraries. Figure 1 displays the structure diagram of the simulation model.
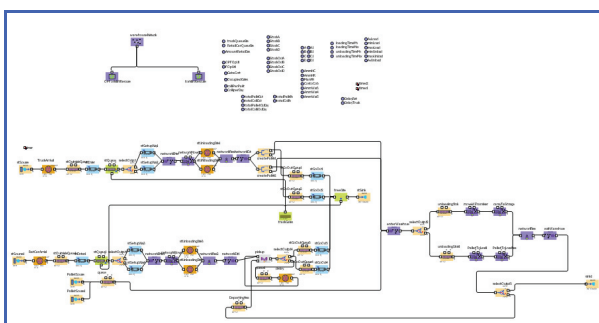
Figure 1: The Structure Diagram of the Simulation Model

3.2. The input parameters

In order to increase the flexibility of the model for scenarios investigation, a number of warehouse management parameters have been implemented within a graphic user interface by using sliding bars as shown in figure 2. Such parameters can be varied both at the beginning of the simulation run and at run-time observing the effect on the warehouse behavior.
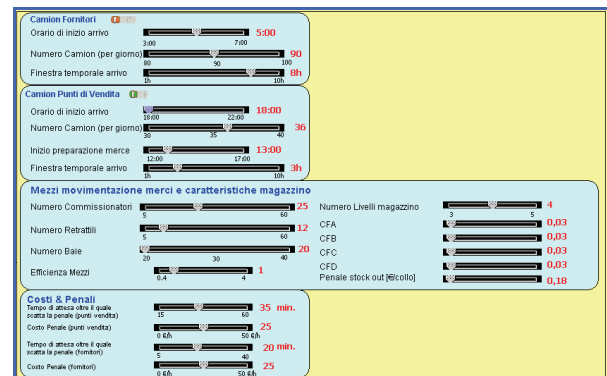
Figure 2: Simulation Model graphic user interface

The Simulation graphic user interface can be subdivided in four different sections:

- *Suppliers' Trucks section*;
- *Retailers' Trucks section*;
- *Warehouse Management parameters section*
- *Logistics Internal Costs section*

The *Suppliers' Trucks section* contains the following parameters:

- the suppliers' trucks arrival time (*TATS*);
- the number of suppliers' trucks per day (*NTS*);
- the time window in which suppliers' trucks deliver products (*TDTS*).

The *Retailers' Trucks section* contains the following parameters:

- the retailers' trucks arrival time (*TATR*);
- the number of retailers' trucks per day (*NTR*);
- the time window in which retailers' trucks deliver products (*TDTR*).
- the time for starting items preparation (*PPT*, items to be delivered to retailers).

The Warehouse Management parameters section contains the following parameters:

- shelves levels (*SL*);
- number of forklift (*NFT*);
- number of lift trucks (*NMT*);
- number of aisles available for loading and unloading operations (*NB*);
- forklifts and lift trucks efficiency (*TPP*);
- stock-out costs parameters (SC_i).

The *Logistics Internal Costs section* contains the following parameters:

- the time after which the warehouse has to pay fines to retailers (*PTR*);

- the time after which suppliers have to pay fines to the warehouse (*PTS*);
- the fine cost for retailers (*FCR*);
- the fine cost for suppliers (*FCS*).

3.3. The output parameters
The output section, see figure 3, provides all the parameters necessary to evaluate and monitor the warehouse performances. The evolution of the performance measure can be observed at run-time.
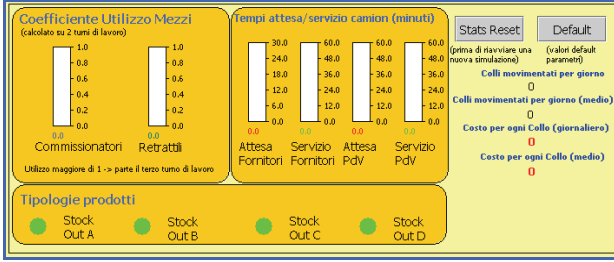
Figure 3: The simulation model Output Section

The performance measures implemented in the simulation model are described as follows:

- The forklifts utilization level (*CFT*);
- the lift trucks utilization level (*CMT*);
- the service level provided to suppliers' trucks (*LSST*);
- the service level provided to retailers' trucks (*LSRT*) ;
- the waiting time of suppliers' trucks before starting the unloading operation (*WTSU*);
- the waiting time of retailers' trucks before starting the loading operations (*WTRL*);
- the packages delivered per day, actual (*PDD*) and average value (*APDD*);
- the daily cost for each package, actual (*DCP*) and average value (*ADCP*).

4. SCENARIOS INVESTIGATION: INPUT OUTPUT RELATIONSHIP
As mentioned into the introduction, the objective of the paper is to investigate the effect of some critical parameters on the warehouse internal logistics costs.

The input parameters (factors) taken into consideration are:

- the number of suppliers' trucks per day (*NTS*);
- the number of retailers' trucks per day (*NTR*);
- the number of forklift (*NFT*);
- the number of lift trucks (*NMT*);
- the number of shelves levels (*SL*).

The variation of such parameters creates different operative scenarios characterized by different resources availability, allocation and utilization. The performance measures being taken into consideration are:

- the packages delivered per day, actual (*PDD*) and average value (*APDD*);
- the daily cost for each package, actual (*DCP*) and average value (*ADCP*).
- The forklifts utilization level (*CFT*);
- the lift trucks utilization level (*CMT*).

The experiments planning is supported by the Design of Experiments (Full Factorial Experimental Design is adopted). Table 1 consists of factors and levels used for the design of experiments.

Table 1: Factors and Levels of DOE

Factors	*Level 1*	*Level 2*
NTS	80	100
NTR	30	40
NFT	6	24
NMT	12	50
SL	3	5

As shown in Table 1, each factor has two levels: in particular, Level 1 indicates the lowest value for the factor while Level 2 its greatest value.

In order to test all the possible factors combinations, the total number of the simulation runs is 2^5. Each simulation run has been replicated three times, so the total number of replications is 96 (32x3=96).

The simulation results have been studied, according to the various experiments, by means of the Analysis Of Variance (*ANOVA*) and of graphic tools.

5. SIMULATION RESULTS ANALYSIS
As before mentioned, the simulation results have been analyzed by means of *ANOVA* and graphic tools. The ANOVA partitions the total variability of the performance indexes in different components due to the influence of the factors reported in Table 1. In this way, it is possible to understand which factors affect the performance indexes, or, in other words, to introduce an analytical relationship (called *meta-model* of the simulation model) between each performance index and the factors being considered.

Let Y_i be the i-th performance measure and let x_i be the factors. The equation 1 expresses the i-th performance measure as linear function of the factors.

$$Y_i = \beta_0 + \sum_{i=1}^{5} \beta_i x_i + \sum_{i=1}^{5} \sum_{j>i}^{5} \beta_{ij} x_i x_j +$$

$$+ \sum_{i=1}^{5} \sum_{j>i}^{5} \sum_{h>j}^{5} \beta_{ijh} x_i x_j x_h +$$

$$+ \sum_{i=1}^{5} \sum_{j>i}^{5} \sum_{h>j}^{5} \sum_{k>h}^{5} \beta_{ijhk} x_i x_j x_h x_k +$$

$$+ \sum_{i=1}^{5} \sum_{j>i}^{5} \sum_{h>j}^{5} \sum_{k>h}^{5} \sum_{p>k}^{5} \beta_{ijhkp} x_i x_j x_h x_k x_p + \varepsilon_{ijhkpn}$$

(1)

where:

- β_0 is a constant parameter common to all treatments;

- $\sum_{i=1}^{5} \beta_i x_i$ are the five main effects of factors;

- $\sum_{i=1}^{5} \sum_{j>i}^{5} \beta_{ij} x_i x_j$ are the ten two-factors interactions;

- $\sum_{i=1}^{5} \sum_{j>i}^{5} \sum_{h>j}^{5} \beta_{ijh} x_i x_j x_h$ represents the three-factors interactions;

- $\sum_{i=1}^{5} \sum_{j>i}^{5} \sum_{h>j}^{5} \sum_{k>h}^{5} \beta_{ijhk} x_i x_j x_h x_k$ are the three four-factors interactions;

- $\sum_{i=1}^{5} \sum_{j>i}^{5} \sum_{h>j}^{5} \sum_{k>h}^{5} \sum_{p>k}^{5} \beta_{ijhkp} x_i x_j x_h x_k x_p$ is the sole five-factors interaction;

- ε_{ijhkpn} is the error term;
- n is the number of total observations.

In particular the analysis carried out by authors aims at:

- identifying those factors that have a significant impact on the performance indexes (sensitivity analysis)
- evaluating the coefficients of equation 1 in order to have an analytical tool capable of expressing the performance measures as function of the most critical factors

In the next section, the authors propose results analysis for the first two performance measure (APDD and ADCP).

5.1. Simulation results analysis for the packages delivered per day (*APDD*)

Table 2 reports the design matrix and the simulation results in terms of number of packages delivered per day. The first four table columns show all the possible combinations of the factors levels while the last column reports the results provided by the simulation model for the first performance parameter (note the last column of table 2 reports the simulation results of one replications; in effect the authors replicated three times each scenario).

The Pareto Chart (see figure 4) of the effects allows to evaluate the predominant effects: in this case they are the first order effects and some effects of the second and third order.

As known from the ANOVA theory the non-negligible effects are characterized by *p-value* ≤α where *p* is the probability to accept the negative hypothesis (the factor has no impact on the performance index) and

Table 2: Design Matrix and simulation results (APDD)

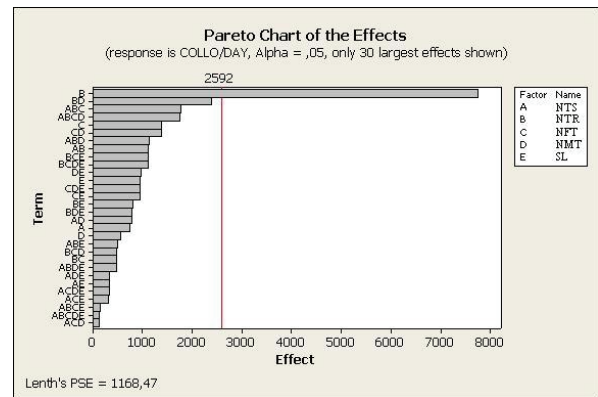| NTS	NTR	NFT	NMT	SL	APDD
80	30	6	12	3	30370
80	30	6	12	5	30345
80	30	6	50	3	30439
80	30	6	50	5	30457
80	30	24	12	3	30421
80	30	24	12	5	30358
80	30	24	50	3	30387
80	30	24	50	5	30488
80	40	6	12	3	40574
80	40	6	12	5	40501
80	40	6	50	3	40603
80	40	6	50	5	40580
80	40	24	12	3	40551
80	40	24	12	5	40568
80	40	24	50	3	40553
80	40	24	50	5	40541
100	30	6	12	3	38528
100	30	6	12	5	37181
100	30	6	50	3	30361
100	30	6	50	5	30399
100	30	24	12	3	30388
100	30	24	12	5	30405
100	30	24	50	3	30416
100	30	24	50	5	30387,6
100	40	6	12	3	35846,1
100	40	6	12	5	37186,2
100	40	6	50	3	40498,8
100	40	6	50	5	40532,1
100	40	24	12	3	40550
100	40	24	12	5	35447,4
100	40	24	50	3	40530
100	40	24	50	5	40563,6

Figure 4: The Pareto Chart for the APDD

α=0.05 is the confidence level used in the analysis of variance. The most significant factors are:

- NTS;
- NTR;
- NFT;
- NMT;
- NTR*NMT;

- NTS* NTR* NFT.

The ANOVA has been repeated for the most important factors, the results are reported in table 3.

- the first column reports the sources of variations;
- the second column is the degree of freedom (*DOF*);
- the third column is the Sum of Squares;
- the 4th column is the Mean Squares;
- the 5th column is the Fisher statistic;
- the 6th column is the p-value.

Table 3: ANOVA Results for the most significant factors

Source	DF	AdjSS (10^{-7})	AdjMS (10^{-7})	F	P
Main Effects	4	50,30	125,75	23,22	0
2-Way interactions	1	45,24	4,52	8,35	0
3-Way interactions	1	24,84	2,48	4,59	0,04
Residual Error	25	13,53	0,54		
Total	31				

Results confirm that factors are correctly chosen because their p-value is lower than the confidence level adopted. The input-output meta-model for the APDD is as follows:

$$APDD = 21777,2 + 21,46 * NTS + 348,74 * NTR +$$
$$-167,083 * NFT - 423,71 * NMT +$$
$$12,51 * (NTR * NMT) +$$
$$+ 0,028 * (NTS * NTR * NFT) \tag{2}$$

Equation 2 is the most important result of the analysis: it is a powerful tool that can be used for correctly defining, in this case, warehouse average number of packages daily delivered in function of the system available resources.

5.2. Simulation results analysis for the average daily cost per package (*ADCP*)

The same analysis has been carried out taking into consideration the average daily cost per packages (*ADCP*). Table 4 reports the design matrix and the simulation results. The normal probability plot allows to evaluate the predominant effects: in this case they are the first order effects and some effects of the second order.

Table 4: Design Matrix and simulation results (ADCP)

| NTS	NTR	NFT	NMT	SL	ADCP
80	30	6	12	3	1,38
80	30	6	12	5	1,33
80	30	6	50	3	0,48
80	30	6	50	5	0,483
80	30	24	12	3	3,06
80	30	24	12	5	3,91
80	30	24	50	3	2,27
80	30	24	50	5	0,623
80	40	6	12	3	1,38
80	40	6	12	5	13,82
80	40	6	50	3	0,45
80	40	6	50	5	11,54
80	40	24	12	3	4,69
80	40	24	12	5	5,3
80	40	24	50	3	3,69
80	40	24	50	5	2,89
100	30	6	12	3	3,05
100	30	6	12	5	4,31
100	30	6	50	3	0,53
100	30	6	50	5	6,72
100	30	24	12	3	5
100	30	24	12	5	6,28
100	30	24	50	3	0,64
100	30	24	50	5	0,62
100	40	6	12	3	3,72
100	40	6	12	5	8,18
100	40	6	50	3	1,06
100	40	6	50	5	8,97
100	40	24	12	3	2,7
100	40	24	12	5	11
100	40	24	50	3	0,48
100	40	24	50	5	0,47
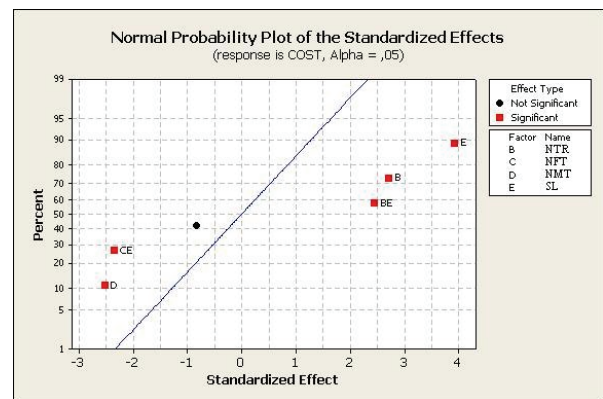
Figure 5: The Most Significant Effects for ADCP

The most significant factors (represented with a red square) are:

- NTR;
- NMT;
- SL;
- NTR*SL;
- NFT*SL.

Figure 6 shows the trend of the average daily cost for packages delivered (ADCP) in function of the main effects:
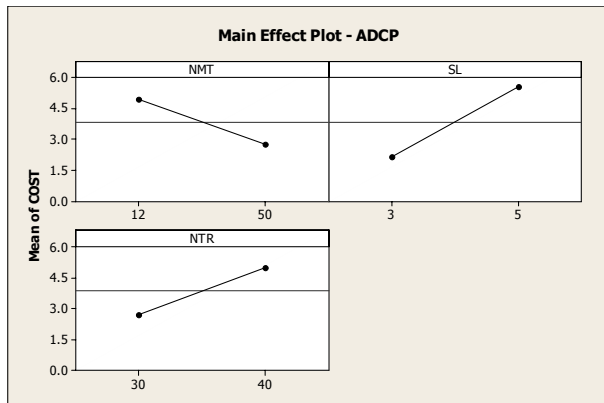
- NTR;
- NMT;
- SL.

Figure 6: ADCP versus Main Effects

As shown in Figure 6, when the number of lift trucks increases, the average daily cost for packages delivered decreases; the contrary happens when the shelves levels and the number of retailers' trucks.

Figure 7 shows the plots concerning the interaction effects between some couples of parameters (i.e NTR-NFT, NFT-SL).
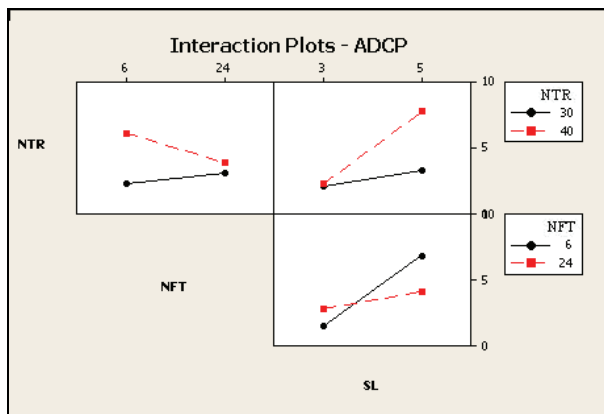

Figure 7: Interaction Plots for ADCP

Results obtained by means of DOE and ANOVA allow to correctly choose the correct resources allocation in order minimize the logistics internal costs. In effect an accurate combination of the number of forklifts and lift trucks, help to keep under control logistics internal costs. The validity of the results, obtained thanks to ANOVA, is still confirmed by residuals analysis. The starting hypothesis which ensure the validity of the ANOVA (observations normally and independently distributed, observations with the same variance for each possible combination of the factors levels) have been verified by using Normal Probability plots, residuals versus the order of the fitted data plots, and histograms of the residuals.

6. CONCLUSIONS

This research work focuses on the investigation of the effect of some critical parameters (the number of incoming trucks from suppliers, the number of outgoing trucks to retailers, the number of forklifts and lift trucks, etc.) on the number of packages delivered per days and on the warehouse internal logistics costs.

To this end the authors developed a simulation model supported by a smart graphic user interface and an output section capable of monitoring the most important warehouse performance measures.

The analyses carried out highlight how the warehouse performance measures (average number of packages delivered per day and average daily cost for each package) are affected by the warehouse resources availability and by the arrival of suppliers' and retailers' trucks. In addition, input-output analytical relations, have been evaluated by using the ANOVA. Such analytical relations become a powerful tools for warehouse design and management.

REFERENCES

Ahire, S. L., Schmidt, C. P., 1996. A model for a Mixed Continuous-Periodic Review One-Warehouse, N-Retailer inventory system. *Operational Research*, 92, 69 – 82.

Amato, F., Basile, F., Carbone, C., Chiacchio, P., 2005. An Approach to control automated warehouse systems. *Control Engineering Practice*, 13, 1223–1241.

Bruzzone A.G., Massei M., Brandolini M., 2006. Simulation Based Analysis on Different Logistics Solutions for Fresh Food Supply Chain. *Proceedings of SCSC2006, Calgary, Canada, July 30-August 3*

Bruzzone A.G., Bocca E. 2006. Logistics and Process Solutions for Supply Chain of Fresh Food In Retail. *Proceedings of HMS2006, Barcelona, October*

Bruzzone, A.G., Longo, F., Brandolini, M., 2006. Enhancement Process Based on Simulation Supporting Efficiency & Organization, *Proceedings of Summer Computer Simulation Conference*, July 30th – August 03rd, Calgary Canada.

Bruzzone, A.G., Longo, F., Viazzo, S., Mirabelli, G., Papoff, E., Massei, M., Briano, C., 2004. Discrete event simulation applied to modelling and analysis of a supply chain. *Proceedings of Modeling and Applied Simulation Conference*, Bergeggi (SV) (Italy).

Chen, F., Samroengraja, R., 2000. A Staggered Ordering Policy for One-Warehouse, Multiretailer Systems. *Operations Research*, 48 (2), 281 – 293.

Christofides, N., Colloff, I., 1973. The Rearrangement of Items in a Warehouse. *Operations Research*, 21 (2), 577 – 589.

Cormier, G., Gunn, E.A., 1996. Simple Models and Insights for Warehouse Sizing. *Operational Research*, 47, 690 – 696.

Curcio, D., Longo, F., Mirabelli, G., 2007. Inventory policies comparison in a manufacturing process using modeling & simulation. *Proceedings of the International Mediterranean Modelling Multiconference,* pp. 237 – 242. October 04 – 06, Bergeggi (Italy).

De Sensi, G. , Longo, F., Mirabelli, G., 2007. Inventory policies analysis under demand patterns and lead times constraints in a real supply chain. *International Journal of Production Research*, doi: 10.1080/00207540701528776

Eben-Chaime, M., Pliskin, N., Sosna, D., 1997. Operations management of multiple machine automatic warehousing systems. *Production Economics*, 51, 83–98.

Eben-Chaime, M., Pliskin, N., Sosna, D., 2004. An integrated architecture for simulation. *Computer and Industrial Engineering*, 46, 159–170.

Hill, R.M., 1989. Allocating Warehouse Stock in a Retail Chain. *Operational Research*, 40 (11), 983 – 992.

Hoare, N. P., Beasley, J. E., 2001. Placing Boxes on Shelves: A Case Study. *Operational Research*, 52 (6), 605 – 614.

Lee, C., Çetinkaya, S., Jaruphongsa, W., 2003. A Dynamic Model for Inventory Lot Sizing and Outbound Shipment Scheduling at a Third-Party Warehouse. *Operations Research*, 51 (5), 735 – 747.

Longo, F., Mirabelli, G., Papoff, E., 2006. Material Flow Analysis and Plant Lay-Out Optimization of a Manufacturing System. *International Journal of Computing*, 5(1), 107-116.

Longo, F., Mirabelli, G., 2008. An Advanced Supply Chain Management Tool Based on Modeling & Simulation, *Computer and Industrial Engineering*, Vol 54(3), 570-588.

Longo, F., Mirabelli, G., Papoff, E., Modeling Analysis and Simulation of a supply chain devoted to support pharmaceutical business retail, *Proceedings of the 18th International Conference on Production Research, July 31 – August 4 , Salerno, Italy.*

Liebeskind, A., 2005. *How to Optimize Your Warehouse Operations*. 1st ed. USA: Industrial Data & Information Inc.

Montgomery, D.C. and Runger, G.C., 2003. *Applied statistics and probability for engineers*. 3rd ed. London: McGraw-Hill.

Mulcahy, D.E., 1993. *Warehouse Distribution and Operations Handbook*. 1st ed. USA: McGraw-Hill.

Nahmias, S., 1982. Perishable Inventory Theory: A Review. *Operations Research*, 30 (4), 680 – 708.

Napolitano, M., Gross, J., 2003. *The Time, Space & Cost Guide to Better Warehouse Design*. 2nd ed. USA: Distribution Group.

Ratliff, H.D., Rosenthal, A.S., 1983. Orderpicking in a rectangular warehouse: a solvable case of traveling salesman problem. *Operations Research*, 31, 507 – 521.

Sanjay, J., 2004. Supply chain management tradeoffs analysis, *Proceedings of the Winter Simulation Conference*, pp. 1358-1364. Washington (USA).

Silver, E.A., Pyke, D.F., Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*. 3rd ed. USA: John Wiley & Sons.

AUTHORS BIOGRAPHIES

ENRICO BOCCA achieved the Logistics and Production Engineering degree in 1999 with a thesis on "Automated Reorder System in Supermarket Chains", summa cum laude. During 2000-2001 he was involved in Customer Satisfaction analysis for Retail Business Sector. He completed during 2003 the thesis in Management Engineering working on "Innovative Technology Management in Retail Sector", summa cum laude. During 2003 June-July he participated to IEPAL experienca coordinating an International Team of Engineering and MBA Students (from France, USA, Turkey) working in a Prefeasibility study related to People Counting in Shopping Centers; this initiative was sponsored by FIPSE US Dept.of Education, DGEACEC, LSC. September-December 2003 he worked as consultant on different initiatives: Business Plans related to R&D Projects; Logistics and Project Management for CFLI. He worked on a joint project DIPTEM/NASA in Kennedy Space Center for Modelling Test Lab Crane Operations, the demonstrator was presented in I/ITSEC2005 Orlando. He worked as researcher in Simulation Team of DIPTEM with special attention to Logistics in Retail Business. Currently he is charge as Responsible of M&D R&D in MAST.

DUILIO CURCIO was born in Vibo Valentia (Italy), on December the 15th, 1981. He took the degree in Mechanical Engineering from University of Calabria (2006). He is currently PhD student at the Mechanical Department of University of Calabria. His research activities include Modeling & Simulation and Inventory Management theory for production systems and Supply Chain design and management. He collaborates with the Industrial Engineering Section of the University of Calabria to research projects for supporting Research and Development in SMEs.

FRANCESCO LONGO took the degree in Mechanical Engineering from University of Calabria (2002) and the PhD in Industrial Engineering (2005). He is currently researcher at the Mechanical Department (Industrial Engineering Section) of University of Calabria. His research interests regard modeling & simulation of manufacturing systems and supply chain management, vulnerability and resilience, DOE, ANOVA. He is Responsible of the Modeling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES), member organization of the MS&Net (McLeod Modeling & Simulation Network) He is also member of the Society for Computer Simulation International and Liophant Simulation.

ALBERTO TREMORI Alberto Tremori is an Electronic Engineer, he took the degree from Genoa University; he acquired experience on the development of Simulators to support process re-engineering, logistics, transportation systems and warehouse automation. Alberto started his career by developing skills in programming (i.e. C/C++) and then moved to resource management. In 1997 he was involved in joint Simulation Researches Projects among National Simulation Center, Istitute for Simulation and Training and Kennedy Space Center and McLeod Institute of Simulation Science. During 1997/98 he worked as freelance consulting in Quality Management applied to Port Services, obtaining the first ISO9000 certificate for Mooring Nautical Service in Europe and completing the certification process of all 67 Italian Commercial Ports. Along the years he participated to several International Conferences as speakers in Europe, Asia and North America. In 2007 he served as Exhbition Chair for EuroSIW2007 in Santa Margherita Ligure, Italy; the same year he participated to the team demonstating DIPTEM/MISS/NASA joint researches in NASA Booth at I/ITSEC, November, Orlando. Alberto accumulated experiences along the years working in major companies, he operated in IBM as project Manager for the AS/400 system business area and in Xerox as Commercial Manager of Large Systems for Public Agencies. He was commercial manager in IDC and he worked also in several IT Companies, Logistics and Consulting Firms (think3, Solid Works, CFLI) as Business Development Responsible. He is currently serving as President of MAST. He is member of the Simulation Team of DIPTEM Genoa University and Liophant Simulation.

ANALYSING INVENTORY MANAGEMENT PERFORMANCE BY SIMULATING SUPPLY CHAIN MANAGEMENT STRATEGIES

Gökhan Yüzgülec [(a)], Markus Witthaut [(b)], Bernd Hellingrath [(c)]

[(a, b, c)] Fraunhofer-Institut für Materialfluss und Logistik (IML), Dortmund, Germany
[(c)] University of Paderborn, Germany

[(a)]goekhan.yuezguelec@iml.fraunhofer.de, [(b)]markus.witthaut@ iml.fraunhofer.de, [(c)]bernd.hellingrath@iml.fraunhofer.de

ABSTRACT
SCM strategies like information sharing achieve improvements of the performance in supply chains; however, due to the usual unbalanced improvement allocation they are not implemented very often. In order to analyse the benefits of SCM strategies, simulation experiments were carried out. In particular a five-stage linear supply chain (four suppliers, one plant, one retailer) was modelled and evaluated concerning the effects of the company-crossing information sharing strategy (planning-relevant and company-internal transmission of information to other involved companies) in contrast to sequential information exchange (pure information exchange about respective orders). Depending on the demand volatility as well as on the utilisation situation of the companies, effects of the inventory, the degree of utilisation as well as the adherence to delivery dates was analysed. Experiments showed that suppliers can profit especially through inventory reduction; the plant mainly profits through a more homogenous production utilisation and a higher adherence to delivery dates.

Keywords: Supply Chain Management, Simulation of SCM Strategies, Information Sharing, Inventory Management

1. INTRODUCTION

New strategies for the collaborative management of supply chains aim to fulfil increased customer requirements and strive at the same time to minimise the production and logistics through harmonised planning execution processes of the different companies of a supply chain. The core idea of these SCM strategies is to improve the customer satisfaction through better logistics service, especially in respect to lead times, due date reliability and product availability whereas at the same time the product and logistics costs are reduced as well. Examples of these collaborative strategies are Supply Chain Monitoring (SCMo) (Odette Recommendation SCMo 2003), collaborative demand and capacity management (Kuhn and Hellingrath 2002; Bock et. al. 2003; Krog et. al. 2002), Vendor Managed Inventory (VMI) (Waller et. al. 1999), information sharing (IS) (Cachon and Fisher 2000) and CPFR (Seifert 2003).

The realisation of such an SCM strategy leads to changes of the material flow, the planning and execution management processes, organisation structures, as well as supporting information systems. These changes involve costs for the introduction and the subsequent application of the respective SCM strategy. Moreover, the future costs and the potential benefits may differ between the participating enterprises. Consequently, companies hesitate to introduce new SCM strategies because they can not evaluate their costs and benefits in advance. Another obstacle for the proliferation of these strategies is the fact that even a single SCM strategy has a considerable design scope: With VMI, e.g., both supplier and customer have to agree on the minimal and maximal inventory levels as well as the planning frequency. So, companies within a supply chain face the question which is the most beneficial SCM strategy and how should the strategy parameters be defined.

Consequently, for configuration and evaluation of SCM strategies efficient instruments for cost-benefit analysis are needed. A reliable cost-benefit-assessment requires a detailed consideration of material flow and information processes factors such as transport lead times and their statistical distribution or the lead time for planning and information transfer. Hence, a dynamic time-related modelling and analysis such as event-based discrete simulation is required. The respective methods need to model the physical supply chain aspects including network structures and resources. Further on, it is necessary to model the information and planning processes of the various SCM strategies. An event-discrete simulation tool for the modelling of order-to-delivery processes in supply networks is OTD-NET, which was developed by Fraunhofer IML. OTD-NET allows detailed modelling of complex production networks with all their constraints and planning, execution and material flow processes (Wagenitz 2007).

Within the scope of the collaborative research centre "Modelling of Large Logistics Networks" (Collaborative Research Centre 559 1998-2008, funded by the German Research Foundation, www.sfb559.uni-dortmund.de) we have evaluated the effects of different

SCM strategies. One of these strategies is information sharing, which is characterised by a transmission of inventory or point-of-sales data from a customer up-stream of the supply chain to suppliers and sub-suppliers. An overview of simulation studies concerning information sharing shows inconclusive results (Schmidt and Knolmayer 2006): The studies show that several partners of a supply chain can benefit from information sharing; however, the degree of the effect varies considerably. Some studies describe improvements for suppliers only; other studies state that advantages by information sharing result also for customers, for example in terms of higher service levels. These contrary findings of the simulation studies trace back to the probable fact that often verification and validation of the models are missing, relevant measure variables are neglected, and partly model designs are highly simplified.

2. SIMULATION EXPERIMENTS

Using the OTD-NET supply chain simulation software developed at Fraunhofer IML, we carried out experiments for a multi-stage linear supply chain within the automotive industry. The purpose of these experiments was to derive how IS influences the performance of the supply chain in comparison to a conventional strategy in which information from a customer is exchanged only the direct supplier.

The investigated supply chain consists of one retailer, one plant, and one build-to-order supplier (BTO) and three build-to-stock suppliers (BTS). Figure 1 shows the respective supply chain structure. For the analysis scenario two different experiments were carried out in order to analyse the effect of the SCM strategies "Information Sharing" in comparison to the "Conventional Strategy (CS)", in which information is transferred only from stage to stage with a certain delay. Furthermore, the following scenarios have been investigated:

- Five different distributions (uniformly distributed) of the end customer demand were considered in order to analyse the sensitivity of the SCM strategies to the demand volatility. The initial situation is a constant end customer demand. In addition, volatilities of the end customer demand for +/- 10% (EDV 10), +/- 25% (EDV 25), +/- 35% (EDV 35) and for +/- 45% (EDV 45) per month were analysed.
- A further dimension of variation for the performed experiments was the ratio of the available production capacity to the capacity demand induced by the end customer. Therefore, experiments were carried out for nearly full load situations (average end customer demand/utilisation = 98% (U 98) respectively 95% (U 95) of the production capacity) as well as for relaxed load situations (end customer demand/utilisation = 80% (U 80)).

- Finally, two different scenarios concerning the delivery time (short delivery time from tier to tier – less than 12 hours (DT 1) – versus delivery times of one month (DT 2)) were considered. The simulation period covers one year, from January 1 until December 31.
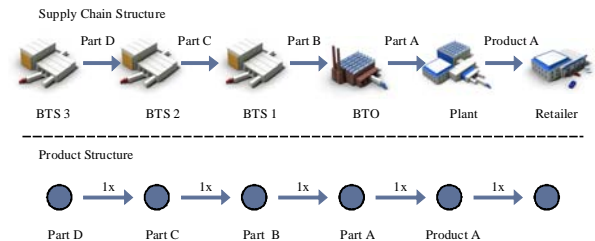

Figure 1: Supply Chain and Product Structure of the Created Model

As figure 1 shows, among the actors only an exchange of one product or one part takes place. The BTS supplier 3 is the source of the material flow in the supply chain; the retailer is the sink of the material flow. The BTS supplier 3 produces units of part D to stock and ships the ordered amount to the BTS supplier 2. This supplier makes units of part C to stock and forwards the amount of ordered units to BTS supplier 1. BTS supplier 1 proceeds with product B in analogous way but with the exception that its sink concerning the material flow is the BTO supplier. The BTO supplier works closely together with the plant. This supplier produces part A per order, activated by the plant. The plant produces in turn per order activated by the retailer. The retailer constitutes the interface to the end customer and transfers the orders to the plant once a month.

The target of each single company of the supply chain is a preferable homogenous production utilisation (nearby full load), the minimising of inventory for parts and for finished products as well as an assurance of a preferable high adherence to delivery dates.

For the evaluation of the effects of the SCM strategies following performance indicators are specified in the simulation experiments for all considered companies:

- Utilisation:
  Utilisation means production utilisation of the respective bottleneck resource of the companies. Here the average utilisation was defined.
- Inventory of finished products and parts:
  Average goods receipt and goods issue inventory were collected. Work-in-Progress inventory were not considered.
- Adherence to delivery dates:
  Degree of performance of the orders at the end customer (retailer in the existing model). For this, arrival time of ordered products A in the goods receipt at the retailer was used. The adherence to delivery dates is a value between

0 (no delivery was delivered on time) and 1 (all deliveries were delivered on time). In the following, however, the adherence to delivery dates will be converted into percentage and described in percent.

In the model it is assumed that for BTS supplier 3 there are infinite parts available which are needed for production if required.

3. SIMULATION RESULTS
3.1. Effects of IS on the Inventory

Figure 2 shows that at all suppliers' inventory level are highly varying. At the beginning of the period of consideration the inventory of the suppliers increase and are again on the decrease, shifted about approx. three months to the respective upstream supplier. During the months 6 to 8 the increase of the inventory at the BTO supplier proceeds almost linearly. Even at the end of the year the BTO supplier cannot decrease its inventory. However, the plant keeps its inventory almost constant and relatively low throughout the whole year. This is possible since the BTO supplier simply builds to order, orders activated by the plant. In a way the plant uses the stock of the BTO supplier and can therefore keep internal inventory relatively low.
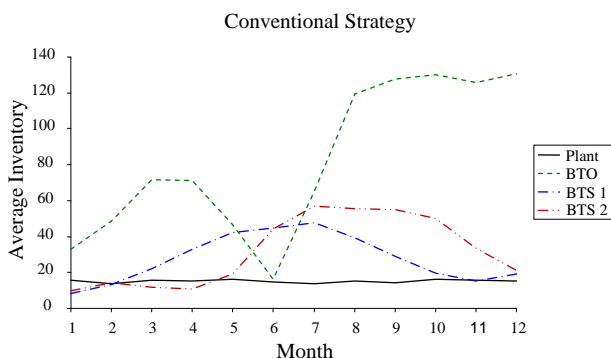

Figure 2: Companies' Inventory Situation with CS (Scenario EDV 35 / U 98)

On the other hand figure 3 shows the inventory situation of the actors by utilisation the information strategy throughout the supply chain. It becomes clear that inventory level at all suppliers is in average much lower than with the conventional strategy. Whereas the inventory situation of the plant changes hardly, the suppliers' inventory situation is considerably improved. Indeed, the BTO supplier's inventory increases over the period of consideration continuing; however, there are no inventory variations any more. Also BTS suppliers' inventory curves proceed relatively constant throughout the whole year – merely in August inventory decreases for a short time. Their inventory situation looks very similar, just with some little varieties.
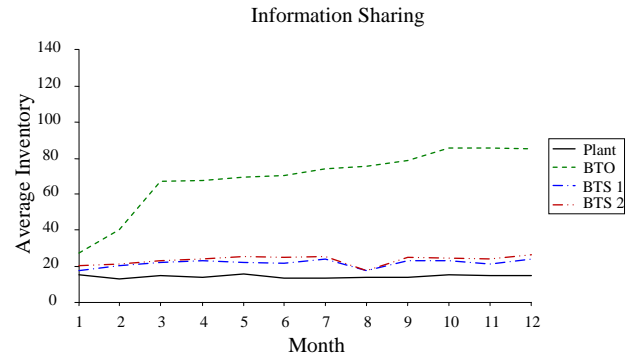
Figure 3: Companies' Inventory Situation with IS (Scenario EDV 35 / U 98)

Figure 4 shows these coherences with the aid of the example of BTS supplier 1 for different demand volatilities (10% and 35%) as well as for different degrees of utilisation (80% and 98%):
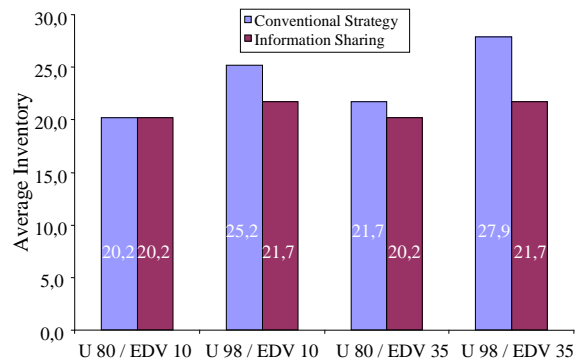
Figure 4: Inventory Situation of BTS supplier 1 for Different Scenarios

The strategy-related variation parameter "end customer demand information" has in the context of information sharing the biggest influence on the inventory. By transmission of the current end customer demand (retailer demand in the existing model) to all business partners of the logistics network, company-internal sales forecast becomes much more precise compared to conventional prognoses, based on past-related orders of the respective customer. Gained simulation results show that IS has different influences on the inventory situation of the companies; depending on which demand variation as well as depending on which production utilisation is prevailing. Most significant improvements through IS were expected when having high demand variations and high degree of utilisation of the actors at the same time. Furthermore, especially the upstream supply chain, the suppliers, profit by the inventory improvements.

3.2. Effects of IS on the Plant Utilisation

The analysis of the simulation results concerning the parameter degree of utilisation again made clear that the information sharing strategy achieves improvements compared to the conventional strategy. By transmission

of information in-between the companies the respective utilisation can be formed more evenly.

However, the smoothing of the production utilisation carries low weight when another initial situation is focused. Therefore, a slight demand variation anticipates simply a little optimisation of the load situation. This is clearly shown in figure 5 and figure 6 taking the plant as an example in turn. While achieving a strong smoothing by the application of IS with the scenario EDV 35 / U 98 (figure 5), the improvement with the scenario EDV 10 / U 98 (figure 6) is considerably lower.

Further analysis showed that information sharing has no influence on the utilisation situation of the companies when having long delivery times. Different scenarios showed no changes of companies' production utilisation. The evaluation of the utilisation situation for different scenarios illustrate clearly that an information sharing concept is only worthwhile when a high end customer demand fluctuation is existent.
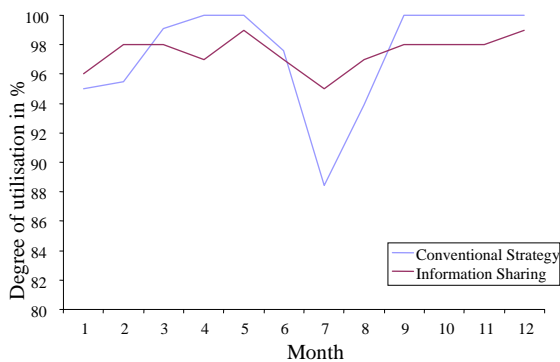

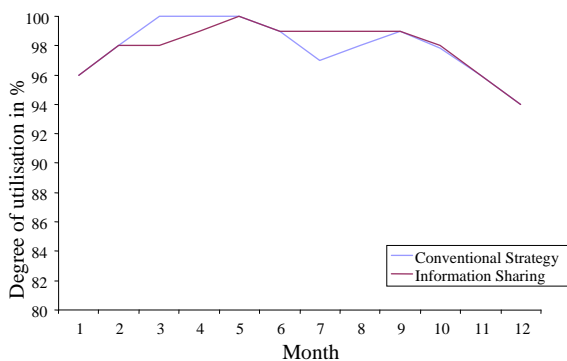Figure 5: Production Utilisation of the Plant, Scenario: EDV 35 / U 98

Figure 6: Production Utilisation of the Plant, Scenario: EDV 10 / U 98

In contrast to the demand distribution, the level of the degree of utilisation has no apparent influence on the extent of the smoothing of the production utilisation. This means that already at low degree of utilisation of the companies, smoothing effects can be achieved.

3.3. Effects of IS on the Adherence to Delivery Dates

The adherence to delivery dates is a very important parameter because among other things this parameter concerns the end customer and co-decides decisively the end customer's future buying pattern. A bad adherence to delivery dates effects a high or frequent delay in delivery of the ordered products or parts. A high adherence to delivery dates in contrast means that ordered products arrive at the appointed time at the customer, thus not earlier either.

Gained results show that the adherence to delivery dates of the logistics network is essentially depending on the utilisation situation of the respective company. During low production capacity utilisation (e.g. 80%) for example no or only a low improvement of the adherence to delivery dates can be achieved. However, when companies have full order books, ultimate load or even full load, by IS the adherence to delivery dates of the supply chain can be improved. But also the demand volatility has an influence on the adherence to delivery dates, even though a little less than the utilisation situation of the actors. The results of adherence to delivery dates with different scenarios are illustrated in figure 7.
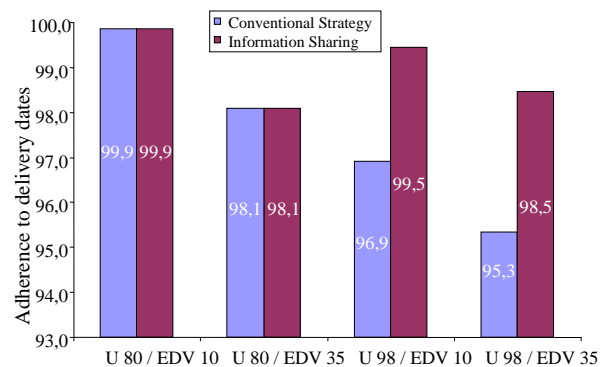
Figure 7: Adherence to Delivery Dates of the SC for Different Scenarios

Figure 7 shows that the adherence to delivery dates of the SC at a companies' degree of utilisation of 80% cannot be increased – neither at low (10%) nor at high (35%) demand volatility. When companies' production capacities are however almost utilised (ultimate load 98%), by information sharing a considerable improvement of the adherence to delivery dates of the SC can be seen. Especially, having high demand volatility (35%) in addition, information sharing seems to be very worthwhile concerning "adherence to delivery dates".

Table 1 shows exemplarily that for extended delivery times (one month) the adherence to delivery dates of BTS supplier 1 and BTS supplier 2 improves by application IS. With extended delivery dates the adherence to delivery dates is not as good as during the initial situation (12 hours) - the delay in delivery is therefore higher. By the information sharing strategy in average the adherence to delivery dates can be improved by 2% in contrast to the conventional

strategy. These results implicate that with long delivery times IS is of avail concerning the adherence to delivery dates of the entire logistics network.

Table 1: Comparison of Adherence to Delivery Dates with an Extended Delivery Time of One Month

Adherence to delivery dates	U 98			
	CS		IS	
	EDV 35/DT 1	EDV 35/DT 2	EDV 35/DT 1	EDV 35/DT 2
January	80%	78%	82%	80%
February	96%	95%	99%	97%
March	97%	95%	100%	97%
April	98%	97%	100%	98%
May	97%	96%	100%	98%
June	96%	94%	100%	96%
July	97%	97%	100%	99%
August	95%	95%	100%	96%
September	98%	97%	100%	98%
October	99%	97%	100%	98%
November	97%	97%	100%	98%
December	95%	94%	100%	96%
Average	**95%**	**94%**	**98%**	**96%**

Table 2 clarifies the percentage improvements of the average adherence to delivery dates through IS. Interesting is that the extent of utilisation has a bigger influence on the improvement of the adherence to delivery dates than the extent of the demand volatility. For instance, it can be noticed that the improvement of the adherence to delivery dates increases considerably when production utilisation of the companies increases while improvement decreases when demand volatility (10% - 45%) declines at the same time. This coherence can be explained with the fact that for low production utilisation there is enough rest capacity available to satisfy fairish high demand volatility in due time. During high utilisation degree, already a little demand volatility can lead to a delay in delivery. Therefore, the information sharing strategy has a significant effect during high and comparatively slight effect during low production utilisation of the companies.

Table 2: Improvement of Adherence to Delivery Dates for Different Scenarios

Demand Volatility / Degree of Utilisation	10% Difference %	25% Difference %	35% Difference %	45% Difference %
80%	0,00%	0,00%	0,00%	-0,83%
95%	-0,64%	-2,83%	-3,25%	-3,97%
98%	-2,63%	-2,97%	-3,36%	-4,04%

3.4. Effects of IS on All Factors

In principle the major advancement concerning the three analysed parameters "inventory", "degree of utilisation" and "adherence to delivery dates" can be achieved with a high production utilisation of the companies. Concerning each individual company of the supply chain following information concerning the parameters can be registered:

- Plant:
  The plant profits from the implementation of IS almost exclusively for the parameters degree of utilisation and adherence to delivery dates. The production utilisation is more homogeneous compared to the conventional strategy. The increased adherence to delivery dates of the supply chain is especially an advantage for the plant. On the other hand, the plant inventory can be only decreased marginally. Regarding this parameter the IS strategy is inconsiderable for the plant.

- BTO supplier:
  The BTO supplier achieves its profit in the context of IS predominantly by the decrease of inventory. During low demand volatility the BTO supplier can improve the inventory situation considerably. The improvement of the utilisation situation however is relatively low and therefore categorised as inconsiderable.

- BTS supplier:
  The BTS suppliers profit in different respects. By application of IS, inventory can be reduced and the utilisation situation can be improved considerably.

- All companies:
  The adherence to delivery dates of the entire supply chain can be improved by information sharing. Noticeable in particular is the reduction of the delay in delivery with a high production utilisation of the actors and high end customer demand volatility at the same time.

Therefore, the biggest effect of information sharing is shown during high production utilisation of the companies of the supply chain as well as during high end customer demand volatility. Based on the simulation results it can be registered that the main profiting actors of an IS concept are upstream tiers of the logistics network. Simulation results show that the percentage improvements of the average inventory of the BTS suppliers throughout all scenarios are above those of the entire supply chain. By reason of this significant reduction of the average BTS suppliers' inventory, the BTS suppliers are those who profit most from implementing the IS strategy.

4. CONCLUSIONS AND OUTLOOK

The analysis of the simulation results shows that information transmission between the companies of the supply chain leads to an optimisation of company-internal parameters (inventory, degree of utilisation) but also of parameters throughout the network (adherence to delivery dates). The simulation results indicate that particularly upstream stages of a supply chain benefit through the information sharing strategy, especially through inventory reduction. On the other hand, the major profit for downstream stages of a supply chain, e.g. a final assembly plant, is the increase in production utilisation homogeneity as well as the improvement in adherence to delivery dates. The simulation results also demonstrate that the degree of the consequences of information sharing is depending on both the volatility of the end customer demand and the situation of the plant utilisation. If both are comparative high at the same time, significant improvements result out of the

information sharing strategy for all companies in the supply chain.

In addition to the presented experiments in this work, in which production utilisation for all companies was assumed identically, different production utilisations could simulate a much more realistic situation. Furthermore, not only inventory costs and costs for non-deliveries but also transport costs, production costs etc. could be considered in the model an in this way allocation decisions in the SC could be influenced strongly.

REFERENCES

Bock, D., Breithor, T., Laforsch, M., Weingarten, U., 2003. *Studie Supply Chain Collaboration: Unternehemensübergreifende Zusammenarbeit.* Bremen: Bundevereinigung Logistik (BVL)

Cachon, G.P., Fisher, M., 2000. Supply Chain Inventory Management and the Value of Shared Information. *Management Science* v.46: pp.1032-1048.

Collaborative Research Centre 559 "Modelling of Large Logistics Networks", 1998-2008. www.sfb559.uni-dortmund.de.

Krog, E. H., Richartz, G., Kanschat, R., Hemken, M., 2002. Kooperatives Bedarfs- und Kapazitätsmanagement der Automobilhersteller und Systemlieferanten. *Logistik Management*, v. 4: pp. 45-51.

Kuhn, A., Hellingrath, B., 2002. *Supply Chain Management – Optimierte Zusammenarbeit in der Wertschöpfungskette.* Berlin, Heidelberg, New York: Springer.

Odette International Limited, 2003. *Supply Chain Monitoring Recommendation.* Available from: www.odette.org/html/Pubs%20More%20Info/SC Mo%20QandAsV4.pdf.

Schmidt, R., Knolmayer, G., 2006. *Ein Vergleich von Simulationsstudien zu Information Sharing und Vendor Managed Inventory.* Available from: www.ie.iwi.unibe.ch/publikationen/berichte/resour ce/WP-185.pdf.

Seifert, D., 2003. *Collaborative Planning, Forecasting, and Replenishment: How to Create a Supply Chain Advantage.* New York: Galileo Business.

Wagenitz, A., 2007. *Modellierungsmethode zur Auftragsabwicklung in der Automobilindustrie.* Dissertation, Technische Universität Dortmund.

Waller, M., Johnson, M.E., Davis, T., 1999. Vendor managed inventory in the retail supply chain. *Journal of Business Logistics* v. 20: pp.183-203.

AUTHORS BIOGRAPHY

Dipl.-Ing. Gökhan Yüzgülec is a Mechanical Engineer and works at the Fraunhofer-Institute for Material Flow and Logistics (IML) in Dortmund, Germany. His main working areas are organisation concepts and supply chain management (SCM). He puts special emphasis on the interplay of strategies for the planning and monitoring of logistics networks with various organisation forms. Furthermore, he is engaged in the measuring and evaluation of the combinations of strategies and organisation concepts since it is an essential factor for the success of logistics networks.

Dipl.-Inform. Markus Witthaut is a Computer Scientist and leading the Supply Chain Management (SCM) Team at the Fraunhofer-Institute for Material Flow and Logistics (IML) in Dortmund, Germany. His main working areas are new business processes, integration concepts an information technology in the field of (SCM). Furthermore he is a part-time lecturer on software engineering at the Applied University of Dortmund.

Prof. Dr.-Ing. Bernd Hellingrath is heading the department Enterprise Modelling at the Fraunhofer-Institute for Material Flow and Logistics (IML) in Dortmund. He is teaching as a professor at the University of Paderborn in the area of modelling and planning of production and logistics networks. His main research work lies in the areas of modelling methods and IT systems for Supply Chain Management and the application of simulation technology in logistics.

SPS TOOLS FOR CAPITAL PLANNING PROJECT ANALYSIS

Nathan Boskers[(a)], Simaan M. AbouRizk[(b)]

[(a)] Springwood Developments Ltd.
[(b)]University of Alberta

[(a)] nathan@springwood.ca, [(b)]abourizk@ ualberta.ca

ABSTRACT

This paper presents the development of a special purpose simulation (SPS) template for the capital planning on large-scale infrastructure construction projects. The template targets outputs of greatest interest to the planner: project duration and cost estimate, using the Box Jenkins method of forecasting. The developed tool was then tested in a case study involving roadway restructuring in Edmonton, Canada. The case study is detailed here for purposes of illustration. Seven models were simulated and analysed, showing the relative influence of fixed rates and mean rates on forecasting, and that subjective analysis procedures can cause disadvantageous decision-making. The case study was also used to perform risk analysis for the City of Edmonton.

Simulation models; intelligent forecasting; capital projects; construction project planning

1 INTRODUCTION

Computer simulation is the process of designing a mathematical-logical model of a real world system and experimenting with the model on a computer (AbouRizk et. al. 1995). Computers ideal tools for modeling real world systems because they can process large amounts of data in short periods of time. Halpin developed the first simulation model for construction purposes around 1973 (AbouRizk et. al. 1992). The developed model was called CYCLONE, standing for cyclic operations networks. The CYCLONE models were based on resources and their interactions with each other.

Due to the success of CYCLONE, much further research was done in the area of computer simulation for construction purposes. Some researchers enhanced CYCLONE, or developed new models, such as INSIGHT, UM-CYCLONE, and others. Carr (1979) developed a system called MUD (Model for Uncertainty Determination). This model simulated network scheduling to estimate activity durations. Dabbas and Halpin (1982) developed a similar simulation, which estimated activity durations, by combining the CYCLONE and a CPM scheduling software system. Lutz and Halpin (1990) developed a simulation to monitor process cycles and stage buffers using microCYCLONE. Liu and Ioannou (1992) developed a discrete-event simulator system, which used an object-oriented design. The simulator could track resources, construct models using a graphical interface, capture resources, define different resources, and link with other planning systems. The system they developed was called COOPS (Construction Object-Oriented Process Simulation System). A knowledge-based simulation was created by Odeh et. al. (1992), called CIPROS. This system was an object-oriented system used for developing discrete-event simulation networks. The CIPROS system enabled users to relate construction plans and specifications to a construction plan.

For project selection purposes, Ruwanpura et. al. (2002) developed a Special Purpose Simulation (SPS) tool. The SPS tool incorporated influence diagrams, with the various components representing the input data. The project parameters were analyzed using Monte Carlo analysis, and a project NPV distribution was produced. Again, like the methods developed by Ye and Tiong (2000) and Lu et al. (2000), a confidence level was used on the NPV distribution to determine the projects attractiveness. Ye and Tiong's study into the various project analysis methods showed that the NPV method is the most suitable method. This can be verified by the NPV method's wide use in industry and in current research. As well, Monte Carlo analysis has become widely used and accepted as a method to account for uncertainty in project costs.

Hajjar and AbouRizk (1999) developed a simulation environment called *Simphony*, for the purpose of building special purpose simulation (SPS) tools. AbouRizk and Hajjar (1998) defined SPS as a computer-based environment built to enable a practitioner who is knowledgeable in a given domain, but not necessarily in simulation, to model a project within that domain in a manner where symbolic representations, navigation schemes within the environment, creation of model specifications, and reporting are completed in a format native to the domain itself. Special purpose simulation (SPS) tools developed in *Simphony* have proven to be very useful for the purposes of analyzing and optimizing capital investment projects. Templates created in *Simphony* use modelling elements for the purpose of

building project specific simulation models. The planning template has seven modelling elements. Each modelling element contains all, or a combination of, input parameters, outputs, and statistics. In some cases, modelling elements can obtain input values from other elements on different hierarchical levels, as well as report outputs to other modelling elements on different hierarchical levels.

With respect to capital investment projects, project analysts have a great need for accurate knowledge of future cash flows. Future cash flows are used to make important decisions, such as the procurement method to use for the construction phase of a project, or what kind of financing should be used for a project, and so on. With the aid of accurate forecasting techniques, such as simulation, cash flow analysis can be done more objectively, which will result in the ability to make more informed planning decisions.

2   BOX JENKINS FORECASTING

The Box-Jenkins methodology for forecasting is used for the purposes of forecasting non-stationary time-series. A non-stationary time-series is simply a time-series that has changing parameters over time (i.e. the mean and variance are not constant). The Box-Jenkins methodology suits itself well to capital investment project planning, or overall lifecycle project planning in general, because such things as Consumer Price Indexes (CPIs) and Construction Price Indexes, which are used in forecasting cash flows, are by nature non-stationary time-series. Hence, the Box-Jenkins forecasting techniques can be well utilized for the forecasting of those indices.

Before an understanding of the steps of the Box-Jenkins methodology is possible, a general understanding of stationary time-series analysis is required. Next, stationary time-series analysis will briefly be discussed. The Box-Jenkins model is for analyzing non-stationary time-series. The Box-Jenkins model simply combines the idea of differencing and the ARMA model into one single model, called the Autoregressive Integrated Moving Average (ARIMA) model. The general ARIMA model can be defined as the following:

$$\phi_P(B)Y_t = \theta_q(B)a_t \tag{1}$$

where

$$Y_t = \begin{cases} (1-B)^d(1-B^s)^{d1}z_t & \text{if } d > 0 \text{ and/or } d1 > 0 \\ \\ z_t - \mu & \text{if } d = 0 \text{ and } d1 = 0 \end{cases} \tag{2}$$

The general ARIMA model can be expressed as ARIMA (p, d, d1, q).

An addition to the general ARIMA model can be added to incorporate a seasonality component into the model. Seasonality implies a repeating pattern in the time-series over a seasonal cycle. This type of pattern is incorporated into the ARIMA model as follows:

$$\phi_p(B)\phi_{p1}(B^s)Y_t = \theta_q(B)\theta_{q1}(B^s)a_t \tag{3}$$

where Y_t is the same as for the general ARIMA model.

The complete ARIMA model is then expressed as ARIMA (p, p1, d, d1, q1, q).

2.1   Box Jenkins Methodology

Using the techniques described above, four steps can be followed to aid in the task of forecasting future cash flows. The four steps are: 1) model identification, 2) parameter estimation, 3) diagnostic checking, and 4) forecasting. Figure 1 shows how these steps are carried out.
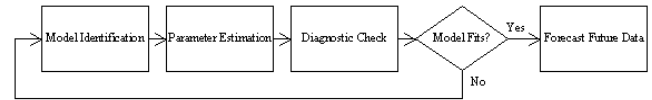
Figure 1: Box Jenkins forecasting methodology

2.1.1   Step 1: Model Identification

The first step in the Box-Jenkins methodology involves determining a tentative model to use in the succeeding steps of the methodology. In order to determine the parameters of the tentative model, the analyst must analyze the plotted data series, the plotted SACs and the plotted SPACs. Using these plots, the analyst must first determine the amount of differencing required, and then from this differenced model, determine the amount of regular and seasonal back shift that is required.

To determine the amount of differencing to use, the analyst should generate many combinations of SACFs and SPACFs by using various combinations of d and d1 values. Many plots should be generated, as the analyst does not know beforehand which differencing is required. As mentioned earlier, if a series is stationary, the plot of data will not have an increasing or decreasing mean, and the SACF and SPACF will tend to dampen quickly.

To determine the orders of the ARIMA model, reference tables can be used. The decision on the ARIMA parameters is subjective, so a few tentative models may be chosen to carry through to the next steps.

2.1.2 Step 2: Parameter Estimation

The next step in the Box-Jenkins methodology is to estimate the model's parameters ϕ and θ. The most appropriate values for the parameters are those that minimize the residuals of the model, a_t. The residuals represent the difference between the actual historical data points, and the estimated historical data points using the tentative ARIMA model with estimated parameters, i.e. $a_t = z_t - \widehat{z}_t$. When the residuals are minimized, the tentative model produces results most similar to the actual historical data.

The most efficient means of obtaining the parameter estimates is by utilizing a computer program. The procedure for determining the parameters is basically a nonlinear optimization problem. The objective function of the problem is to minimize the sum of the residuals squared, i.e. $a_t{}^2$, and the decision variables are the values of the model parameters. The reason a computer program is needed to perform this analysis is due to the many iterations required to test the parameter estimates. Due to the problem being iterative in nature, initial estimate values for the parameters are required for the first iteration. Gaynor and Kirkpatrick (1994) suggest using 0.1 as a good starting value for all ϕ s and θ s being estimated.

2.1.3 Step 3: Diagnostic Checking

The third step in the Box-Jenkins methodology is to test the tentative model with the estimated parameters for its closeness of fit, which is done by running tests on the residuals, a_t. The test that is used is the Chi-Squared test.

The Chi-Squared test checks to see if there is any systematic error in the residual SAC. The first step in performing the Chi-Squared test is to calculate the test Chi-Squared statistic, Q, as follows:

$$Q = n \sum_{k=1}^{K} r_k{}^2(\widehat{a})$$

(4)

where n is the total number of observations minus the maximum back shift, K is the number of residual SAC values that have been calculated, and $r_k{}^2(\widehat{a})$ is the residual SAC of the series $(\widehat{a}_t)$ at lag k. The second step in performing the Chi-Squared test is to determine a Chi-Squared distribution, X^2 with K-p-q degrees of freedom, where p and q are the AR and MA orders. The third and final step is to compare the calculated value Q versus the Chi-Squared distribution X^2. If Q is greater than X^2, the model is inadequate, and if Q is less than X^2, the model is adequate.

If the Chi-Squared test failed, further improvement to the tentative model is required. The analyst must determine how to improve the model and repeat steps 1 through 3 over and over until an adequate model is obtained. The first thing the analyst should do is look at the pattern of the SACF of the residuals obtained in step 3. This pattern will help indicate the improvement required for the inadequate model. For example, if the SACF shows a large spike at the first lag, yet no MA order is present, this is a good indication that a MA with q=1 should be added to the model.

2.1.4 Step 4: Forecasting

Once an adequate model has been determined, the analyst can proceed with the final step of forecasting future data. Using the determined model, the ARIMA model equation is rearranged to solve for z_t. In order to better demonstrate how this process works, consider an ARIMA (2, 0, 1, 0, 0, and 0) model. Also, assume that the parameters were estimated to be $\widehat{\phi}_1 = 0.25$ and $\widehat{\phi}_2 = 0.45$. The model would be written and rearranged as follows:

$$(1 - 0.25B^1 - 0.45B^2)(1 - B)^1 z_t = a_t$$

(5)

$$(1 - 1.25B^1 - 0.2B^2 + 0.45B^3)z_t = a_t$$

(6)

$$z_t = 1.25z_{t-1} + 0.2z_{t-2} + 0.45z_{t-3} + a_t$$

(7)

Converting equation (7) into a forecasting equation, let T be the current time period, and l be the period in the future at the end of the current period. Equation (7) then becomes:

$$\widehat{z}_T(l) = 1.25z_{T+l-1} + 0.2z_{T+l-2} + 0.45z_{T+l-3} + a_{T+l}$$

(8)

Equation (8) can then be used to forecast the desired future data from the historical data.

3 MODELLING ELEMENTS

3.1 Input Parameters

The input parameters in the modelling elements are parameters that require the template user to manually enter a value. The inputted values will be used by the template to calculate the output. All the modelling elements and their respective inputs are listed in Table 1.

The project planner element is the highest element on the hierarchical structure of elements. The input parameters required in this element are those that are related to the overall project analysis. For example, the discount rate value is entered as a percentage in this element. As well, outputs generated from the modelling elements on lower hierarchical levels are reported to this element and used as input. The project planner has the

following input parameter (see Table 1) for simulation: Discount Rate (%). The discount rate value is what is used in the computation of the overall project's NPV.

Table 1: Modelling elements with input variables

Modelling Element	Description of the Input Variable	Deterministic (D) or Stochastic (S)		Subjective Variable
		D	S	
Project Planner	Discount Rate (%)	X		
Collector	Collector Description	N/A	N/A	
	Overhead (%)	X		
Activity	Model Duration?	N/A	N/A	Yes/No
	Forecast Type	N/A	N/A	Select from list
	Activity Description	N/A	N/A	
	Activity Duration		X	
	Activity Direct Cost ($)		X	
	Activity Indirect Cost ($)		X	

The collector element is the element used to represent a project's work package. This element is below the project planner element in the hierarchical structure, but above the activity element. The input parameters in this element relate to the sub-work packages and activities inside the given work package. The collector element has the following input parameters (see Table 1) for simulation: Collector Description, and Overhead (%). The overhead input parameter is entered as a percentage number, and is multiplied by the total costs obtained from the outputs of the elements in the lower hierarchical levels. The purpose of this input parameter is because often times at the conceptual planning stages of a project, the overhead cost estimate ("overhead" meaning such things as project office staff and other office resources) is given as a percentage of the direct project work package cost.

The activity element is the element used to represent a project's activities. This element is below the collector element in the hierarchical structure. The input parameters in this element relate to the costs and durations of the activities. The input parameters entered here affect the overall project analysis output. The input parameter duration affects the timing of all succeeding activities and succeeding work package, which affects the total project

NPV. The input parameter cost affects the overall time period cash flow, which affects the total project NPV. The activity element has the following input parameters (see Table 1) for simulation: Model Duration, Forecast Type, Activity Description, Activity Duration, Activity Direct Cost, and Activity Indirect Cost.

The start, finish, in, and out elements are all used for the purpose of guiding the path of the simulation model. None of these elements have any input parameters.

3.2 Output and Statistics

The planning template performs calculations and computations to produce the desired output and statistics an analyst would use for the analysis of a project. *Simphony* allows the template designer to create template specific output and statistics if desired, as well as provides built-in functional features.

The only output in the planning template that is seen by the analyst is in the finish element. The output here was specifically incorporated into the planning template by the template designer, so that the analyst could determine the overall project finish time for each simulation run. The other output in the planning template is produced by *Simphony* and used as input in other elements, where it is needed for further computations. This output is irrelevant to the analyst, and so hence is not given in visual form.

The statistics that are calculated and represented in graphical form are all done in the project planner element. The first statistic is the NPV. The NPV is collected for each simulation run, and the statistics are calculated on the collected data. The statistics on the collected NPVs are used by the analyst to analyse the project. The statistics given then are the project's maximum, minimum, and mean NPV. In addition to this, a CDF is produced. The CDF is what the analyst can use as an indicator at a specific level of confidence to tell if the project is worth investing in or not. The final statistic is the project time. This statistic is generated for the purposes of knowing the possible ranges of the project duration, so that planning can be made accordingly.

4 CASE STUDY

The Yellowhead Trail and 156 Street Grade Separation project, which is owned by the City of Edmonton, was used for a case study to test the developed SPS tool. At the time of modelling this project, the project was at the conceptual design phase. The purpose of the study was twofold: first, to test the developed SPS tool, and second, to perform risk analysis for the City of Edmonton so that they could have a more reliable cost estimate for the project. A more accurate cost estimate would enable the

City to properly assign financial resources to the project, and not tie up any additional capital that would not be required.

The planning template is an SPS tool developed in the *Simphony* environment specifically for the purposes of analyzing and optimizing capital investment projects. The planning template has a graphical user interface that allows for the easy creation of capital investment project models. It can be reused over and over again for the analysis of countless projects.

Please note that this paper uses the term 'range estimate,' which should be understood as the application of Monte Carlo techniques to project costs and revenues.

4.1 Developed Cost Estimate Models

Seven models were developed for the analysis process. The first model was the fixed cost estimate provided from the City. This model did not incorporate any uncertainty, nor did it incorporate any future costs due to inflation. This model was used as the base model. The second model was the fixed estimate plus the incorporation of an inflation rate of 3%. The third model was a range estimate of the costs, without any inflation costs incorporated. The fourth model was a range estimate and network analysis model, with an inflation rate of 3%. The fifth model was a range estimate and network analysis model, with a forecasted inflation rate. The sixth model was a range estimate and network analysis model over a 15-year period, with an inflation rate of 3%. The seventh and final model was a range estimate and network analysis model over a 15-year period, with a forecasted inflation rate.

The first two models were used for the purpose of comparisons with the other developed models. These models are what current industry would use in project analysis. The first model does not incorporate any uncertainty, and assumes the inputs (future cash flows) are known with absolute certainty. As well, the first model does not incorporate any escalation costs due to inflation. The second model is the same as the first, except it incorporates inflation by subjectively selecting an inflation rate of 3%.

The third model developed was strictly a range estimate of the costs. This model was done using the Range Estimate Template in *Simphony*. The costs were in current dollars (today's dollars), and were directly inputted into the simulation model from the data that was provided by the City. The contingency value that was provided from the City for the CN Structure and the Yellowhead Trail Structure were not included in the estimate costs. It was assumed that the purpose of this contingency was to account for risk and uncertainty in the estimate values, due to the early stage of the design work. Since the purpose of range estimating analysis is to account for risk and uncertainty in project costs, it was concluded that by including the contingency values, the risk and uncertainty associated with structure costs would be over accounted for.

The fourth model developed was done by creating a network model of the project's work packages and activities. The network model was based on the Microsoft Project bar chart schedule, provided by the City. In the analysis done, the work packages were defined by the 10 construction activities on the schedule. Activities were defined as work done inside each work package. Some assumptions had to be made as to which activities belonged to which work packages. It was assumed that the 'Excavation' and 'Fill' activities from the owner's estimate belonged to the 'Fill Placement' work package on the schedule. It was assumed that the '50mm Grind', '50mm Overlay', 'Slab-on Islands', and 'Clean & Grubbing' activities were part of the 'Paving Interchange' work package. It was assumed that the 'Landscaping' work package items 'Naturalized Wetland', 'Standard Turf', '150mm Topsoil', 'Wetland Fills', and 'Wetland Soils' would occur between April 1, 2004 and August 4, 2004. It was assumed that 'Formal Planting in Beds', 'Street Tree Planting', and 'Individual Axial Planting' activities would occur between July 1, 2005 and January 1, 2005. And finally it was assumed that 'Asphalt Trail', 'Benches', 'Picnic Tables', and 'Waste Receptacles' activities would occur between March 18, 2007 and October 1, 2007. The reason these assumptions had to be made was because of the escalation of costs due to inflation calculations that were performed on them. Varying the occurrence of an activity would subsequently result in altering the total project cost. As well, it should be noted that due to the lack of detail in the schedule provided, all activity costs occurring within a work package had to be spread evenly over the work package duration. For the fourth model developed, a subjective inflation rate of 3% was used.

The fifth model that was developed was identical to the fourth model developed, as explained above, except for one notable difference. The inflation rate that was used in this models analysis was not subjectively selected; rather it was forecasted using forecasting techniques. The same assumptions made in the network set-up from the fourth model were also assumed in this model. The results of this model are a project total cost in actual dollars.

The sixth model that was developed was identical to the fourth model developed, as explained previously. The difference between the fourth and sixth models was in the modelled project's time of occurrence. The fourth model was modelled to start August 1, 2002, while the sixth model was modelled to occur beginning October 1, 2011. This was done was to compare the long term effect of using a forecasted inflation (which is model seven) versus a subjective inflation rate (which is model six) on the total project cost.

The seventh and final model that was developed was identical to the sixth model, except that the inflation rate used here was forecasted using forecasting techniques.

It should be noted that all the input data used for all the developed models was provided by the City of Edmonton. The data includes durations, quantities, and costs, given as parameters for triangular and uniform distributions.

4.2 Analysis Procedure

Upon determining the project network sequencing, the project network was input into the developed SPS tool in the *Simphony* modelling environment for the discussed models. Figure 2 shows a graphical image of the network created at the highest hierarchical level. Once the network was inputted, all the project activity parameters, such as duration and cost, were defined. Once all the parameters were defined, all the developed simulation models were run for 100 simulation runs, and statistics calculations were performed on the collected data.
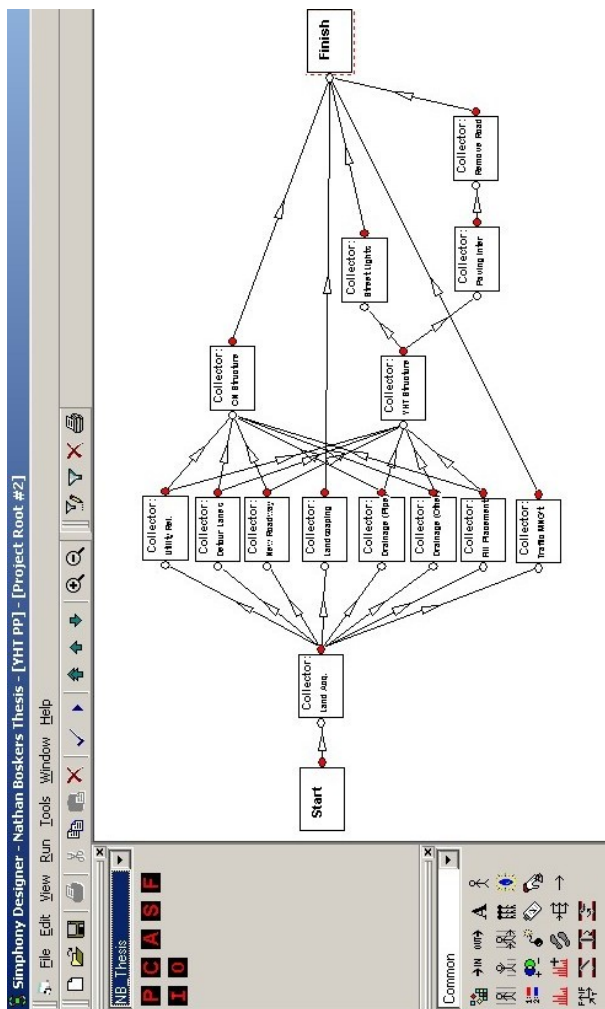

Figure 2: Case study network diagram

4.3 Summary of Simulation Results

The results of the simulation analysis are summarised in Table 2. Figure 3 shows the cumulative distribution function (CDF) for the range estimate plus forecasted inflation costs.
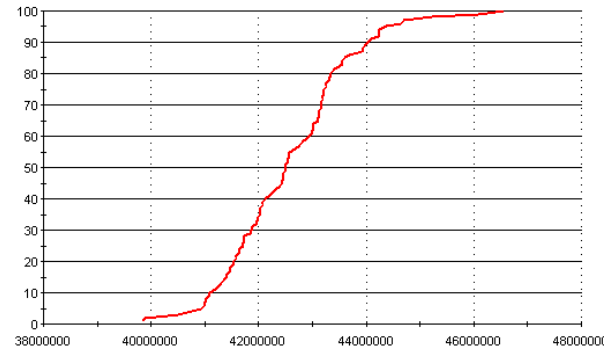

Figure 3: CDF for range estimate with forecasted inflation

4.4 Discussion

As indicated earlier, the purpose of the case study was twofold. The first reason was to test the developed SPS tool. Upon viewing the output of the seven models, some observations can be made. The first observation is that the mean range estimate costs for all cases are higher than the fixed estimate costs. This means that if the fixed estimate were used for budgeting purposes, the actual costs of the project would most likely be overrun. The second observation that can be made is that the forecasted inflation rate produced outputs that were consistently less than the subjective inflation rate of 3%. It can be observed as well, that as the project life increases (as in the case of the 15 year project), the difference between the total project cost for the forecasted model and the arbitrary model increases. This means that by subjectively selecting an inflation rate, the cash flows of the project can be significantly different from the cash flows of the forecasted rate. This could result in the analyst consistently overcompensating for the cost of inflation, or not compensating enough, depending on the subjective rate that is used. In the case study, the costs determined using the subjective rate would over account for the cost of inflation, and hence the City would budget more capital than necessary to the project.

The second reason for the case study was to perform risk analysis for the City of Edmonton. The results obtained indicated that the data provided by the City was not ideal in the sense that the fixed estimate totals were too similar to the range estimate totals. The main reason for this is due to the input data that was provided from the

City. For example, the work package 'Land Acquisition', which is the largest work package in value, was a fixed number, with no variance at all. Other large work package items had relatively small variance, which as well could contribute to the small overall project cost variance. According to a study performed by AbouRizk et al. (2002) for City of Edmonton projects, the suggested variance for preliminary design estimates for road construction should be + 25%. The results of the analysis show that the total project cost ranges from approximately $39.8 million to $46.6 million, with a mean of $42.6 million. The variance here is about $4 million, which is about 9.4%.

This actual variance determined from the analysis is much less than the suggested value made by AbouRizk et al. (2002). This implies that the input data provided by the City was not as generous in the size of the ranges as it should have been, considering the phase of the design. In order to ensure that the data provided for the input parameters for the project are accurate, the activities that make up 80% of the total project cost should be checked, as they are the activities that contribute most to the project's overall cost.

5   CONCLUSION

Range estimating and network analysis techniques were used to analyse the Yellowhead Trail and 156 Street Grade Separation project. Seven simulation models were developed for the purposes of testing the developed SPS tool, as well as performing risk analysis for the City of Edmonton. The output generated yielded some interesting observations, the first being that the fixed estimates were always less than the range estimates, and the second being that the forecasted inflation rate produced cash flows that were significantly different than the arbitrary rate selected, showing that subjective project analysis procedures could lead to improper decision making.

REFERENCES

AbouRizk, S.M., Babey, G.M., and Karumanasseri, G. (2002) "Estimating Capital Projects: An Empirical Study of Accuracy Levels for Municipal Government Projects." *Canadian Journal of Civil Engineering* 29 (5): 653-661.

AbouRizk, S.M., and Hajjar, D. (1998). "A Framework for Applying Simulation in Construction." *Canadian Journal of Civil Engineering*. 25(3), 604-617.

AbouRizk, S.M., Halpin, D.W., and Lutz, J.D. (1992). State of the Art in Construction Simulation. *Proceedings of the 1992 Winter Simulation Conference*, Arlington, V., 1271-1277.

AbouRizk, S.M., Shi, J., McCabe, B., and Hajjar, D. (1995). "Automating the Process of Building Simulation Models." *Proceedings of the 1995 Winter Simulation Conference*, Piscataway, NJ., 1032-1038.

Carr, R.I. (1979). "Simulation of Construction Project Duration." *Journal of Construction Engineering and Management*, 105(2), 117-128.

Dabbas, M., and Halpin, D.W. (1982). "Integrated Project and Process Management." *Journal of the Construction Division*, 109(1), 361-373.

Gaynor, P.E. and Kirkpatrick, R.C. (1994), Introduction to Time Series Modeling and Forecasting in Business and Economics, McGraw-Hill, Inc.

Hajjar, D., and AbouRizk, S. (1999). "Simphony: An Environment for Building Special Purpose Construction Simulation Tools." *Proceedings of the 1999 Winter Simulation Conference*, Phoenix, AZ., 998-1006.

Liu, L.Y. and Ioannou, P.G. (1992). Graphical Object-Oriented Simulation System for Construction Process Modeling. *Proceedings of the Eighth Conference on Computing in Civil Engineering*, Dallas, TX., 1139-1146.

Lu, Y.C., Wu, S., Chen, D.H., and Lin, Y.Y. (2000). "BOT Projects in Taiwan: Financial Modeling Risk, Term Structure of Net Cash Flows, and Project at Risk Analysis." *The Journal of Project Finance*, 5(4), 53-63.

Lutz, J.D. (1990). Planning of Linear Construction Using Simulation and Line Balance. Ph.D. Dissertation, School of Civil Engineering, Purdue University, W. Lafayette, IN.

Odeh, A.M., Tommelein, I.D., and Carr, R.I. (1992). Knowledge-Based Simulation of Construction Plans. *Proceeding of the Eighth Conference on Computing in Civil Engineering*, Dallas, TX., 1042-1049.

Ruwanpura, J.Y., Powers, G., Dolhan, G., and Chu, M. (2002). Simulation Based Project Selection Decision Analysis Tool. *Proceedings of the 2002 Winter Simulation Conference*, Not yet published.

Ye, S., and Tiong, R.L.K. (2000). "NPV-at-Risk Method in Infrastructure Project Investment Evaluation." *Journal of Construction Engineering and Management*, ASCE, 126(3), 227-233.

Table 2: Simulation results

Work Package	5 yr. Fixed Estimate Cost (No Inflation)	5 yr. Fixed Estimate + 3% Inflation	5 yr. R.E. (No Inflation) (Mean)	5 yr. R.E. + Forecasted Inflation (Mean)	5 yr. R.E. + 3% Inflation (Mean)	15 yr. R.E. + Forecasted Inflation (Mean)	15 yr. R.E. + 3% Inflation (Mean)
Land Acquisition	$ 8,533,680.00	$ 8,728,126.97	$ 8,533,680.00	$ 8,641,018.69	$ 8,728,126.97	$ 10,429,176.00	$ 11,429,684.30
Utility Relocations	$ 2,325.60	$ 2,454.24	$ 2,325.60	$ 2,407.67	$ 2,454.24	$ 2,906.02	$ 3,213.89
Detour Lanes	$ 819,840.00	$ 865,189.57	$ 896,359.16	$ 904,812.98	$ 945,389.56	$ 1,122,916.99	$ 1,240,402.05
New Roadway	$ 5,831,448.00	$ 6,154,015.39	$ 6,215,068.16	$ 6,412,024.31	$ 6,565,788.42	$ 7,787,411.98	$ 8,596,883.59
Storm Drainage	$ 3,612,039.19	$ 3,806,610.06	$ 3,697,465.00	$ 3,823,845.11	$ 3,891,688.72	$ 4,610,990.48	$ 5,097,794.14
Fill Placement	$ 2,887,536.00	$ 3,096,609.21	$ 4,277,634.77	$ 5,391,153.82	$ 5,425,960.53	$ 6,606,391.97	$ 7,028,727.97
CN Structure	$ 3,556,275.00	$ 3,747,797.71	$ 3,374,765.73	$ 3,635,517.26	$ 3,765,093.42	$ 4,355,306.43	$ 4,929,017.02
YHT Structure	$11,201,868.80	$11,935,219.98	$10,913,811.45	$11,705,536.85	$12,062,921.45	$ 14,151,983.99	$ 15,802,506.06
Street Lights/Signals	$ 3,176.40	$ 3,574.16	$ 3,176.40	$ 3,445.75	$ 3,574.15	$ 4,159.40	$ 4,680.45
Paving Interchange	$ 375,162.00	$ 422,635.76	$ 410,578.84	$ 446,252.74	$ 462,682.11	$ 544,406.49	$ 606,484.22
Remove Detour Road	$ 245,280.00	$ 277,464.25	$ 316,581.14	$ 348,410.62	$ 353,769.10	$ 411,859.82	$ 463,236.72
Landscaping	$ 951,215.75	$ 1,022,410.87	$ 955,767.08	$ 1,009,258.42	$ 1,037,879.19	$ 1,206,193.54	$ 1,357,614.03
Traffic Management	$ 240,876.00	$ 262,732.35	$ 240,876.00	$ 255,152.32	$ 262,732.35	$ 307,990.58	$ 344,054.10
Total Project Cost	$38,260,722.74	$40,324,840.54	$39,838,089.33	$42,578,836.54	$43,508,060.22	$ 51,541,693.68	$ 56,904,298.52
Variance	-	-	$ 1,377,553.96	$ 1,193,838.58	$ 1,191,478.87	$ 1,413,444.80	$ 1,434,936.43

USE OF DISCRETE EVENT SIMULATION FOR A LONG RANGE PLANNING OF AN EXPEDIDITON SYSTEM

Leonardo Chwif[a], Afonso Celso Medina[b], José Arnaldo Barra Montevechi[c], Marcos Ribeiro Pereira Barretto[d]

[a] Simulate Tecnologia de Simulação
[b, d] Universidade de São Paulo
[c] Universidade Federal de Itajubá

[a] leonardo@simulate.com.br, [b] afmedina@uol.com.br, [c] montevechi@unifei.edu.br [d] mrpbarre@usp.br

ABSTRACT

Discrete Event Simulation is widely used in many areas both in industrial and non-industrial ones. It can be used as an operational tool or as a planning tool. Historically it is most used as a planning tool. Focusing discrete event simulation as a planning tool, generally it is used as a tactical planning tool, and its results should be implemented in days, weeks or months. There are fewer cases that employ discrete event simulation as a long range planning tool. This works is regard a long range planning study of an expedition system by means of discrete event simulation of a factory that produces customized metal sheets (special steels). After the study was performed it was possible to confirm that the initial configuration works for the forecasted growth of flow of the expedition system and that discrete event simulation can also be used as a long range planning tool.

Keywords: Discrete Event Simulation, Long Range Planning, Expedition System

1. INTRODUCTION

Discrete Event Simulation is widely used in many areas such as Public Systems (health care, military), manufacturing, call center, transportation systems, computer system performance and others (Banks et al. 1996). It can be used as an operational tool (e.g. as a factory scheduling tool) or as a planning tool (to validate a new system configuration for example). Historically it is most used as a planning tool, when the simulation project is punctual and is used to support a decision. After they supported a decision they "can be thrown away". But there are also "operational simulation models" in which the model should be used in an ongoing basis, i.e. they are reutilized. Refer to Pidd and Robinson (2007) for a detailed classification of simulation model practices.

Focusing discrete event simulation as a planning tool, generally it is used to validate a concept or to optimize the performance of a system before its implementation. So it is generally used as a tactical planning tool, and its results should be implemented in days, weeks or months. There are fewer cases that

employ discrete event simulation as a long range planning tool, with the results to be implemented in years.

This works is regard a long range planning study of an expedition system by means of discrete event simulation of a factory that produces customized metal sheets (special steels). For confidentiality the name of the company should remain unmentioned. The major concern here is to answer if the proposed expedition system will support the growth of volumes until 2020.

The scope of simulation comprises the entrance of the expedition trucks (there are 3 basic kinds of trucks) inside the border of the factory, the entrance weighting process (at the balance), the loading process, the truck cover process and the exit weighting process (at the balance). The actual expedition system has 1 balance, 1 pit for loading and 1 pit for covering the truck (the configuration of the truck used demands to cover it after the loading process). The proposed system for a 235% growth of the flow would be 1 or 2 balances, 3 pits for loading and 2 pits for covering, and this should be confirmed. The team decided to confirm this project by means of discrete event simulation.

This paper is organized as following: section 2 makes a brief review of simulation methodology; section 3 presents the conceptualization phase and data collection; section 4 describes implementation issues; section 5 deals with the analysis of simulation results and section 6 makes the conclusions.

2. SIMULATION METHODOLOGY

Simulation Methodology or Process, describes the steps of activities that the modeler should perform in order reach a successful outcome. A classic one is described in Law and Kelton (2000) and it is depicted in figure 1.

Briefly, step 1 is problem definition, when it is necessary to define the objectives of the study and other specific issues. Step 2 concern with the data collection (if exists) and the creation of the conceptual model. Step 3 validates the conceptual model, while step 4 implement the model (i.e. create a computerized model). Runs should be made (Step 5) and a validation process should take place (step 6) to guarantee that the

model represents the real system. Experimentation is focus of the steps 7 to 10 with the documentation and presentation of the study in step 10.
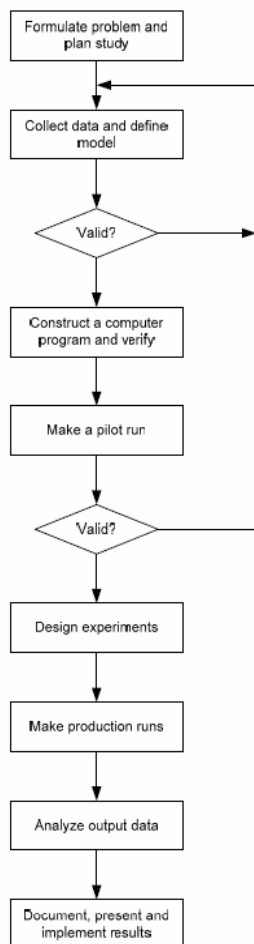

Figure 1: Process of Simulation
(Adapted from Law and Kelton, 2000)

However, in practice, a simulation study does not advance in such a linear and step by step way. In fact there are parallel activities such as implementation of the model and verification, and returns. For instance, the conceptual model or the data collection (or both) should be reviewed if the validity of the model could not be proved. Robinson (2004) confirms this non-linearity presented in simulation studies, despite of the simulation text books shows always linear processes. Another point is that all of this systematic process can be of no use, if the problem is no well understood and identified (Paul et. al 2005).

In this project we used the simulation methodology by applying the following steps: development of conceptual model, data collection, model building (actual), model verification and validation, model building (future) and analysis. It can simply be divided into 3 parts, such as: model conceptualization, model implementation and model analysis. So the in the next tree sections, we will cover each of these macro steps.

3. MODEL CONCEPTUALIZATION AND DATA COLLECTION

So first the conceptual model was discussed with the project team in a meeting of almost 5 hour (the project team was composed by an external simulation consultant, 2 company's engineers and 1 logistical supervisor).

Basically the model took into account the expedition flow of trucks and the flow of raw materials. In this latter case, despite the focus of the study is the expedition trucks, the raw material and administrative material trucks will be also considered in the model, because they share the same resources (balance). However they will have an overall permanency time inside the simulation, instead of the detailed modeling of its logistic process. Therefore the main process was of the expedition trucks and it comprehends the following steps:

1. Entrance of the expedition trucks (there are 3 basic kinds of trucks) inside the border of the factory;
2. Entrance weighting process (at the balance);
3. Travel time to the pit
4. the loading process, at the pit;
5. the truck cover process, at the cover point;
6. Travel time to the balance
7. Final weighting process (at the balance).

As other factors that affects directly the results of the simulation were: the time window for the arrival of trucks, seasonality factor (to consider the concentration of flow at the end of the month), and the growth factor (as described before this would reach 235% by the year 2020). In this model it was also considered the truck limit inside the expedition area, which controls the flow of trucks (if the number of trucks reached this limit, they should wait outside). It was also considered the rain profile, since when it rains, the covering process is made inside the loading pit, thus affecting its utilization. This was considered by defining the probability of rain in each hour trough the day. Regarding the profile of arrivals during the day, we waited to the data analysis to see how it will be modeled.

After defining the conceptual model, the data collection phase took place. In fact as mentioned by Pidd (1996), it is the model that drives the data and not vice versa. So we began to collect only data after the scratch of the initial conceptual model.

This lasted a period of 2 month and the following data was collected: trucks arrival flow (both expedition as raw materials trucks), time for the weighting process (with empty and full loaded truck), total permanency time for raw materials trucks, travel times (from the balance to the pits and vice versa), time for loading and time for covering (considering all kinds of trucks). It was found that the pattern of trucks is almost constant during a day (there was no peak concentration) so it was modeled simply with an exponential distribution.

These raw collected data were processed into a fitting software (Stat:fit) which generated the mathematical distributions. Figure 2 shows some Stat:fit analysis.
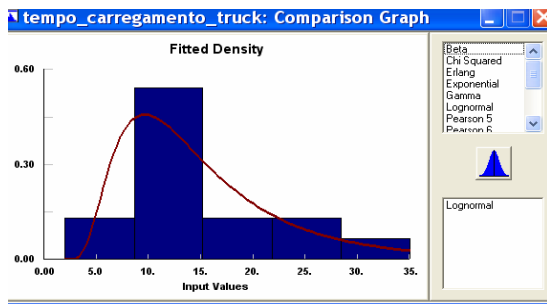
Time to Load a Truck a the Pit (medium Truck)
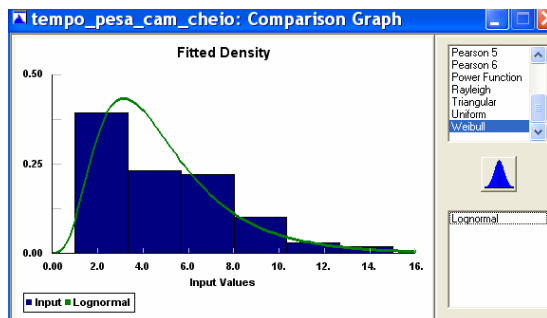
Time to Weight a Loaded Truck

Figure 2-Stat:Fit Analysis

The profile of the rain of São Paulo Region was found in Pezzopane et al. (1995). Table 1 shows the basic rain profile considered, in the months of January and February, that corresponds to the months with the highest pluviometric rate.

Table 1: Basic Rain Profile

Rain Probability	
from 00:00 to 1:00	9.7
from 1:00 to 2:00	11
from 2:00 to 3:00	9.4
from 3:00 to 4:00	8.1
from 4:00 to 5:00	9
from 5:00 to 6:00	9
from 6:00 to 7:00	7.7
from 7:00 to 8:00	6.8
from 8:00 to 9:00	6.1
from 9:00 to 10:00	4.8
from 10:00 to 11:00	5.5
from 11:00 ao 12:00	6.8
from 12:00 a 13:00	8.1
from 13:00 to 14:00	12.3
from 14:00 to 15:00	14.5
from 15:00 to 16:00	16.8
from 16:00 to 17:00	14.5
from 17:00 to 18:00	16.8
from 18:00 to 19:00	18.7
from 19:00 to 20:00	19.7
from 20:00 to 21:00	18.1
from 21:00 to 22:00	15.8
from 22:00 to 23:00	14.2
from 23:00 to 24:00	12.3

After the data was collected, we discussed also some considerations:

- Loading resources (man, forklift trucks) will be modeled without constraint, i.e., they will always be available at the pit.
- The travel time from the balance to the pit can be considered equal to the return (from pit to balance)
- The weighting time does not depend on the truck but on the state of the truck (empty or full)
- For the sake of the analysis it was considered the typical mix of trucks, i.e:22% of small truck, 62% medium truck and 16% large truck.

Next step was model building. This was done with the aid of Simul8 Simulation Software; which can be seen in next section.

4. MODEL BUILDING AND V&V

So the model of the actual configuration was developed (1 balance, 1 loading pit and 1 covering pit) in Simul8 Simulation Software vs. 2007. Since this model is not complicated it took around 8 hours to build it all. Figure 3 shows the initial configuration of the model, while table 2 shows the input data, which can be customized by the user. In this case all the times inside the model was migrated from Stat:fit to Simul8, but we made them not customizable for the client (they cannot alter its value).
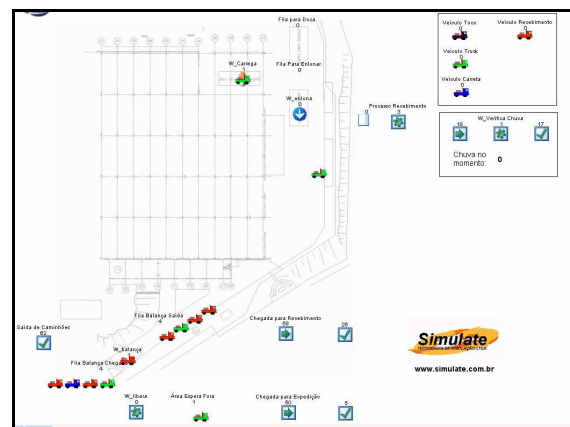
Figure 2 – Initial Model

Next phase was to validate its results against real world. It was set up a validation meeting with the team which last 3,5 hours. All the performance measures such as waiting times, queues size, utilizations were validated except one – the waiting time for loading. The simulation was shown that the truck waits less than in reality. So it was necessary to recollect the data, specially this measure to verify what went wrong. After one and half week of recollecting data, two points was discovered: that the initial queuing time was not measured and calculated by difference and that exist some wait due to other factors that the loading time.

These factors are many, including the time spend by the driver to make other duties, time to wait for the forklift trucks, but since we hypothesize that there will not be any constraints at the pit in the future configuration, these "waiting times" was not considered for the model. At this point the model was validated and we can proceed to the analysis phase. So the initial model was expanded to account for the extra number of pits and cover area. This is shown in figure 3.

Table 2 – Input Data Table

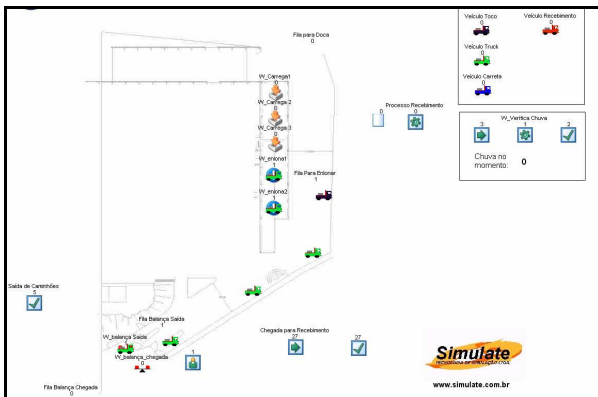| Simulation Data		
Expedition Truck Flows		
Average Truck per Hour	1.6	
Time Window (from-to)	2	23
Raw Material Truck Flow		
Average Truck per Hour	1.7	
Time Window (from-to)	6	21
Sazonality Factor	1.2	
Growth Factor	1	
Mix	%	
toco (1)	22	
truck (2)	62	
carreta (3)	16	
Truck Limit inside pit area	4	
Enable Rain?	0	
Rain Profile		
see table 1		

Figure 3 – Future Model

5. ANALYSIS
For the results analysis the main measures of performances adopted were:

- Minimum, Average and Maximum Time in System;
- Average and Maximum queue Size and queuing time for the balance entrance queue;
- Average and Maximum queue Size and queuing time for the balance exit queue;

- Average and Maximum queue Size and queuing time for the wait outside the factory (if the limit is reached);
- Average and Maximum queue Size and queuing time for the pit queue;
- Average and Maximum queue Size and queuing time for the cover area queue;
- Balance utilization;
- Pit utilization;
- Cover area utilization.

Figure 4, depticted these main results of one analyis, providing the confidence interval for 20 replications.
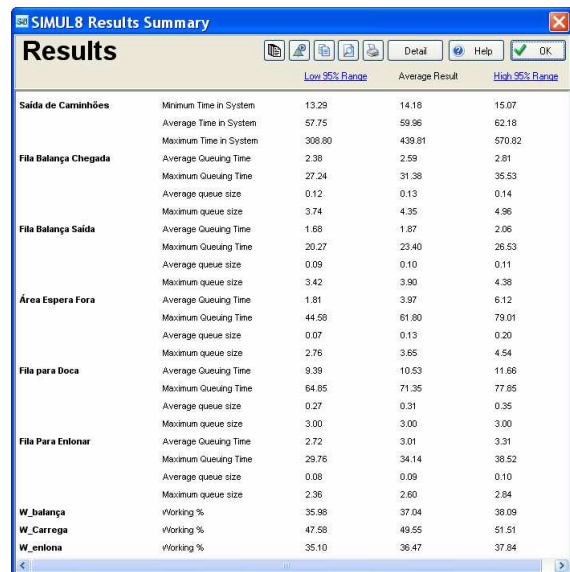

Figure 4 – Simulation Results

The first analysis performed was a saturation analysis, and this tries to answer the following question: "Considering the actual configuration (1 loading pit and 1 covering pit), how much increase of traffic does the system support". We found that the actual configuration can be sustained for 3 years, and after that the performance of the expedition will degrade. Then the model was expanded to consider up to 3 loading pits, up to 2 covering pits and up to 2 balances (see Figure 3).

Several scenarios was build to account to the combination of the following variables: increase of flow, number of balances, number of loading pits, number of covering pits, presence or absence of rain, rain profile (normal or heavy) and mix of trucks (small, medium and big). Since the Basic Rain Profile (which we called normal) affects little the utilization of the pits, we decided to create a scenario with "heavy rain", that increases the probability of rain within a given hour of the day.

The results showed that with the growth of flow forecasted to 2020, it was necessary 2 balances, 3 loading pits and 2 covering pits (exactly the configuration initially proposed). In this case, if we not take into consideration some exceptions, the system performed very well - utilization of pits is around 50%

and waiting times is around 1 min (medium) and 30 minutes (maximum). Only in some very particular circumstances when there are heavy rain or there are a concentration of arrival of big trucks the system become saturated, but in normal operation (98% of the time), the system is well sized.

6. CONCLUSIONS AND FINAL TAUGHTS

This paper covers the study of the expansion of an expedition system of an industry that produces metal sheets up to 2020. This study was conducted by means of discrete event simulation. After applying the simulation methodology, it was possible to assure that the initial configuration will work for the forecasted growth of flow of the expedition system, confirming the engineers' expectations. It is also confirmed that discrete event simulation can also be used as a long range planning tool.

The model was given to the company and they can run it on the Simul8 Run Time Version – Simul8 Viwer. So they can perform further analysis beyond the ones covered in this study.

REFERENCES

Banks, J, Carson, J.S. and Nelson, B.L. 1996. *Discrete-Event System Simulation*, 2nd edn. Upper Sadle River, NJ: Prentice-Hall.

Law, A. M., and Kelton W. D. 2000. *Simulation Modeling and Analysis*. Third edition. Boston: McGraw-Hill.

Paul, R. J., Eldabi, T., Kuljis, J., Taylor, S.J.E. Is problem Solving, or Simulation Model Solving, Mission Critical, *Proceeding of the 2005 Winter Simulation Conference*, pp. 547-554. December 4-7. Orlando-FL (USA).

Pezzopane, J.E.M, Sentelhas, P.C, Ortolani, A.A. Moraes, A.V.C. 1995. Hourly Rainfall Pattern for three sites in the State of São Paulo, Brazil: A Subsidy to Agricultural Practices Planning: Sci. Agric. Piracicaba, 52(1):70-77, jan/abr.

Pidd, M. 1996. Five simple principles of modelling, *Proceedings of the 1996, Winter Simulation Conference*, pp. 721-728. December 8-10, Coronado (CA)- USA.

Pidd, M. and Robinson, S. 2007. Organizing Insights into Simulation Practice, *Proceeding of the 2007 Winter Simulation Conference*, pp. 771-775. December 9-12. Washington D.C (USA).

Robinson, S. 2004. Simulation: *The practice of Model Development and Use*, John Wiley & Sons, Ltd.

AUTHORS BIOGRAPHY

LEONARDO CHWIF graduated in Mechanical Engineering (Mechatronic Specialization) in 1992 at the University of Sao Paulo and achieved his M.Sc. degree in 1994 and his Ph.D. in Simulation in 1999 from the same University. He spent a session at Brunel University as a research visitor at the Centre for Applied Simulation Modelling. Upon graduation Dr. Chwif joined the Brazilian branch of Mercedes-Benz truck manufacturer and then he joined the Brazilian branch of Whirlpool Corporation. Currently he is C.E.O of Simulate Simulation Technology, a consultancy Simulation Company located in São Paulo, Brazil. His e-mail address is <leonardo@simulate.com.br>

AFONSO CELSO MEDINA graduated in Naval Engineering (Transports Specialization) in 1988 at the University of Sao Paulo and achieved his M.Sc. degree in 1996. Currently he is a researcher of the Naval and Ocean Engineering Department of the University of Sao Paulo. Mr. Medina also is a associate of Simulate Simulation Technology. His research interests include simulation of transportation systems and facility location problems. His e-mail address is <afmedina@uol.com.br>

JOSÉ ARNALDO BARRA MONTEVECHI is a Titular Professor of Instituto de Engenharia de Produção e Gestão at Federal University of Itajubá, in Brazil. He holds the degrees of Mechanical Engineer from Federal University of Itajubá and M.Sc. in Mechanical Engineer from Federal University of Santa Catarina, and Doctorate of Engineering from Polytechnic School of University of São Paulo. His research interest includes Operational Research, Simulation and Economic Engineering. His e-mail address is <montevechi@unifei.edu.br>.

MARCOS RIBEIRO PEREIRA BARRETTO graduated in Electronic Engineering in 1983. He achieved his M.Sc. degree in 1988 and his Ph.D. in 1993, both from the University of Sao Paulo. He is a professor and researcher of Mechatronic Departament from University of São Paulo. His email is <mrpbarre@usp.br> .

SUPPLY CHAIN PERFORMANCE UNDER TRANSIENT DEMAND INCREASES: A CASE STUDY SUPPORTING SUPPLY CONTRACT NEGOTIATION

Ivor Lanning[a], Cathal Heavey[b]

[a] [b] Enterprise Research Centre, University of Limerick, Limerick, Ireland

[a]ivor.lanning@ul.ie, [b]cathal.heavey@ul.ie

ABSTRACT

The advent and continued growth in the electronics and EMS (electronics manufacturing services) industry has precipitated research in this area in recent times. This paper addresses issues surrounding the development of supply contracts in the resulting outsourcing arrangements. A case study is presented which investigates exceptional occurrences in the form of transient demand upturns with little or no forewarning. The EMS provider wished to examine inventory holding and service level issues in this environment. The case study is described in some detail and initial experimentation is carried out using discrete event simulation to analyze how this demand, coupled with problematic lead time and sub-supplier capacity constraints affects the system. Results indicate that parameter interactions are non-trivial. In addition useful information from simulation modeling can aid in negotiations of supply contracts where OEMs wish to maintain service levels under reasonable supply chain costs.

Keywords: Simulation, Inventory Control, Transient Demand, Supply Contracts.

1. INTRODUCTION

In environments where production outsourcing exists, OEMs need to have assurance of supply. Along with high quality and low prices, meeting demand on time and in full has become a way of gaining a competitive advantage in a growing electronics market. EMS providers currently make up a major link in the electronic goods supply chain. The EMS market grew by 13.8% in 2005, which was over twice the growth experienced in the worldwide electronics market (6.1%). Electronics assembly, valued at $810 billion in 2005 has been predicted to reach a value of $1.1 trillion by 2010. If this growth was accurate the EMS industry would grow from $190 billion in 2005 to $328 billion in 2010, (DATAMONITOR 2006). This growth has precipitated investigations in the literature into how EMS providers operate such as (Zhai, Shi et al. 2007).

Here, the issue of supply assurance during occurrences of transient demand was addressed with the aim being to support contract negotiation between an OEM and contract manufacturer/EMS. Specifically, transient upturns in end-item demand with little or no

warning experienced by an OEM and relayed to an EMS provider is addressed. Supply contracts are used to define the operating conditions that the participating members agree to and for these conditions they set out the acceptable supplier service levels. (Das and Abdel - Malek 2003), note that a supply contract is the key document defining the parameters under which the parties operate and it is the robustness and flexibility of such documents that will determine the success and stability of the partnership. A contract manufacturer faced with stable demand requirements from an OEM with given assurances of supply from its suppliers can agree to meet service level requirements laid out in the contract. Under these conditions the OEM is reasonably assured of supply. However, the occurrence of unforeseen transient behaviour in the supply chain conditions can affect this stability. Through collaboration the OEM can work with the contract manufacturers and sub suppliers to attain the specifications of the supply contract as is the case with the computer manufacturer Dell. Dell successfully practices heavy reliance on outside suppliers and contract manufacturers Kraemer and Dedrick 2005). It is also noted that Dell invest heavily in advanced manufacturing and logistics capabilities incorporating advanced information systems to help use information as a substitute for inventory, with positive market results.

Supply chains are subject to numerous factors that can affect performance. (Cigolini, Cozzi et al. 2004), is one paper that suggests a framework with which, managers can decide what environmental characteristics to attend to when designing their supply chain management policy. While some are controllable by specific participants in the supply chain, others tend to lie of their control, such as demand (disregarding seller initiated disruptions such as discounts, for example). Such factors are labelled environmental conditions. The demand parameter constitutes an important input to models in general in supply chain management as it is almost ubiquitous regardless of the kind of modelling approach. (Bookbinder and Cakanyildirim 1999), applied the simplest form of demand input in a model which investigated random lead times and expedited orders in (Q, r) inventory systems. This model consisted of a two-echelon system with a known and constant demand rate. This is generally referred to as a constant

deterministic demand parameter and is simply a point estimate of periodic demand. The other main form of demand modelling is stochastic demand as employed by (Ayanso, Diaby et al. 2006). Here, demand in each period consists of a sample taken from an empirical distribution with chosen parameters. Further to this is non-stationary demand which can be both deterministic and stochastic. In this case the point estimate or mean of the distribution can shift from period to period. This is often seen when there are seasonal variations in the base level of demand for a product. In addition to this we have found some papers however that either model demand or at least consider the importance of it with greater detail based often on actual market information across a number of industries.

(Hwarng, Chong et al. 2005), undertook a large study on a real, complex supply chain. One of the aims was to evaluate the impact of simplifying demand assumptions under different supply chain configurations. Raw data collected over a period of more than three and a half years was used, taken from a pharmaceutical company. The demand data for a single product was used to derive the demand inputs for the model. Distribution fitting software, Stat::Fit was used to determine the most appropriately fitting distributions and parameters for the data and these were input to the model. Chi-Square and Kolmogorov- Smirnov values were given to indicate the goodness of fit of the distributions to the data. The results of the experiments indicated a difference in output when distributions based on case study data were used as opposed to using empirical distributions. Higher backlogs resulted of up to approximately 5% along with average stock increases up to 70%. However because of the specificity of the paper on a supply chain with quite a highly complex structure and unique data the results cannot readily be applied to many other supply chains. The approach was a singular case study, without an attempt to generalize to other scenarios. This was deliberately done to serve the purpose of highlighting the dangers of arbitrarily choosing levels of stochastic inputs.

Another demand input method that has received minor research in comparison to the two main approaches is detailed in (Bartezzaghi, Verganti et al. 1999a). They investigated the ability of three forecasting methods to deal with irregular and sporadic - termed lumpy - demand. The forecasting techniques examined were; exponentially weighted moving average, early sales and order over planning. The latter two methods make use of future demand information rather than just historical data and are described in some detail in the paper. Performance of each of the methods was judged on the mean absolute deviation (MAD) of the forecast error. The results helped form a framework whereby each forecasting method was given a domain of applicability relating to the prevailing conditions of lumpy demand.

Transient demand characterized by exceptional, unpredicted increases (pulses) can result in massive reductions in service levels, i.e. meeting demand on

time in full (OTIF). In addition, costs to the system due to backorders can dramatically increase and persist long after the transient demand has passed. While demand input is quite comprehensively investigated in supply chain management literature, demand of this nature, has been previously unfound. The demand pulses which characterized the system can be defined by 3 main elements as shown in Figure 1; magnitude, duration and shape. Time is depicted in weeks unless otherwise noted.

Figure 1 illustrates that the pulse can be any real positive multiple of the nominal demand in magnitude added to the nominal forecasted demand and may also extend over an interval of a number of weeks thus extending duration. Figure 2 further illustrates that the pulse can take varied forms. Shown here are two examples which equate to; (a) a set of 4 spikes of equal magnitude and, (b) a 'ramp' pulse of increasing amounts over a time of 4 weeks. A constant deterministic figure, nominal demand derived from the OEM forecast made up the demand in all the other periods of the model. In addition, there may be some kind of positive demand lead time that gives advance notice of the pulse. This advanced demand information (ADI) can positively affect supply chain performance to varying degrees as noted by (Bourland, Powell et al. 1996), depending on how it is utilized.
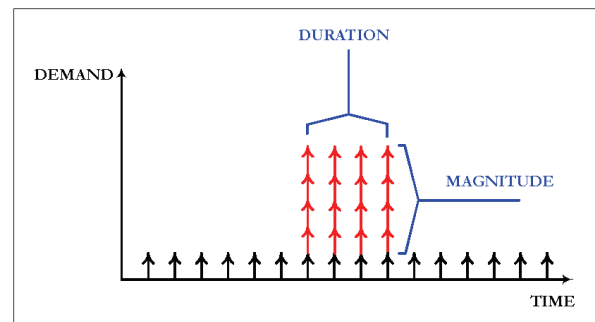
Figure 1: Demand Characterization

Figure 2: Additional Demand Pulses

These demand pulses are a form of unexpected exceptional events. As noted by (Hwang and Tang 2004), exceptions can be of two types;

- Expected,
- Unexpected.

An expected exception is known to occur occasionally or periodically and can be handled routinely with measures to facilitate them put in place at design time. However, other exceptions such as the demand pulses defined and applied in this

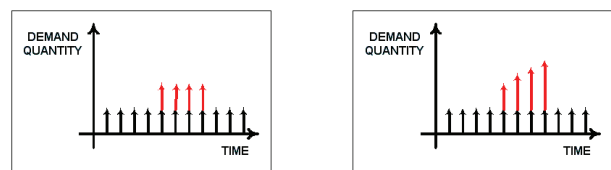experimentation are a result of unforeseeable changes in the supply chain environment. This demand is experienced by a production supply chain based on a case study model of an EMS provider. The segment of the supply chain in question consisted of the OEM experiencing end-item demand passed on to the EMS who in turn sourced components for manufacture from a range of sub-suppliers. The problem area in the supply chain involved the assembly of an electronic medical device for a large OEM from various components for shipping to the OEM for final configuration for individual end-item markets. Figure 3 illustrates the supply chain and the focal point of the investigation. A turnkey arrangement contract existed between the OEM and contract manufacturer. This is defined by (Kim 2003), as an arrangement where the OEM pre-approves the sub-supplier list for the contract manufacturer from whom the components for the product must be sourced. Therefore, supplier selection was not available to the contract manufacturer in this supply chain similar to (Smith, Agrawal et al. 2003), for example. This is not uncommon in the area of life sciences related products where strict legislation regularly exists for both design and production.
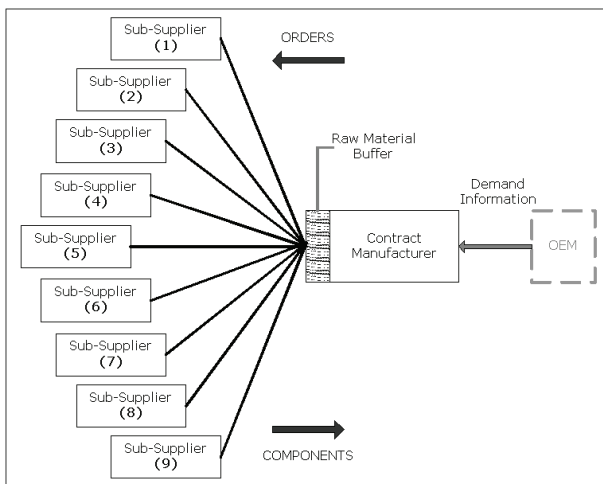

Figure 3: Case Study Supply Chain

This forced the contract manufacturer to operate with a set of sub-suppliers not of its own choosing. Thus, meeting OEM demands became a more complex issue for the contract manufacturer. The contract manufacturer was forced to plan its production around the production capacities and lead times of its sub suppliers. For these reasons it was found that the input factors of sub-supplier capacity and component procurement lead time also impacted heavily on the performance of the supply chain.

The study of lead time variability in supply chains is almost as popular as that of demand. There is an important segregation that must be made in the study of lead times in supply chain literature. Many papers such as (Chu, Yang et al. 2005), view lead times as an output from a supply chain system and that an aim of supply chain management should be to minimize lead times. In addition (Holweg, Disney et al. 2005), note that long

lead-times in vehicle manufacture in Europe are a vital hindrance in the increased production of build-to-order products. For this reason many papers look toward reducing lead-times thus viewing them as an output of the system. This is the case in (de Treville, Shapiro et al. 2004), which aims at improving supply chain performance by reducing lead-times.

However, lead-times are also inherent in supply chain systems. For example, while production lead-times can be reduced or affected by a number of means such as production lot size for example (Ben-Daya and Hariga 2004), transportation lead-times faced by companies often cannot be attended to in a similar fashion. Similar to (Waters 1992), (Chu, Yang et al. 2005), clarify that lead times are made up of a number of smaller components each with differing attributes and each having differing abilities of being reduced. Lead time components such as transportation periods via sea shipping methods for example, fall under the previously referenced term of environmental conditions in the supply chain. Essentially, short of advances in the field of transporting physical stocks from one location to another, it is not possible to affect these lead times and therefore firms must incorporate them into their planning process.

There were 13 components noted as being key to the manufacture of the product in the case study as indicated in Table 4.1. Sub-supplier production capacity is expressed as a percentage of the nominal forecasted demand. In general, components for the device were classified as either *A B* or *C* depending on criteria regarding their importance.

Table 1: Case Study Lead Times and Capacities

Item	*a*	*b*	*c*	*d*	*e*	*f*
Lead Time (Wks)	8	7	7	13	12	4
Capacity %	119	167	167	248	159	167

g	*h*	*i*	*j*	*k*	*l*	*m*
8	12	22	20	2	2	8
110	222	120	248	167	167	248

These criteria covered such aspects as component price, contribution to the product and also specificity to particular products/ease of redeployment to other products. Class-*A* components were the most important and were the only component type considered worth investigating in the context of case study according to the contract manufacturer. At the time of initiating the case study, the contract manufacturer actually had little information regarding production capacity, even from the sub-suppliers of its 13 class-*A* components associated with this product. Within these 13 components the associated lead times were as low as 2 weeks and as high as 22 weeks with a broad dispersal in between. Such varied lead times complicated coordination of shipments and buffer allocations at the contract manufacturer. While lead-times were known and treated as deterministic in the system they were sub

supplier dictated. Lead times between the sub supplier and supplier were defined as the interval between the placing of an order with the sub-supplier until the shipment arrives at the supplier, similar to (Oke and Szwejczewski 2005).

2. SIMULATION MODEL

The information from the case study was used to build a discrete event simulation model to analyze the supply chain performance. The supply chain model was built using an object orientated simulation package, eM-Plant. There were a number of reasons for choosing this piece of software. eM-Plant models in an object oriented framework which allows for the easy replication of objects such as the similar but not identical sub-suppliers in the supply chain in question. The ability to use the ActiveX object to link to MS Excel allowed for easy transfer of the output data for further analysis. Finally, the flexibility of manipulating the standard objects' behaviour using eM-Plant's programming language (SimTalk) was fundamental to the choice of the package to model the supply chain. SimTalk is a C based programming language that allows the model builder to customize the functions and interactions of the objects which make up the model. In order to model the supply chain, interaction with the contract manufacturer was undertaken including;

- Meetings both on and off-site with key personnel at the contract manufacturer facility,
- The general plant manager was the initial point of contact in the company and provided much of the background information to the case-study setting,
- The planning and inventory manager then became the project specific point of contact. This person was in a position to supply company data and information on inventory movement.

The simulation model consisted of 9 sub-suppliers supplying non-competing components into a single contract manufacturer, as illustrated in Figure 3. The OEM involvement in the model was restricted to a supply of demand information received by the contract manufacturer; there was no processing at the OEM. In addition it was assumed that there was a set of inexhaustible external suppliers to the sub-suppliers. 13 class-*A* components were supplied to the contract manufacturer. One sub-supplier furnished 5 component types and the remaining sub-suppliers each supplied 1 type. Each component type was subject to an individual lead-time as defined above. Buffer minimum requirement levels applied to all component types regardless of the associated lead times or capacities. Assumptions regarding some of the processes in the supply chain had to be made in lieu of more detailed production and supply chain information;

- It was assumed that 100% of the components shipped to the contract manufacturer arrived in usable quality,

- Production at the contract manufacturer could not exceed the minimum component in-stock quantity in any given period. Production of an end-item unit only commenced when sufficient quantities of all components were available,
- Backorders occurred at all locations,

Based on the problem formulation above a number of objects were developed in the simulation environment. Figure 4 shows the main graphical user interface (GUI) for the model at the highest level.
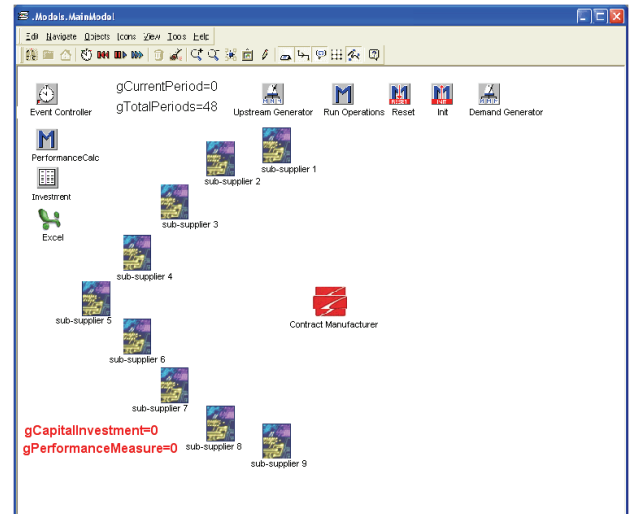

Figure 4: Simulation Model Main GUI

In the same way that the main model interface shown above, is a collection of objects located within a frame, each of those objects (sub-suppliers and contract manufacturer) are similarly a frame with a collection of objects within them. By double clicking on one of these objects a separate frame is opened and access is given to the individual objects that are constituents of it.

Figure 5 illustrates the contract manufacturer frame. Its behaviour as a player in the supply chain is defined by the objects within it, and their relationships to each other and the objects within the sub-suppliers etc. The contract manufacturer has three main functions;

- demand is experienced and orders are allocated to the master production schedule (MPS),
- components are received and added to the "instock" (raw material) buffer,
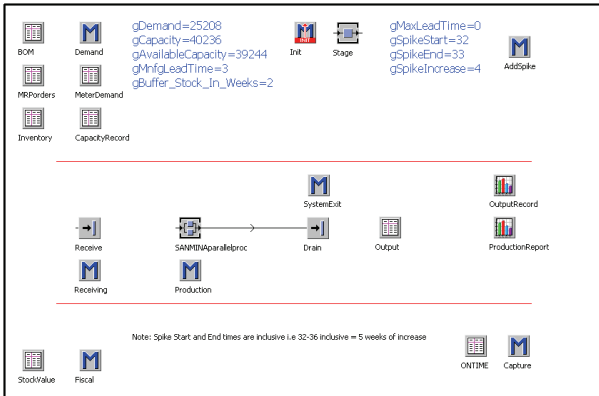- production of end-item units takes place.

Figure 5: Simulation Model Contract Manufacturer Object

Historical information was gathered over time in each simulation run regarding production, backorders and costs at the contract manufacturer facility. At the end of a simulation run the information was collated for extraction to MS Excel. This was carried out using the code in the *PerformanceCalc* and *Excel* objects shown in Figure 4.

The sub-suppliers to the model were all of a similar construction to each other but had attributes that differentiated them from each other. These differentiating attributes were the production capacity, the lead time to the contract manufacturer and the component type produced. In the case of one of the sub-suppliers they in fact supplied multiple components as described previously. The replication functionality of the simulation tool enabled us to generate all nine sub-suppliers relatively easily from one parent object. Figure 6 depicts an example of a sub-supplier. The sub-suppliers produced the components used to make the end-item product. Orders were processed based on the MPS requirements generated at the contract manufacturer. Production took place based on these orders with allowances for production capacity. Backorders in any given period i were added to demand for period $(i + 1)$. However, as with all backorders in the system there was no differentiation between freshly generated demand and outstanding backorders. The cumulative total of required parts was the effective demand figure. This did not affect the model as each sub-supplier only fed the single contract manufacturer and therefore prioritization was not an issue.
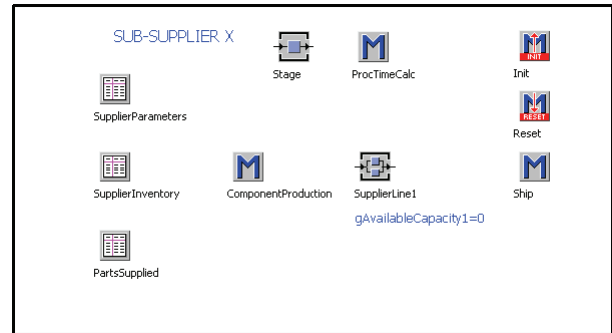

Figure 6: Simulation Model Sub-Supplier

The sub-supplier object tracks the list of part types supplied by the particular sub-supplier. In addition the functions of component production and shipping are handled and the historical inventory and cost data is recorded in the entity for results. Similar to the contract manufacturer the *PerformanceCalc* and *Excel* objects are used to extract the information from the individual sub-suppliers at the end of each simulation run. The functions of the sub-supplier can be surmised below;

- Orders for the period are picked up from the contract manufacturer MPS and backorder requirements if any are added resulting in the effective demand for the period,
- Production is launched equal to effective demand, production capacity allowing,
- Inventory and backorders are updated,
- WIP that has completed its cycle of processing and shipping triggers the release and is moved into the raw material buffer (instock) at the contract manufacturer.

The number of periods defining the model run-length was set by a global variable in the main model object. In addition there was code which initialized the model at the beginning of each run. All of the experimental data for each run (scenario) of the model was held in an internal database within the simulation model. Before each run was made the particular inputs were initialised in the model, setting the buffer levels, lead times and capacities. The demand was then created for the entire simulation run, based on nominal demand as detailed earlier. A periodic generator governed the triggering of each of the actions carried out in the model. A group of code "Run Operations", (shown in Figure 4) then called the subsequent code in their order of occurrence. Finally a "Reset" function ensured that the model was essentially cleaned at the end of a run in preparation for the initialization of the next run.

Part of the initialization regarded the warm-up period for the model, illustrated with a simplified example in Figure 7. Orders to the sub-supplier with the longest lead time began in period 0 and the entrances of orders were then staggered for the other components based on the lead time. Thus, the first order of components all arrived at the contract manufacturer at the same time. At this point the model was coded to

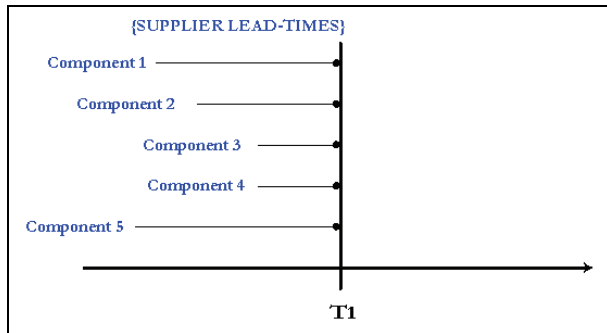being production at the contract manufacturer and a steady supply of components then arrived period by period.

Figure 7: Forward Planning Vs. Lead Times

2.1. Performance Measures

As stated by (Kaipia and Tanskanen 2003), the only meaningful service level is from the retailer to its customers because this is what measures the performance of the supply chain as a whole. Because of the perceived importance of service level in the eyes of the OEM, the contract manufacturer needed access to information on backlogs in the system. The backlogs were created when backorders occurred. A service level metric was used to judge the system performance that was developed in conjunction with the plant manager at the contract manufacturer.
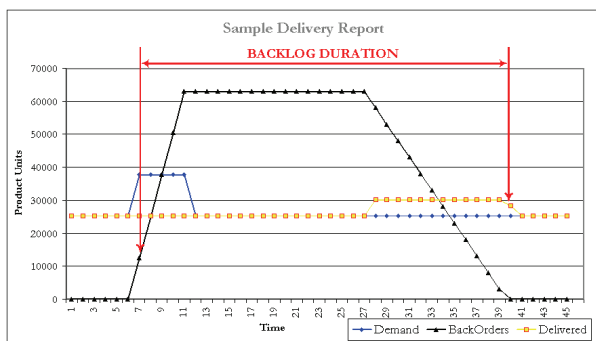

Figure 8: Definition: Backlog Duration

Figure 8 illustrates the key performance indicator (KPI), 'Backlog Duration'. This is defined as the interval in time from the point where backorders in the system appear as a result of the demand pulse to the point where the backorders return to zero. It is anticipated that the backlog duration is affected by component procurement lead times, sub-supplier capacities, buffer minimum requirements and the form of the demand pulse.

3.   CASE STUDY EXPERIMENTATION

The case study experiments were based on the numerical data supplied by the contract manufacturer for the current supply chain configuration. The first requirement of the contract manufacturer was a simple one; to examine the effect of various minimal buffer requirements on backlog duration.

3.1. Case Study Results

The results corresponding to the four demand input scenarios across the buffer sizes are shown in Figures 9 to 12. In scenario A – Figure 9, when the supplier was faced with a demand pulse of 1 week in magnitude in addition to the regular demand, the absence of buffer stock resulted in all required goods backlogged not being outputted for 29 weeks. However, a minimal buffer requirement of just 1 week of components ensured that the delay was reduced to the minimum length of time, equating to the production lead-time at the supplier, 3 weeks.

Scenario A	Spike = 1 Week	
Buffer Size in Weeks	Investment Value	Backlog Duration
0	0	29
1	948,554	3
2	1,897,108	3
3	2,845,663	3
4	3,794,217	3
5	4,742,771	3
6	5,691,326	3

Figure 9: Scenario A

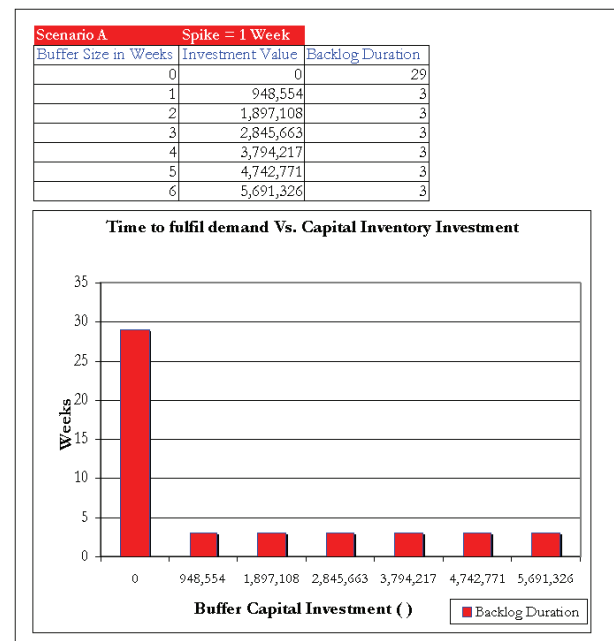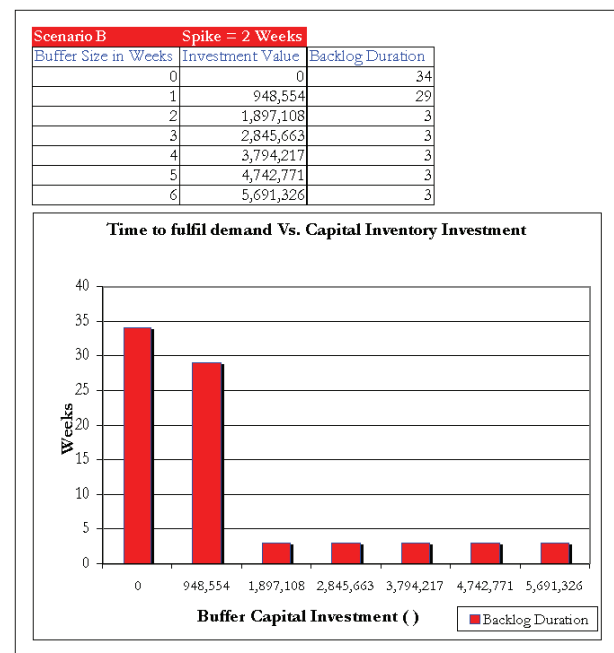| Scenario B	Spike = 2 Weeks	
Buffer Size in Weeks	Investment Value	Backlog Duration
0	0	34
1	948,554	29
2	1,897,108	3
3	2,845,663	3
4	3,794,217	3
5	4,742,771	3
6	5,691,326	3

Figure 10: Scenario B

Minimal buffer requirements greater than 1 week resulted in similarly reduced backlogs but simultaneously resulted in over inventory and thus unnecessary costs at the contract manufacturer echelon. Most notable in scenario B – Figure 10 was the only marginal increase in backlog duration due to a doubling of the demand pulse.

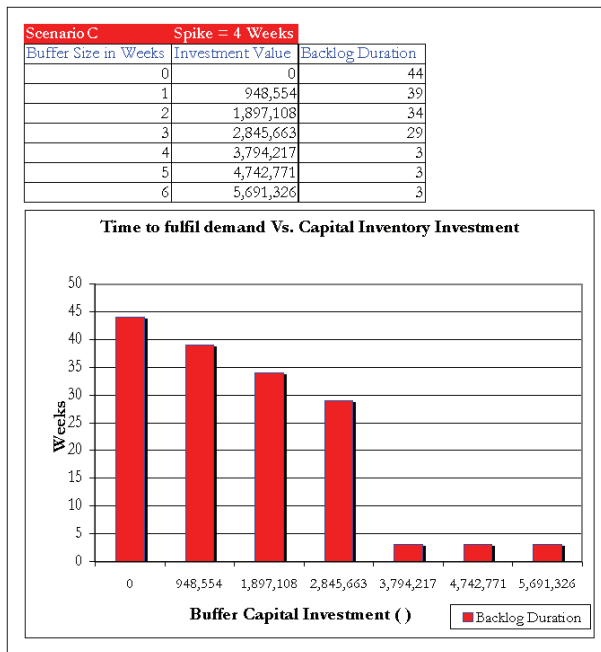| Scenario C	Spike = 4 Weeks	
Buffer Size in Weeks	Investment Value	Backlog Duration
0	0	44
1	948,554	39
2	1,897,108	34
3	2,845,663	29
4	3,794,217	3
5	4,742,771	3
6	5,691,326	3

Figure 11: Scenario C

While a 1-week pulse created a backlog duration of 29 weeks, shown in Figure 10, the 2-week pulse only marginally increased the backlog by 5 weeks. This trend is similarly seen across the other scenarios depicted in Figures 11 and 12.
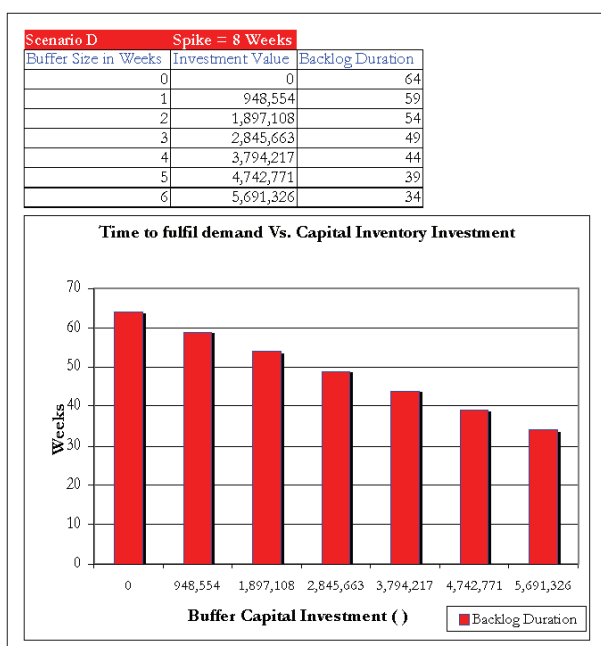
Scenario D	Spike = 8 Weeks	
Buffer Size in Weeks	Investment Value	Backlog Duration
0	0	64
1	948,554	59
2	1,897,108	54
3	2,845,663	49
4	3,794,217	44
5	4,742,771	39
6	5,691,326	34

Figure 12: Scenario D

It was clear from the data outputted in the case study that for every week's worth of inventory held, the backlog duration was reduced by 5 weeks under the demand impulse. This was most notably illustrated in Figure 12. The results also indicate that regardless of buffer levels, sub-supplier production capacities or lead times; if the contract manufacturer was production capacitated, this would override any measures taken to reduce the backlog durations. This was intuitive but important for further experimentation when assessing the interactions of the various input parameters.

4. CASE STUDY CONCLUSIONS

The results of the case study indicated potential benefits of modelling supply chain environments at the contract negotiation stage, specifically incorporating negotiations surrounding transient demand upturns or other exceptional occurrences. Without proper precautions, even the small demand pulses can lead to inordinate backorder intervals. The case study provided a snapshot of a multi sourced supply chain under a set of input parameters taken from a segment of an industrial supply chain and in this respect was quite case specific. With fixed capacities and lead times it was evident that the increasing buffers led to linear reductions in the backlog durations to the system. After examining the results in detail and taking the replenishment rates at the buffers into account, there were found to be non-trivial interactions between capacities and lead times in conjunction with minimum buffer quantities.

Future work has been carried out on the system incorporating more detailed input and output analysis. Costs have been expanded upon to include actual inventory costs in the system over time coupled with backorder costs between the contract manufacturer and the OEM. In addition, stochastic demand and lead time inputs have been considered in the supply chain. This includes the use of previously tested empirical distributions from literature and also the analysis of real data from the supply chain. This input data has been analysed using the statistical software package MINITAB to uncover possible trends in demand data and/or the prevalence of certain empirical distributions when modelling demand data. Work in this area is ongoing.

REFERENCES
Ayanso, A., Diaby, M., et al. (2006). "Inventory rationing via drop-shipping in Internet retailing: A sensitivity analysis." *European Journal of Operational Research* **171**: 135-152.

Bartezzaghi, E., Verganti, R., et al. (1999a). "A simulation framework for forecasting uncertain lumpy demand." *International Journal of Production Economics* **59**: 499-510.

Ben-Daya, M. and Hariga, M. (2004). "Integrated single vendor single buyer model with stochastic demand and variable lead time." *International Journal of Production Economics* **92**: 75-80.

Bookbinder, J. H. and Cakanyildirim, M. (1999). "Random lead times and expedited orders in (Q, r) inventory systems." *European Journal of Operational Research* **115**: 300-313.

Bourland, K. E., Powell, S. G., et al. (1996). "Exploiting timely demand information to reduce inventories." *European Journal of Operations Research* **92**: 239-253.

Chu, P., Yang, K.-L., et al. (2005). "Improved inventory models with service level and lead time." *Computers and Operations Research* **32**: 285-296.

Cigolini, R., Cozzi, M., et al. (2004). "A new framework for supply chain management: conceptual model and empirical test." *International Journal of Operations and Production Management* **24**(1): 7-41.

Das, S. K. and Abdel - Malek, L. (2003). "Modelling the flexibility of order quantities and lead-times in supply chains." *International Journal of Production Economics* **85**: 171-181.

DATAMONITOR (2006). Company Profile Sanmina-SCI Corporation.

de Treville, S., Shapiro, R. D., et al. (2004). "From supply chain to demand chain: the role of lead time reduction in improving demand chain performance." *Journal of Operations Management* **21**(6): 613-627.

Holweg, M., Disney, S. M., et al. (2005). "Towards responsive vehicle supply: a simulation-based investigation into automotive scheduling systems." *Journal of Operations Management* **23**: 507-530.

Hwang, S.-Y. and Tang, J. (2004). "Consulting past exceptions to facilitate workflow exception handling." *Decision Support Systems* **37**: 49-69.

Hwarng, H. B., Chong, C. S. P., et al. (2005). "Modelling a complex supply chain: understanding the effect of simplified assumptions." *International Journal of Production Research* **43**(13): 2829-2872.

Kaipia, R. and Tanskanen, K. (2003). "Vendor managed category management - an outsourcing solution in retailing." *Journal of Purchasing And Supply Management* **9**: 165-175.

Kim, B. (2003). "Dynamic outsourcing to contract manufacturers with different capabilities of reducing the supply cost." *International Journal of Production Economics* **86**: 63-80.

Kraemer, K. L. and Dedrick, J. (2002). "*Dell Computer: Organization of a Global Production Network - http://www.crito.uci.edu/GIT/publications/ pdf/dell.pdf*." Centre for Research on Information Technology and Organizations Retrieved 01/11/2005, 2005.

Kraemer, K. L. and Dedrick, J. (2005). "*Dell Computer: Organization of a Global Production Network - http://www.crito.uci.edu/GIT/publications/ pdf/dell.pdf*." Centre for Research on Information Technology and Organizations, 2005.

Oke, A. and Szwejczewski, M. (2005). "The relationship between UK manufactureres' inventory levels and supply, internal and market factors." *International Journal of Production Economics* **93-94**: 151-160.

Smith, S. A., Agrawal, N., et al. (2003). SAM: A decision support system for retail supply chain planning for private-label merchandise with multiple vendors. In: J. Geunes, P. M. Pardalos and H. E. Romeijn. *Supply Chain Management: Models, Applications and Research Directions*. New York, Springer**:** 163-199.

Waters, C. D. J. (1992). Inventory Control and Management. New York; Chichester, Wiley.

Zhai, E., Shi, Y., et al. (2007). "The growth and capability development of electronics manufacturing service (EMS) companies." *International Journal of Production Economics* **107**: 1-19.

AUTHORS BIOGRAPHY
Ivor Lanning

Ivor is a University of Limerick graduate with a first class honours degree in Industrial Engineering. His research is funded by the Irish Research Council for Science, Engineering and Technology (IRCSET) under the EMBARK initiative Ivor's area of study is in the computer-aided simulation of supply chains with particular emphasis on information sharing to reduce inefficiencies and increase customer service. Ivor's core competencies and research interests include Simulation Modelling of Discrete Event Systems, Modelling and Analysis of Supply Chains and Manufacturing, Decision Support Systems.

Cathal Heavey

Cathal Heavey is a Senior Lecturer in the Manufacturing and Operations Engineering department in the College of Engineering. He lectures in the areas of operations research, information technology, supply chain modelling and discrete event simulation. He is Joint Director of the Enterprise Research Centre. Cathal's Core Competencies and Research Interests include: Simulation Modelling of Discrete Event Systems, Modelling and Analysis of Supply Chains and Manufacturing Systems and Enterprise Modelling. In his research he has specialised in the area of modelling and optimisation of manufacturing and supply chain systems using both analytical and simulation techniques.

WORKLOAD FORECAST ALGORITHM OPTIMIZATION FOR RE-ORGANIZING RETAIL NETWORK

Agostino G. Bruzzone[a], Simonluca Poggi [b], Enrico Bocca[c], Francesco Longo[d], Francesca Madeo[e], Sabrina Rondinelli [f]

[a] [b] McLeod Institute of Simulation Science
DIPTEM University of Genoa
Via Opera Pia, 15, Genoa, 16145, ITALY

[c] MAST s.r.l
Piazza Lerda, 1, Genoa, 16158, ITALY

[d] [e] [f] Modeling & Simulation Center - Laboratory of Enterprise Solutions (MSC – LES)
M&S Net Center at Department of Mechanical Engineering
University of Calabria
Via Pietro Bucci, Rende, 87036, ITALY

[a]agostino@itim.unige.it, [c]enrico.bocca@mastsrl.eu, [d]f.longo@unical.it

ABSTRACT
This research is focused on developing an innovative approach for optimizing workload forecast algorithms in point of sale for retailers; this paper proposes a real case as validation framework and the procedures for optimizing and fine tuning the predictive algorithms for improving their performances. The analysis is based on different time series (i.e. sales, customers, working hours, etc.) correlated by the predictive algorithms. The paper proposes a metrics devoted to measure the performances considering the multivariable framework and the different target functions.

Keywords: forecast algorithm optimization, retail network re-organization

1. INTRODUCTION

Forecasting Models are used commonly to foresee future trends of real processes or phenomena; in effect there are different approaches: it is possible to concentrate on historical data (prediction based on the past), in modeling the reality (prediction based on the present) or even in reprocessing expert expectations (prediction based on the future); in general the most common used algorithms are based on mathematical algorithms re-elaborating historical data; this is mostly due to the fact that the creation of complex models requires much more efforts and customization, while the expert forecasts are usually not easy to be quantitatively elaborated adding further value; viceversa the time series analysis is easy to be adapted for a broad spectrum of applications.

Store organization represents a very important aspect to improve the overall performance of the Retail network; therefore to forecasts the workloads introduce the possibility to plan in advance the use of the resources and to guarantee that the service are always synchronized with the demand. However the requirements to support a planning system for a Retail store are much higher that that ones for an algorithm for predicting sales; in fact a store represent just a single node of a wide network and data statistical influence is not so high; in addition planning requires to consider multiple factors: i.e. activities for personnel are driven both by consumer demand (corresponding to cash barrier workload) and goods acceptance and preparations (corresponding to internal workload).

In addition it is important to mention that the data set are reduced even for the necessity to map each single store department (i.e. general goods, vegetables, cheese, meat, fish etc) and that the planning requirements is requiring to obtain a detailed behavioral evaluation of the workload along the days and the week. Based on this consideration the forecasts algorithms require to be very complex and to integrate different variables and factors to be fused together in order to achieve a better estimation of the system; however it is evident that an effective system in this area need to be able to self-tune and to be tailored for each specific department of each store, due to the specific characteristics of this behavior.

So the authors decided to use time series of different parameters and to combine them by data fusion algorithms, however in this case becomes critical to define the metrics for measuring the performance of the algorithms considering different aspects and to develop a methodology for optimizing the algorithm parameters in order to self optimize the system; this paper presents these aspect in relation to a case derived

from the real application of this approach to a major Italian Retail Network (data have been modified due to the confidential nature of the original sets).

2. TIME SERIES COMPONENT IN RETAIL

Within the classical approaches in historical series analysis, the main components of a phenomena are supposed to results from the aggregation of different components:

- Trend: it represents the course of the historical series in the medium-long period respect the resolution of the data available; it changes in the time, but it doesn't present predictable cycles previously. Actually, don't exist simple specific techniques to analyze the trend, nevertheless if the trend is monotonous increasing or decreasing the analysis result to be very simple. Often an visual observation of the series allows to diagnose the presence of the trend. Sometimes it is necessary to eliminate the trend for further data elaboration, in this case specific methodologies are applied.

- Period components: these corresponds to oscillations around the mean values due to the expansions and the contractions of phenomena. These periodic components have multiple nature and correspond in Retail Store to:

1. Year Seasonal Components: represent growth or decrease of the activity due to specific events (Christmas, Easter etc.) or to specific periods: i.e. store in tourist areas during holiday timeframe are more active, while concurrently that ones in town are characterized by the opposite behavior. These sub-components depends on the location, climatic factor and social-cultural frameworks of the area, so each store have its own characteristics to be defined.
2. Month Periodic Components: during the month the demand have usually a periodic demand behavior due to the mostly concurrent distribution of salary to consumers (near the end of the month), so in Retail sales and related activities are affected by a fluctuation due to this aspect; usually this component have a limited influence on the store workload respect other ones.
3. Week Periodic Components: during the week the demand is very variable (i.e. friday and saturday) so this introduce a very strong periodic components.
4. Day Periodic Components: along the day the demand and activities are changing with peak timeframes.

- Error: it's the irregular component that introduce deviations. All historical series, that are not deterministic, have positive or negative irregularities, produced by stochastic factors. Usually it is acceptable the hypothesis (based on the fact that there are not systematic errors: i.e. data collections) that the Error is a random variable with zero mean, constant variance and not correlated.

In fact in retail historical series treated in this research refer to workload necessary to the different sales points of a major Italian retailer.

As anticipated the products of the great distribution are subject to yearly, monthly and weekly seasonal behavior and so related activities in retail store are affected by these components:

- the yearly seasonal behavior refers to the increase of demand for some kinds of product during particular periods of the year (Christmas, Easter, Summer…) and this corresponds often to an increase of workforce in the departments of these products;
- the monthly periodic component refers to economic factors, in particular to the consume of salaries. In fact, in the first 15 days of the month it's possible to observe a greater demand than the last days;
- the weekly seasonability reflects the habits of family: the first day of week is characterized by ordinary purchases to cover ordinary needs of week, while the purchases in the last days of week are extra purchases related to Sunday. Weekly workforce is characterized by peaks of work in the first day of the week both to respond to the demand both to allocate new promotional material and goods.

In addition during the store life-cycle many events occur that affect the behavior (i.e. competitor opening in the surroundings, strikes, power failure, etc.); the forecasts need to estimate the impact of these past punctual events for correcting the historical time series and to predict the presence in future of someone (i.e. day with extra opening time, special promotional week) for future events.

3. FORECAST MODEL

The historical data for forecasting the retail store workload have been defined based on a structure that characterize each point of sale and each department.

The basic element is the department, each department belongs to a store and each store belongs to a retail cluster.

Each department is characterized the following data:
- TimeFramed Data (TFD): data related to the evolution of the independent variable along the day collected each quarter of an hour (customers, sales, working hours);
- Date and Flow Data (DFD): data related to the evolution of the independent variable along the

weeks/months/years summarized day by day (customers, sales, working hours, arriving pallets);

- Calendar: include the details of the operating hours and the definition of events (past or futures) and their estimated/expected impact on the different independent variables.
- Department Characterization: all the calibration parameters of the algorithms for the specific sector. Since previsions are subject to different stochastic factors, it is necessary to correct them with calibration parameters that reduce the deviation between real data and prevision data.

The forecast model, developed by authors, combines the different data of the 96 quarters of an hour of the day (timeframed) and of the days (data&flow); the results of these two data set some time don't corresponds due to administrative changes and corrections, in addition their comparison allows to complete consistency tests. The data set are expected to be provided by company ERP (Enterprise Resource Planning) and the personnel presence system in multiple session to allow the update of data to be corrected by managers and administrative procedures; in fact in order to improve accuracy on the forecast input data are corrected by specific parameters (i.e. event impact on past data). The model processes input files based on the data structure above described in term of TimeFramed Data, Date and Flow, Calendar and Departments.

Forecasts are generated for each store and structured based on two main data set: FOT (Forecast Organized by Timeline) and FOD (Forecast Organized by Date) corresponding respectively to the evolution along the day and over weeks; they include the target functions related to customers, sales and pallets and working time; in fact the final goal for planning and store organization is to forecast working hours so the other target functions have to be considered as supporting information.

In the case study the authors developed a tailored model for the users integrated with the company ERP able to generate and distribute the forecasts for next weeks; historical and prevision data are available for two years along the day and for five months in term of daily evolution.

In the case study proposed the network is composed by twelve stores each one including five departments:

- department 1: general goods;
- department 2: deli and dairy products;
- department 3: fruits and vegetables;
- department 4: butchery;
- department 5: fishmonger.

FOD and FOT are save in files for each stores of each point of sail of the retail network and distributed to the stores; a specific tool for post-processing the

forecasts in order to include store manager preferences and final tuning was developed by the authors and installed in the stores.

4. THE CALIBRATION PROCEDURE

To identify optimal parameters for algorithms tuning the authors defined a specific metrics based on error indicators able to measure deviation between forecast and real data. In fact forecast algorithms are strongly depending on calibration parameters related to the department that need to be optimized based on specific behavior profile; this need to be identify based on historical data analysis. To identify the optimal parameter sets, in term of deviation reduction between real data and forecasts, the authors developed a calibration procedure based on four different error indicators; these are devoted to measure the forecast errors and to evaluate the algorithm parameter set efficiencies. The metrics developed consider both the conceptual forecast error and the use of forecast itself; based on these consideration the authors identified the following indicators:

- J_1^j is an index, for j-th target function, derived by MAD (Mean Absolute Deviation), for the calculation of the absolute error of daily workload.

$$J^j_1 = \sum_{k=1}^{7 \cdot m} \frac{\left| DFDr^j_k - DFDf^j_k \right|}{7 \cdot m} \qquad (1)$$

- m: number of weeks under evaluation;
- $DFDr^j_k$: real Value of j-th DFD target function on the day k-th;
- $DFDf^j_k$: forecast Value of j-th DFD target function on the day k-th;
- J_2^j is an index, for j-th target function, that absolute measures the error of daily workload.

$$J^j_2 = \sum_{k=1}^{7 \cdot m} \frac{DFDr^j_k - DFDf^j_k}{7 \cdot m} \qquad (2)$$

- J_3^j is an index, for j-th target function, derived by MAD (Mean Absolute Deviation), for the calculation of the absolute error of the daily workload in each quarter of an hour.

$$J^j_3 = \sum_{i=1}^{672 \cdot m} \frac{\left| TFDr^j_i - TFDf^j_i \right|}{672 \cdot m} \qquad (3)$$

- $TFDr^i_k$: real Value of i-th TFD target function on the day k-th;
- $TFDf^i_k$: forecast Value of i-th TFD target function on the day k-th;

- J_4^j is a comprehensive weighted target function that represent an index able to represent the sum of the other indicators.

$$J_4 = k_1 J_1 + k_2 |J_2| + k_3 J_3 \qquad (4)$$

- k_1, k_2, k_3, k_4: factors for balancing the weighted average of the performance indexes.

The calibration parameters are tuned in order to optimize the J_1, J_2, J_3 and J_4. In particular the authors decided to fix a set of correcting factors (k_1 k_2 k_3 k_4) and to proceed in the optimization of the parameters mostly by minimizing J_4.

The calibration procedure, to be repeated for each single department of all the stores, is following:

- Preprocessing analysis on input data set:
1. Calendar, TFD, DFD Format Check;
2. Calendar Consistency Check;
3. Comparison DFD vs. Store Template Profile;
4. Comparison TFD vs. Store Template Profile;
5. Comparison DFD vs. TFD and difference motivation.

- General Optimization:
1. Automated Optimization for each error index over a wide range of analysis;
2. Comparison of the J_1, J_2, J_3 and J_4 error indexes and tuning of k_1, k_2, k_3 and k_4 factors.

- Fine Tuning:
1. Automated Optimization in small range of analysis for J_4;

- Update of Algorithm Parameters;
- Test of the performance achieved on data not used for forecasts (i.e. one or two more weeks).

Based on above described procedure, it is possible to define the best parameters for each sales point of a distribution network.

In fact it is pretty important to divide the historical data in order to keep available for testing last weeks; in order to achieve this results it is suggested, during the test phase, to complete the analysis on the time series with the exception of the last two weeks; these data will be used during the test phase for validating the robustness of the configuration.

As example of calibration it is proposed the case of general good department of one store of the network in the case study proposed.

In the case the general optimization results of the analysis are presented in relation to the change in one of the department parameters.

Table 1: Different Errors Comparison

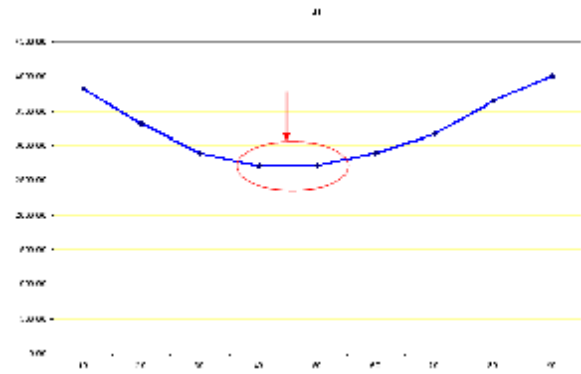| Param.	J1	J2	J3	J4
10	25.05%	24.53%	37.81%	3819.95
20	20.67%	17.74%	33.06%	3316.10
30	17.45%	10.98%	28.92%	2888.20
40	18.38%	4.04%	26.57%	2699.95
50	21.39%	-2.68%	25.99%	2708.40
60	26.73%	-9.61%	26.87%	2892.40
70	31.04%	-15.43%	29.01%	3165.90
80	37.47%	-23.43%	33.03%	3646.55
90	41.99%	-28.39%	36.06%	4002.05
100	50.39%	-37.30%	41.89%	4681.80

Figure 1: Identification of Fine tuning Area for J_4

From graphs and tables it's possible to observe that the J_4 best values are concentrated in range 35 - 55 of the parameter under optimization; completing a detailed analysis in the surroundings the results are the following:
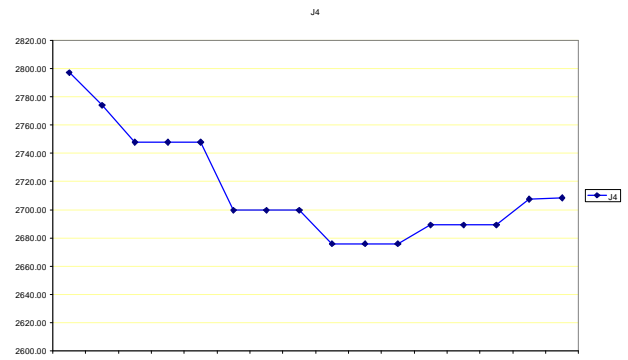

Figure 2: Fine Tuning of J4 in the proposed case

Table 2: Error Indexes during Fine Tuning

Param.	J1	J2	J3	J4
35	17.35%	8.15%	27.91%	2797.25
36	17.54%	6.87%	27.60%	2774.00
37	17.74%	6.02%	27.26%	2747.90
38	17.74%	6.02%	27.26%	2747.90
39	17.74%	6.02%	27.26%	2747.90
40	18.38%	4.04%	26.57%	2699.65
41	18.38%	4.04%	26.57%	2699.65
42	18.38%	4.04%	26.57%	2699.65
43	18.84%	1.94%	26.19%	2675.90
44	18.84%	1.94%	26.19%	2675.90
45	18.84%	1.94%	26.19%	2675.90
46	19.66%	0.21%	26.16%	2689.15
47	19.66%	0.21%	26.16%	2689.15
48	19.66%	0.21%	26.16%	2689.15
49	21.09%	-2.11%	26.04%	2689.15
50	21.39%	-2.68%	25.99%	2707.45
51	21.39%	-2.68%	25.99%	2708.40
52	21.39%	-2.68%	25.99%	2708.40
53	21.39%	-2.68%	25.99%	2708.40
54	23.85%	-5.94%	26.27%	2708.40
55	24.40%	-6.72%	26.36%	2782.15

In the same way all calibration parameters are defined for each target function of each departments of the network.

5. COMPARING PERFORMANCES

The analysis for optimizing the forecast algorithms is pretty articulated; it was identified the necessity to forecast the workload for each quarter of an over over next weeks for each department of each store of the network.

In order to proceed in this sense and provide possibility to have daily up-to-date analysis it was necessary to combine data related both to detailed evolution (immediately available and pretty time consuming for processing) and to day summary (including corrections and easy to process but available later).

In addition the forecast performance changes over the time due to the fact that the system needs periodic tuning; considering the stochastic component it is pretty difficult to maintain a clear understanding of the problems that is articulated over different directions that are synthesized in the following:

- Store/Department Explosion of variables;
- Periodic check for monitoring forecasts performances along the time;
- Day vs. Week prediction for allocating resources.

In addition different target functions have been defined and even if for the case proposed the working hours are the main factor it is evident that the check about goods arrivals (pallets), sales and customers are pretty interesting for planning activities and need to be keep under control for providing a reliable support to store managers.

Due to these reason the authors developed a solution for complete analysis over the different direction of investigation on the forecasting results; this solution was defined as Drawer and it provides the possibility to automate the optimization process by defining the scenario to be investigate in term of timeframe, target function, store, department etc. The Drawer allows to fix the range of analysis and the investigation resolution as well as the optimization algorithms and all its parameters. The system perform in sequence all the different set of steps required by the proposed methodology and provides results in term of tables and graphs for each one in order to allow the analyst to complete a review. In addition the drawer export the files for automated configuration of the forecast algorithm in order to guarantee an easy and periodic update of the forecast algorithm optimization.
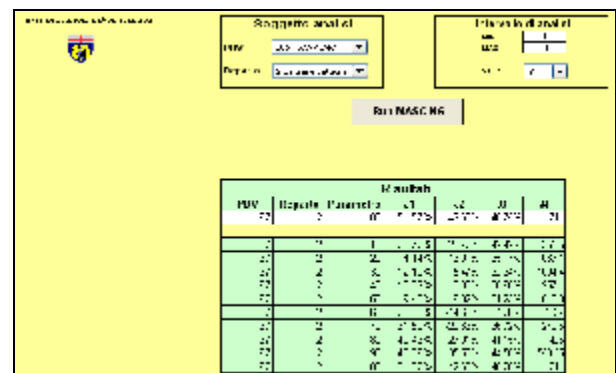

Figure 3: Interface of Drawer

The drawer process directly the source database DFD and TFD available for the forecasting algorithms and analyze the results generated in FOD and FOT databases. The comparison are based on: out_t, out_d analysis; out_t contains comparison of real and forecasts for TFD data, while the out_d focuses on DFD data; it is interesting to note that the analysis of the performance on DFD, corresponding to check the total working hours required for each day of next week need compared the FOT integrated over the day in order to check consistency of detailed planning hour by hour.

Moreover the data in out_t and out_d are used to check the performance indexes evolution.

For instance, in the case in analysis, are proposed two different results as following:

- TFD vs FOT. It's a line diagram, that represents the workload profile required in each quarter of an hour along the days of future weeks based on forecasting algorithms respect real data; this is presented in the following.
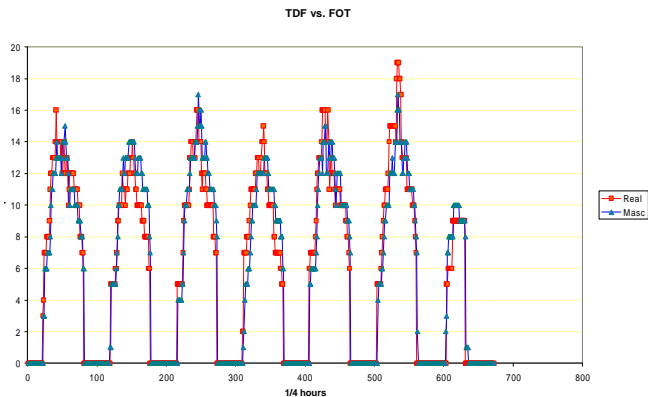
Figure 4: TFD vs FOT

- DFD vs integrated FOT. This bar diagram, that proposes the comparison between daily workload (h/day) during the period of analysis respect the forecasts provided by the algorithm in term of FOT.
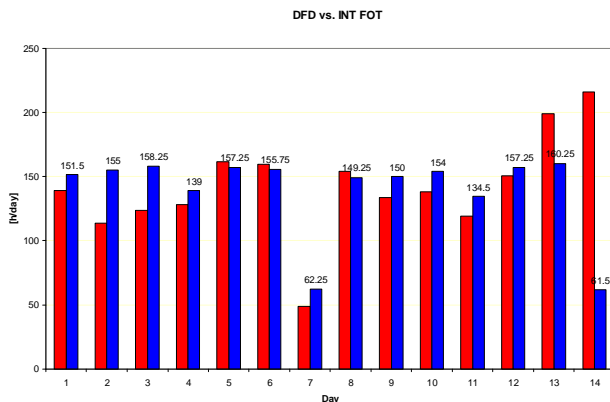

Figure 5: DFD vs integrated FOT

6. CONCLUSIONS

The methodology proposed was applied over a small cluster of stores on a wide retail network as pilot; the results was very promising in term of efficiency of the algorithms and reliability of the methodology.

In fact the proposed example allowed to validate a quick and robust methodology for forecast algorithm tuning that create a periodic updating calibration procedures.

Therefore the forecast model after the parameter optimization improved its performances of about 60%, reducing the deviation between real data and forecasts for all sales points and obtaining results that are very satisfactory based on store expert point of view. Moreover the use of Drawer allows to automate the calibration procedure reducing the efforts required in term of skills and time and guaranteeing the introduction on the company of a systematic procedure

REFERENCES
Bruzzone A.G., Mosca R., Spirito F., Coppa A., Simeoni S. (2001) "Modelling & Simulation for Customer Satisfaction in Retail Warehouse Management", *Proc. of EUROSIM2001*, Delft NL June 26-29

Bruzzone A.G., Mosca R., B.M., B.C., Spirito F., Coppa A.C.M. (2001) "Advanced Modeling for Customer Behaviour Analysis ", *Proc.of MIC2001*, Innsbruck Austria, February

Bruzzone A.G., Brandolini M., Simeoni S., Massei M. (2002) "Virtual Shop through Internet: E- Business for Supermarket Chains", *Proceedings of HMS2002*, Bergeggi Oct 3-5

Bruzzone A.G., Viazzo S., Longo F., Papoff E., B.C., (2004) "Simulation and Virtual Reality to Modelling Retail and Store Facilities", *Proceedings of SCSC2004*, San Jose', CA, July

Bruzzone A.G, Viazzo S., Massei M. (2005) "Computational Model For Retail Logistics", *Proceedings of WMSCI*, July 10-13

Bruzzone A.G., Bocca E., Massei M., Pierfederici L., Poggi S. (2007) "Advanced Models for Performance Control in Supermarket Logistics Management", *Proc. of MAS2007*, Bergeggi, October

Gibson P., Greenhalgh and R. Ken, *"Manufactuting Management: principles and concept"*, Chapman & Hall

Montgomery D.C., L: A. Johnson, J. S. (1990) *"Gradiner, Forecasting and Time Series Analysis"*, Mc Graw Hill,1990

Papoulis A., *"Probability, random variables and stochastic processes"*, MCGraw-Hill, 1984

Roger W. Shmenner, *"Produzione: scelte strategiche e gestione operativa"*, Il Sole 24 ore Libri

A SIMULATION-BASED METHOD FOR THE DESIGN OF SUPPLY STRATEGIES TO ENTER DEVELOPING MARKETS

Christian Schwede[a], Yu Song[b], Brian Sieben[c], Bernd Hellingrath[d], Axel Wagenitz[e]

[a],[b],[c],[d] [e] Fraunhofer Institute for Material Flow and Logistics in Dortmund, Germany

[a]Christian.Schwede@iml.fraunhofer.de, [b]Yu.Song@iml.fraunhofer.de, [c]Brian.Sieben@iml.fraunhofer.de, [d]Bernd.Hellinrath@iml.fraunhofer.de, [e]Axel.Wagenitz@iml.fraunhofer.de

ABSTRACT

In this paper we provide a method to support decision makers selecting strategies for global supply chains during the life cycle of a product. As we focus on developing markets such as China or Latin America, we stick to supply strategies which we call "supply strategies to enter developing markets". The method we present is divided into two steps. Starting with a definition and classification of these strategies a typology based selection method is used, clustering the strategies with respect to their elemental properties. Thus having pre-selected the strategies, due to the dynamic behaviour we use discrete-event simulation to have a closer look on the remaining options. Following this procedure, we are even able to achieve a more detailed setting of parameters such as stocks, capacities or transport frequencies. Within this paper we will focus on the second step of the method.

Keywords: supply chain design, discrete-event simulation, global sourcing, product life cycle

1. INTRODUCTION

The rising importance of emerging markets for western companies becomes especially apparent in the field of international Supply Chain Design (SCD). Hence, in recent years supply concepts such as Completely Knocked Down (CKD) and Part by Part (PbP) have been widely applied in order to supply newly established production sites in markets with steeply improving purchasing power.

However, shifting production steps to low-wage countries is not only motivated by labour cost advantages or the avoidance of import tariffs. Increasingly, strategic aspects such as hedging currency risks and building up a global sourcing base are taken into account which add new levels of complexity to international SCD policies. Since logistical requirements typically shift over a product's life cycle, the need for a continuous re-assessment and redesign of the supply processes is evident.

In this paper, we introduce a method that provides decision makers in industry with feasible guidelines for the design of supply strategies to enter developing markets over a product's life cycle.

The basis for the method is a comprehensive typology of such strategies. By setting the values of the parameters in the typology a limited number of feasible logistical strategies is identified. Then we apply discrete-event simulation, more specific the OTD-NET framework (Order-to-Delivery) created by the Fraunhofer IML in Dortmund (Wagenitz 2007; Toth, Hellingrath, Wagenitz and Klingebiel 2006) to decide between the remaining strategies. Furthermore by demonstrating through simulation that changing the supply strategy when sales are growing can save money, we will support the statement that an iterative re-design of the supply chain does make sense. Within this paper we focus on the second step of the method, while the first one will be discussed in detail in another publication (Sieben et al. 2008).

By following this approach, a transparent method for decision support with regard to supply strategies to enter developing markets can be derived.

The paper is structured as follows: In chapter 2 the so called supply strategies to enter developing markets will be defined and brought into a chronological order according to the product life cycle and market maturity. In chapter 3 the strategies will be classified using three sets of identifying attributes and the principle of the typology for the strategy selection is outlined. In chapter 4 the second step of the method applying simulation will be described and the discrete-event simulation environment OTD-NET will be presented. In chapter 5 a scenario out of the automotive industry to which the suggested method will be applied to is described and the results of the case study are presented. Chapter 6 will resume the research results and provide an outlook of further research and applications.

2. SUPPLY STRATEGIES TO ENTER DEVELOPING MARKETS

Since internationalization is a key factor of success for global enterprises, the establishment of products in emerging markets becomes more and more important. (Laux 1992; Moran 2000; Humphrey and Memedovic 2003; Petri 2004). Dealing with global supply strategies for these developing markets one can identify a certain set of strategies forming a chronological order due to the heights of sales and the market maturity. The main

question is how to enter such a new market and especially what supply strategies are adequate for which type of product.

Assuming a protected market with a poor infrastructure of suppliers and instable sales, the first step to enter this market would normally be to sell completely manufactured products in the foreign country, avoiding risky investments. But when the sales grow higher and more stable decisions concerning a more or less intensive local production in the market's country must be made. Typically production arrangements for semi-knocked down or completely knocked-down kits' (SKD or CKD kits) in the automotive or computer / electronics industries are used as means of entering new (or protected) markets (Moran 2000; Tulder and Ruigrok 1997; Elkin 2006). For instance the colour TV assembly in India since the mid 1980s (Kinra 1995), and the Information and Communication Technology (ICT) products manufactured in China in the 1990s (Hong 2008; Tan 2002). Other more popular examples come from the automotive industry: Japan, since the beginning of World War II (Smitka 1990), South Korea, in the early 1960s (Mukherjee and Sastry 1996) or more recently the motorcycle production in Vietnam in 1998 (Fukase and Martin 1999) and the production of Audi A 6 in Audi's joint venture with First Automobile Works (FAW) in Changchun, China in 2003 (Wagnera and Johnson 2004).

The dynamic learning and the accumulation of production experience in an "infant industry" environment can lead to fully competitive local industries over time. From the export of completely finished products to the completely local production, there are several strategies like CKD that equal in some properties and differ in others (Chung 1998). The strategies and terms for the strategies we apply and present in the following section are commonly used in science and industry (Volpato 2002). To distinguish these strategies from general supply strategies we introduce a special term in the following section.

2.1. Considered strategies

We call the following five supply strategies "supply strategies to enter developing markets" if they are applied to a supplier-customer relationship between a mother company in an industrialized country and a dealer within a developing market of a foreign country.

2.1.1. Completely build up (CBU)

CBU is the abbreviation for Completely Build-up Units, or also known as Fully Build-up Units (FBU) or BUX (Built-Up Export). It refers to the import or export of finished products.

The CBU process is the export of the finished products. The products are fully produced and assembled in the country of origin and exported as a whole piece to the destination market. Dealers or end consumers pick up the products directly from the importer.
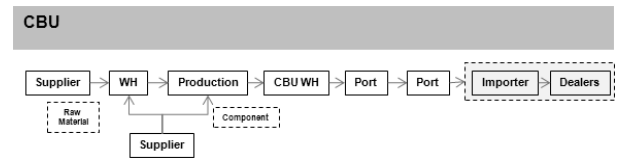

Figure 1: CBU Process Description

2.1.2. Semi Knocked Down (SKD)

Semi Knocked Down (SKD) is based on SKD lots or kits also known as SKD sets. A SKD kit needs the final assembly process to become a complete product (Jun 1987).

That is why in some literature SKD is also called turn screw assembly, by which only a certain module (for example wheels, steering wheels, and bumper) needs to be completely screwed or installed (Kaltofem 1999).

The SKD process can be divided into different variants we name SKD-Basic, SKD-Medium, and SKD-Advanced. All these SKD variants follow the same idea that the exporter collects and packs the parts in the SKD warehouse, while the importer repacks the products in the local OEM factory (SKD-packing and SKD-unpacking). The strategies differ on the exporter side, where the suppliers may provide the export factory with different assembly levels of components.
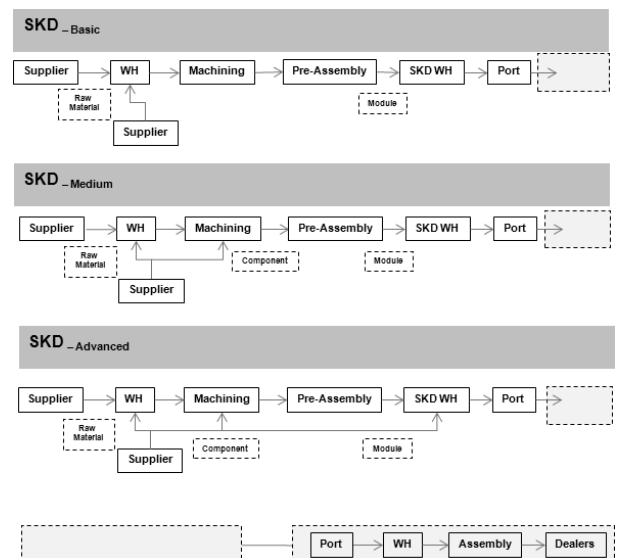

Figure 2: SKD Process Description

2.1.3. Completely Knocked Down (CKD)

Completely Knocked Down (CKD) is based on CKD lots or kits also known as CKD sets. CKD means the product is exported in parts and will be assembled at its destination (Schmitt, Thiesse and Fleisch 2007). CKD sets are managed in the same way as SKD. A CKD kit differs from a SKD kit in that it has less preassembled parts (Jun 1987), which means in the case of CKD, local assembly activities are much more substantial than with SKD (Tulder and Ruigrok 1997).

Both CKD and SKD refer to the practices of selling knocked down kits to foreign affiliates in order

to finish the assembly process due to the consideration of the quality or sophistication of the local suppliers (Grant and Gregory 1997). The benefits of this practice for the exporter include avoidance of high import taxes and/or receive tax preferences for providing local employment (Hong 2008).

The CKD process can be also divided into different variants we name CKD-Basic, CKD-Medium, and CKD-Advanced. The difference between the three variants mainly lies on the importers side, where the factory in the import country takes over more and more responsibilities, coordinating local suppliers and accordingly the planning and assembly in the local factory is more and more sophisticated.

In general, under the CKD mode, from the strategy point of view, the exporter has the control of technology know-how. No matter in which degree and form of CKD, the factory organization structure and information flow are still dominated by the exporter side. From the operation point of view in the exporter warehouse all CKD parts are subject to the CKD-packing in the CKD warehouse before export. In the importer plant, before the assembly, the CKD kits are subject to the CKD-unpacking according to the demands from the actual working stations.
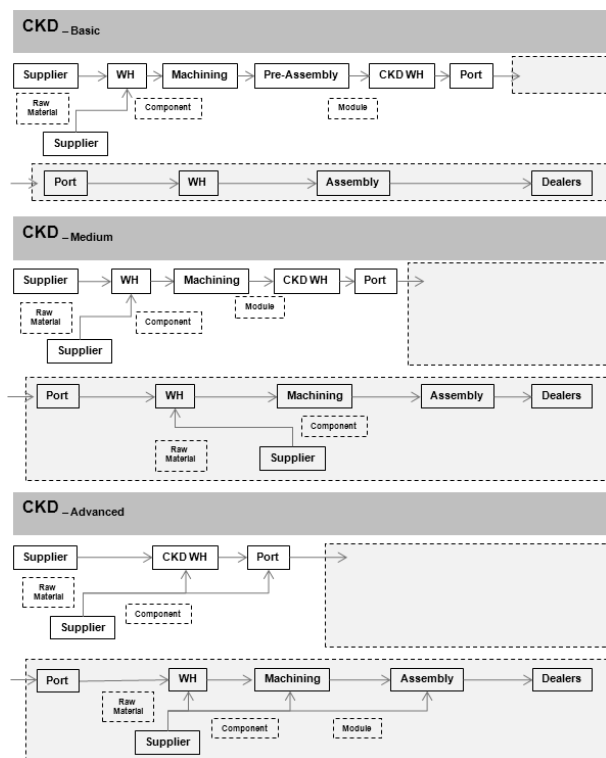

Figure 3: CKD Process Description

2.1.4. Part-by-Part (PbP)

PbP is the abbreviation of Part-by-Part. A PbP component shipment is normally delivered to a more sophisticated assembler who even can have full control of its bill of material and a high local sourcing. Under the mode of PbP, the shipment lines (part numbers and quantities) are not constant and generally not in engineered pack solutions.

Under the PbP mode, the importer takes over the control of technology know-how. The factory organization structure and information flow are also dominated by the importer side. The importer factory makes out the planning for the orders and transfers the information to the exporter accordingly, while the exporter acts as the coordinator for parts collection in between.

Besides, during the warehousing operation in the exporter PbP warehouse there is no repeated unpacking and packing for the PbP parts, which means in the warehouse for PbP only consolidation for the sea shipping has to be performed. When using PbP the decision has to be taken which parts can be delivered directly by the supplier and which are to be consolidated with others.
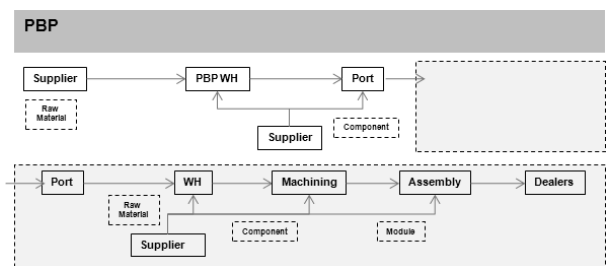
Figure 4: PbP Process Description

2.1.5. Mainly Locally Sourced

100% localized purchasing and production is quite rare, hence here we define a production as mainly locally sourced when the local contents account for more than 90%. Under this mode, the importer factory functions as the mother factory, fully independent in planning and production for the local market, only a limited number of parts will be ordered and transported from certain overseas suppliers.

Figure 5: Mainly Local Sourced Process Description

2.2. Chronological order of the supply strategies

The five strategies described above can be related to the growth of sales of the product (product life cycle) and the market's maturity. This leads to a chronological order of the strategies (Figure 6).

Figure 6: Chronological order of the Strategies
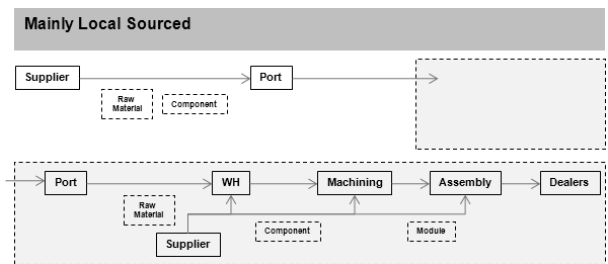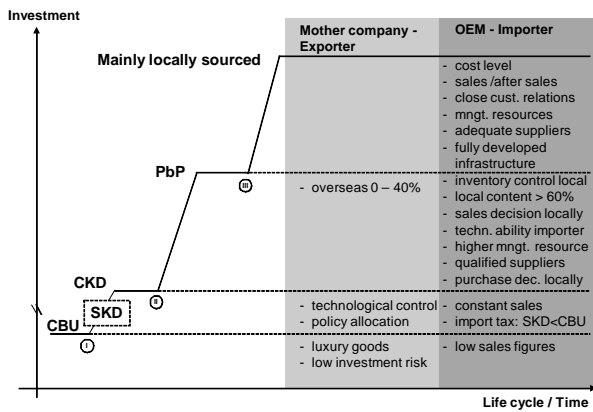
Following this order the mother company in the industrialized country might begin with a CBU strategy assuming that the sales are low and the infrastructure in the target country is poorly developed. Thus avoiding the risk of high investments in building-up own manufacturing infrastructure, supplying the dealer with completely build-up products is more feasible even if the import taxes might be higher.

When the sales increase and the maturity of the market grows it will become more feasible risking an investment and avoiding import taxes. Yet local suppliers are rare, leading to the strategy of SKD or CKD.

With even higher rising sales and established local suppliers avoiding packaging and repackaging of the CKD process becomes possible. The import OEM might even have developed the required know-how to run the planning mainly by itself, avoiding unnecessary high stocks or stock shortages. This leads to the PbP strategy.

The last stage would be reached if the pool of local suppliers is developed so far that a mainly locally sourced production strategy can be executed.

Taking this sequence into account, the main problem that remains is: "When exactly to shift from one strategy to another?"

3. STEP 1: STRATEGY SELECTION VIA TYPOLOGY

In this chapter we will provide a classification of characterising attributes and describe how a typology of the supply strategies to enter developing markets can be used to help a decision maker. However, the description of the actual typology would stretch the point of this paper and thus will be presented in another publication (Sieben et al. 2008). The usage of this typology is the first step of the strategy selection.

3.1. Using a typology

Since the process of selecting an appropriate supply strategy can be characterised as rather complex, from a system oriented point of view typification seems to be an especially suitable way to encapsulate those multiple aspects in a structured and consistent way and hence to provide a decision maker with an omnibus, decision

relevant model (Kluge 2000; Elman 2004; Schieferdecker 2003).

Criteria	Characteristics				
Hard lines					
Local Market Trend for this product	Development	Introduction	Fast Growth	Stable Growth	Maturity
Local Policy on FDI in this industry	Strong local industry protection policy	JV with major share by local partner	JV with major foreign share allowed		100% foreign share allowed
Local Policy on local content ratio	No local content	0-19%	20-40%	41-60%	>60%
Import Tax in importer OEM country	0	1 -15%	16 - 30%	31 - 50%	> 51%
Labour costs in OEM country	Lower than export country	No big difference		Higher than export country	
Local transportation costs in OEM country	Lower than export country	No big difference as the export country		Higher than export country	
Sea transportation price export from OEM country	Cheaper	More expensive		No big difference as import	
Sea transportation price import to OEM country	Cheaper	More expensive		No big difference as import	
Strategy Level					
Product Life Cycle in OEM country	Development	Introduction	Fast Growth	Stable Growth	Maturity
Real Market Demands for this product	Development	Introduction	Fast Growth	Stable Growth	Maturity
Factory Organization	Controlled by Mother Company		Controlled by exporter OEM	Controlled by importer OEM	
Technology controlled by	Mother Company		Exporter OEM	Importer OEM	
Target Market for the finished products	Only domestic consumption		Domestic and export		
Operation Level					
Order from importer	Daily basis		Weekly basis	Monthly basis	
Packaging from exporter	product oriented	module oriented	component oriented	raw materials	
Warehousing before export	Consolidation		No Consolidation		
Shipping and Quality when damage occurs	Special order		no Special order		
Complexity (assembly / production) in OEM country	Not complex at all	Low		Medium	High
Local content ratio in OEM production	0	0 - 40%	40 - 60%	60 - 90%	> 90%
Integrated Logistics Chain Control	Central Control by Mother Company		Decentralised controlled	Centralised control by OEM	
Inventory Transparency in OEM	No transparency when controlled by the mother		No transparency when controlled	Transparent when controlled by OEM itself	

Figure 7: Morphology of supply strategies to enter developing markets

From a typology, a type can be derived which represents a (possible) observation with common distinguishing features and attributes (properties). The five strategies described in chapter 2 can be considered as basic types within a typology. Every typology is based on a morphology (see: Figure 7) that consists of a list of characterising attributes and their potential settings. The different combinations of settings then lead to the possible types (here: strategies).

A decision maker can use the typology to get the appropriate strategy for the current supply situation by choosing settings for the different attributes. Since there are attributes unchangeable by the decision maker and attributes that can be influenced such as those which refer to strategic and those which refer to operational decisions, we have come up with a classification of these attributes, which will be presented in section 3.2. Hence there is also the possibility that the decision maker doesn't know or care about a setting for a certain attribute, this could lead to optimisation potential, when several of these "gaps" make more than one strategy possible. When these optimisation potentials exist they will be taken into the second step of our method and

included into the selection of the optimal strategy via simulation.

3.2. Classification of the strategies' attributes
Here the classification will be described, identifying and classifying attributes of the strategies. In general we order the attributes into three main categories, they are:

- Hard lines
- Operation Level
- Strategy Level

In the following we will develop these three categories in detail.

3.2.1. Hard lines
The hard lines refer to economic conditions and industrial policy in the exporting country and / or importing country, which are hard facts and cannot be changed by business entities, including:

- Local Market Trend for this product
- Local Policy on FDI in this industry
- Local Policy on local content ratio
- Import Tax in importer OEM country
- Labour costs in OEM country
- Local transportation costs in OEM country
- Sea transportation price export from OEM country
- Sea transportation price import to OEM country

3.2.2. Operation Level
The operation level indicates the business environment for the supply process, factory or products in the importing or exporting country, which can be either defined by the local market situation or decided locally in the factory, for example:

- Product life cycle in OEM country
- Real Market Demands for this product
- Factory Organization
- Technology control party
- Target Market for the finished products

3.2.3. Strategy Level
The strategy level mainly contains decisions from the overseas mother company, for example:

- Order from importer
- Packaging from exporter
- Warehousing before export
- Shipping and Quality when damage occurs
- Complexity (assembly / production) in OEM country
- Local content ratio in OEM production
- Integrated Logistics Chain Control
- Inventory Transparency in OEM

4. STEP 2: SIMULATION
After the typology has been applied to a certain supply situation, a subset of the strategies has been selected due to political and strategic parameters, yet only roughly focussing on the process parameters themselves. This focus will now be set using the process simulation. The parameter setting from the typology will be extended by more precise settings and additional parameters needed for the simulation model. Each member of the subset of strategies is then integrated in the basic model, forming distinct scenarios that are simulated, analysed, compared and evaluated with respect to the process parameters. The best strategy can then be selected and applied. In the following sections the simulations environment used to perform the second step of the method and the components that can be used to model the supply scenarios are presented.

4.1. The OTD-NET simulation environment
OTD-NET is a simulation and modelling environment using discrete-event simulation to analyze the order-to-delivery process of huge supply networks. It was developed by the OTD-group in the Fraunhofer Institute for Material Flow and Logistics in Dortmund, Germany.
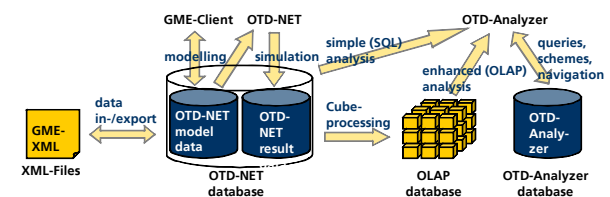

Figure 8: The OTD-NET environment

Figure 8 shows the three modules of the environment: the Graphical Modelling Environment (GME), the Simulator and the Analyzer. Using these three components OTD-NET can be used to model and simulate supply networks based on customer orders or forecasts that are back-propagated through the supply chain. Considering realistic network planning, single orders are satisfied with single products which can be observed throughout the whole process. Economical ratios as capacities, workload, stocks, delivery reliability, process times and costs or capital commitment can be derived and evaluated.

Regarding supply strategies to enter developing markets, OTD-NET can be used to model, analyze and compare different strategies, focusing on ratios such as stocks, transport/ packing costs and capital commitment.

4.2. The components of the process model
To create a process model of a global supply chain, the OTD-NET environment offers several components and options that can be divided into three mayor categories. These components can be combined and set differently according to the focus of the scenario to be modelled.

4.2.1. The network components

The main component of the network is the plant of the original equipment manufacturer (OEM) in either the industrialized or the emerging market's country. This component is defined by the location, capacities, working shifts, lead time, products and the behaviour of the product sequencing. The production sites of the suppliers of the OEM are connected via distribution channels, delivering the parts the main products consist of. These production components have mainly the same attributes as the OEM plant, but can additionally be run as Build-to-Order or Build-to-Stock suppliers. The distribution channels that connect the suppliers with each others or the OEM can use all kinds of transportation devices (trucks, trains, ships, aeroplanes, etc.). They have a capacity, transportation behaviour (either by plan or on demand) and are connected by conditioned routings. In between the distribution channels can be warehouses with capacities.

4.2.2. The product structure

The product structure is a hierarchy with the end products to be sold to the customers at the top. These products have different property groups leading to different variants of a product. Properties consist of parts that can consist of parts themselves, forming a hierarchical bill of material. Parts and properties are produced by the suppliers while the products are manufactured by the plants. Complex dependencies between the properties can be included leading to constraints and restraints between the properties of a product.

4.2.3. Sales planning and customer behaviour

The production in OTD-NET is triggered by customer orders and forecasts. The plant generally produces based on periodic forecasts, propagating the demands back through the network. Producing on forecast can lead to a gap between the products built and the products wanted by the customers. This gap can be used to model the deviation between the forecasted and the actual demand. The customers in the model can be patient, waiting several time periods for their products, or buying products similar to the one originally wanted if there is a certain discount. Furthermore they could not be willing to accept products that are more than a certain time period in the warehouse. Thus, a complex behaviour of customers and planning cycles can be realised.

5. THE SCENARIO
5.1. Description

The scenario we apply the method to is a global supply scenario within the automotive industry. It involves typical processes of global supply chains and focuses on the entry of an emerging market. Furthermore a single strategy cannot be clearly derived. Thus the scenario is appropriate for demonstrating our concept.

In detail a German car manufacturer is selling a certain car in China. Transportation between the continents is done via shipments taking 6 weeks, debarking weekly. Focusing on the second step of the method, we assume that the first step, applying the typology, has yield in a subset of two strategies. Since an assembly plant in China already exists these strategies are CKD – including CKD-packing of all parts of one car into one 40ft container in Germany and CKD-unpacking of this container in China to feed the different lines – and PbP, which in this case means the German suppliers send their parts directly to China. The decision to be made is if it is feasible to consolidate the parts in a CKD-set or if it is better to send them unconsolidated, saving process costs. In Figure 9 you can see the two different supply strategies.
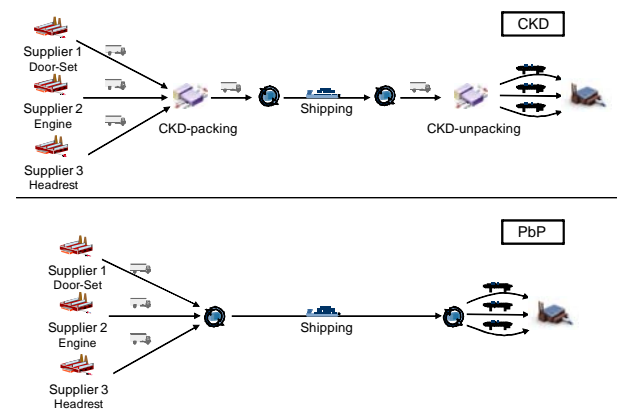

Figure 9: Network structure of the scenario

The scenario contains three suppliers, supplying three parts for the car that have been selected for their size and values. The door-set, supplied by Supplier 1, is the biggest part and has a medium value. Supplier 2 provides the engine with a medium size and a high value, and Supplier 3 the headrests with a small size and small prize. Thus, we can analyse different behaviour towards warehousing costs, expenses for capital commitment and number of parts per 40 ft container of the three parts.

Comparing the two strategies, one can find that in the case of CKD all parts for one car are delivered in one container, leading to a minimal stock in both countries, since exactly the needed amount of parts can be delivered. But the additional costs for CKD-packing and CKD-unpacking have to be taken into account. Considering the other strategy, PbP, these extra costs are obsolete, but since every part is sent separately, a 40 ft container full of part x includes parts for n cars, depending on the size of the part. We assume that sending half filled containers is not feasible at all considering the disproportionally high transportation costs of 1500$ per container. Hence it could happen that, depending on the sales, one container contains parts for more cars than the amount built in one week. This holds especially true, when the transport frequencies are adaptable since applying the transport frequency needed to fill whole containers of smaller parts yields in higher stocks. Thus, on the one hand parts have to be pre-produced in Germany, on the other

stored in China until they are needed. Figure 10 shows the stock behaviour of the two strategies.
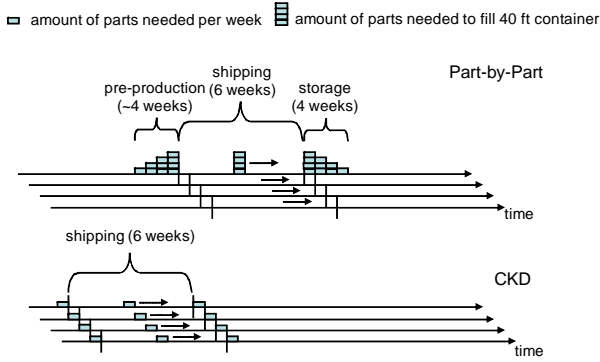

Figure 10: Principal behaviour of the stocks in both strategies

Since the amount needed per week is equal to one unit, but the amount of a certain part to fill one container is four units, the supplier of that certain part, using PbP, has to pre-produce for four weeks, thus sending parts to China only one time a month. Using CKD complete kits are sent, each filling one container, so that the total amount can be divided into four deliveries leading to a minimal stock.

The scenario is now divided into two parts. First of all we will decide which is the most suitable – that is economic – strategy for a certain sales figure by comparing the simulation results for both strategies. Secondly, to point out that selecting different strategies according to the heights of sales makes sense and the re-design process is necessary, we are going to change the sales, to find the point were switching the strategy becomes feasible.

So the questions to be answered are: "Which strategy is more feasible regarding a certain sales figure?", "How many \$ could then be saved per part?" and "At what amount of sales per month is it feasible to deliver a certain part directly to China (using PbP), when considering dynamic system behaviour?"

To find an answer to these questions we will model the two strategies using the OTD-NET environment, apply simulation and calculate the total cost for both cases.

$$C_{PbP} = \text{cstock}("germany") + \text{cclr}("germany") \\ + \text{cshp} + \text{cclr}("china") + \text{cstock}("china") \tag{1}$$

$$C_{CKD} = \text{cstock}("germany") + \text{cwh}("germany") \\ + \text{cclr}("germany") + \text{cshp} + \text{cclr}("china") \\ + \text{cwh}("germany") + \text{cstock}("china") \tag{2}$$

The total costs for the PbP strategy (1) consist of the costs for stock in Germany (3), the costs for customer clearance in Germany (6), the shipping costs (7), the costs for customer clearance (6) in China and the costs for the stock at the factory in China (3).

$$\text{cstock}(country) = \sum_{part} \text{chold}(part, country) + \text{crent}(country) \tag{3}$$

$$\text{chold}(part, country) = \text{avgstock}(part, country) \\ \cdot \text{val}(part) \cdot \text{irate}(country) \tag{4}$$

$$\text{crent}(country) = \max_{t}(\sum_{part} \text{stock}(part, t, country) \cdot \text{size}(part)) \\ \cdot \text{rentsqm}(country) \tag{5}$$

The costs for the stock in a country are the sum of the holding costs of all parts (4) plus the rent for the warehouse (5) calculated based on the maximal space needed.

$$\text{cclr}(country) = \#ships \cdot \text{clrtaxe}(country) \tag{6}$$

The customer clearance is a tax paid per ship (6).

$$\text{cshp} = \#container \cdot \text{containerfee} \tag{7}$$

The costs for the sea transportation (7) are paid per 40 ft container.

The total costs for the CKD strategy (2) are equal to the costs for the PbP strategy except for the additional costs for the CKD-packing and CKD-unpacking – that is the warehousing in Germany (8) and the warehousing in China (8).

$$\text{cwh}(country) = \#container \cdot \text{cpacking}(country) \\ + \sum_{part} \text{chold}(part, country) + \text{crent}(country) \tag{8}$$

The warehousing costs are the sum of the salaries paid for the packing of the containers, the holding costs (4) and the rent for the warehouse (5).

Now that the scenario and the cost functions have been described, the results will be presented in the next section.

5.2. Results

We assume that applying the first step of the method – the typology-based pre-selection – to the scenario described above has yield that two of the strategies to enter developing markets – CKD or PbP in this case – are possible candidates to apply, while the others were already filtered out. Now that there are only two candidates left we apply the second step using simulation to get to the final decision.

As stated in the chapter above three questions to be answered via simulation are: "Which strategy is more feasible regarding a certain sales figure?", "How many \$ could then be saved per part?" and "At what amount of sales per month is it feasible to deliver a certain part directly to China (thus using PbP), when considering dynamic system behaviour?"

To get an answer to the first question the cost function presented in chapter 5.1 is evaluated for the simulation results of both scenarios. A basic sales figure is used for this comparison (Figure 11).

As you can see the amount of products demanded in the whole year for the basic scenario is 4380, which may indicate an early phase of the life cycle.
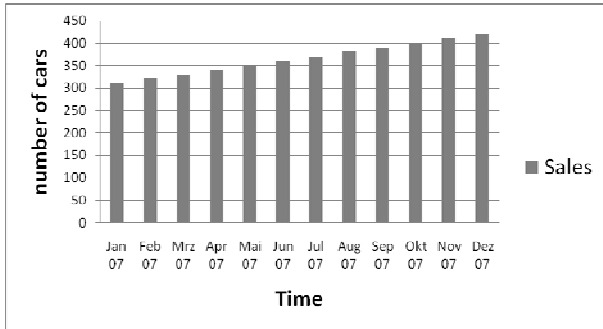

Figure 11: Sales per Month (basic scenario)

Comparing the simulation results of the two scenarios, we made some observations. First of all the cost for the shipping exceed the other costs 7 times (Figure 12) – which confirms our assumptions, that the examination of sending containers that are not completely filled, is unnecessary.
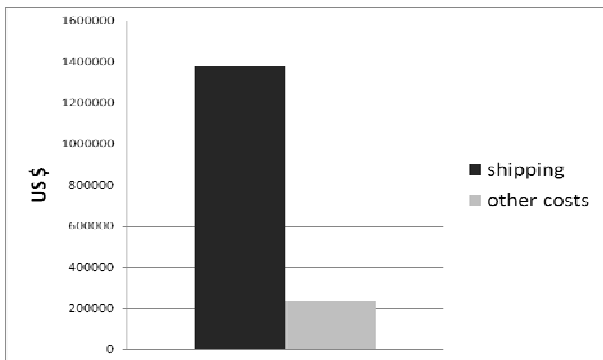

Figure 12: Comparison of shipping costs and remaining costs

Secondly the distribution of the cost of both strategies differ strongly form each other. Figure 13 shows the distribution for the basic sales scenario of the CKD-strategy.
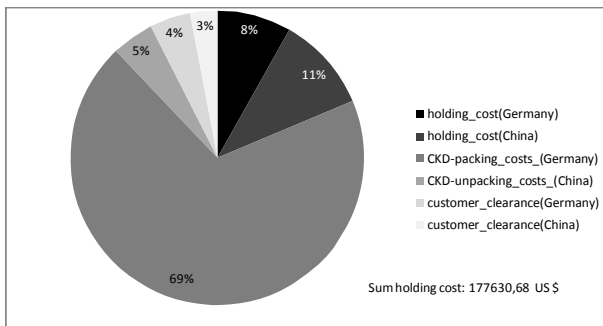

Figure 13: Cost distribution for the CKD-strategy

The costs for the repackaging in Germany exceed the costs in China, which is due to the higher salaries for the warehouse personal. The holding costs are relatively low for this strategy, since only the parts for the production program for one week are stored.
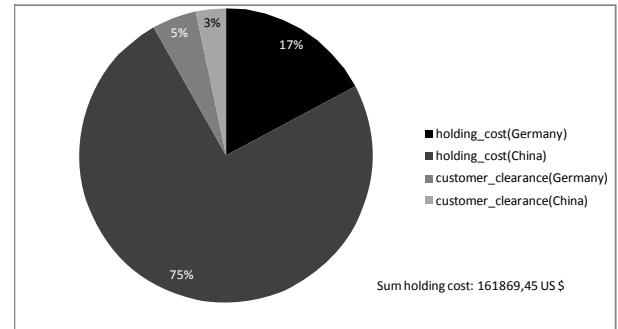

Figure 14: Cost distribution for the PbP-strategy

Figure 14 shows the cost distribution for the PbP-strategy. Compared to the CKD-strategy one can see that storing the parts is the primary cost factor, which confirms the assumptions made in chapter 5 about the indispensability of preproduction in Germany and storing in China.

To answer the first question the comparison of the total costs for both scenarios yield that at this phase the CKD-strategy is the cheaper one. The actual amount of money that can be saved selecting the right strategy can be obtained from the simulation results as well. For this scenario it is a maximum of 23 US $ per product, that is 23 US $ times 4380 products = 100740 US $ for the basic sales scenario.

To answer the last question, which is the right point in the life cycle (that is: the amount of sales reached) to change the strategies, we applied different sales figures to both scenarios. These figures were created by multiplying the basic figure with several different values. The results were as follows.

On the one hand, regarding the CKD-strategy, the distribution of the costs doesn't differ much for the different sales figures. This can be comprehended taking into account that for this strategy the costs are directly linked to the amount of products ordered, since every product has a single container. On the other hand the costs for the PbP-strategy differ plainly. In Figure 15 the holding costs are broken down to the single parts for the PbP Strategy and two sales scenarios.
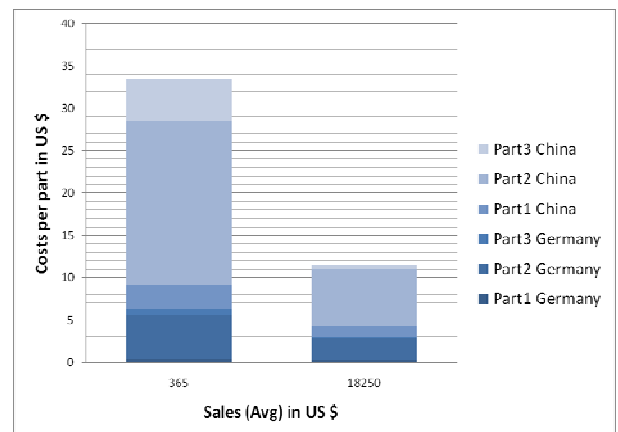

Figure 15: Holding costs per part for two sale figures

In Figure 15 it can be observed that the highest costs have to be allocated for part 2, the engine, which also has the highest value. Nevertheless, when focussing on the proportional amount of money that can be saved per part, increasing the sales of part 3, the headrest, is the most promising option. While the seat's and engine's costs are decreased by 40 % the headrest's costs can be decreased by 90%. That is because the amount of parts that fit in one 40 ft container is much higher for part 3 than for the other two. Hence a container filling for the other two parts lasts a shorter period of time and thus fewer parts must be stored at both sides of the supply process. Therefore the possible amount of money that can be saved through increasing the sales is far higher for part 3, since the other two parts are already transported optimally.

Now that the possibilities to save holding costs by increasing the sales using the PbP-strategy have be illustrated, the comparison of both strategies regarding the total costs for different sales will be shown. Figure 16 presents this comparison.
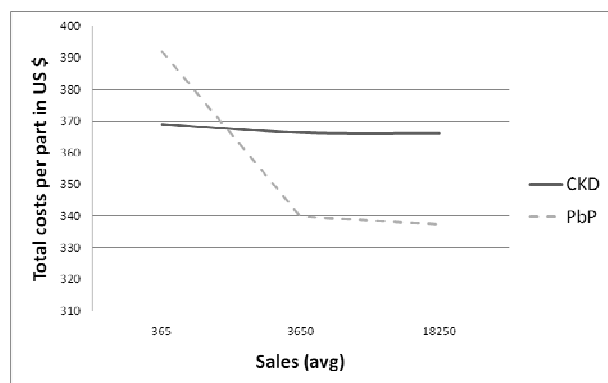

Figure 16: Comparison of the total costs for both strategies and different sales

The costs of the CKD-strategy are more or less constant for different sales, while the costs of the PbP-strategy decrease with the sales. Thus when the sales are low (e.g. at the beginning of a product life cycle) the CKD-strategy is cheaper since the holding costs are minimized. However, at a certain amount of sales it is more feasible to switch to the PbP-strategy (only considering the process point of view). That is because of the saved process cost for the omission of the consolidation on the one hand, and for the decreasing holding cost due to the rising sales on the other hand.

It has been exemplarily shown that this strategy switching point in the life cycle can be named using simulation and that the actual money saved can be numbered. Furthermore, it has been indicated that simulation can be used to rate different supply strategies with respect to the aspect of consolidation of parts. An optimisation of the process costs can be obtained by deciding for single parts depending on their value and size if a consolidation makes sense or not. Additionally, the whole process and the process costs can be made

transparent to the decision maker through modelling and simulation.

6. RESULTS AND OUTLOOK

We started identifying the supply strategies to enter developing markets and outlined a method for the strategic re-design in this application domain. A morphology was presented classifying the strategies and building the basis of a typology that will be presented shortly. Built upon this typology driven pre-selection a simulation-based comparison was presented focussing on the process parameter. The method was validated with a scenario focussing on the process cost of transportation and warehousing of either CKD or PbP. Not only the concrete costs could be named but it could be shown that switching the strategy when comparing different sales figures can save money. The results also show that a focus of further work in this sector should concentrate on minimising the amounts of containers used, since this is the main cost factor. Furthermore other aspects of the process must be taken into account such as:

- applying more fluctuating sales figures
- cost for extra transportation due to lost or damaged goods
- stock heights due to planning inaccuracies comparing planning control by the exporter or the importer

Beyond this special comparison the other strategies have to be taken into account as well. Thus further research will on the one hand include proofs of the concepts for the other strategy comparison and on the other hand include the permanent testing and extending of the typology.

Moreover the application of the method in the industry integrated into a decision support system will be aimed at.

REFERENCES

Chung, M.K., 1998. Globalization strategies of Korean motor vehicle industry: a case study of Hyundai, in: *Actes du GERPISA (Groupe d'Étude et de recherche Permanent sur l'Industrie et les Salariés de l'Automobile)*, 22.

Elkin, M., 2006. In the middle of a chain reaction - BMW, *Automotive Logistics*, July/August, 30-36.

Elman, C., 2004. *Theoretical Typologies in the Qualitative Study of International Politics*, Arizona State University, Department of Political Science.

Fukase, E., Martin, W., 1999. *A Quantitative Evaluation of Vietnam's Accession to the ASEAN Free Trade Area (AFTA)*, Development Research Group, World Bank,Washington, DC.

Grant, E.B., Gregory, M.J., 1997. Adapting manufacturing processes for international transfer,

International Journal of Operations & Production Management, 17 (10), 994-1005.

Hellingrath, B., Toth, M., Wagenitz, A., Motta, M., 2004. Simuliere und herrsche. In: *Beschaffung Aktuell,* 10. Konradin Verlagsgruppe, Echterdingen.

Hong, Y., 2008. Distinctive Characteristics of China's Path of ICT Development: A Critical Analysis of Chinese Developmental Strategies in Light of the Eastern Asian Model, *International Journal of Communication,* 2, 456-471

Humphrey, J., Memedovic, O., 2003. *The Global Automotive Industry Value Chain: What Prospects for Upgrading by Developing Countries*, UNIDO (United Nations Industrial Development Organisation), Vienna.

Jun, Y., 1987. The Reverse Direct Investment: The Case of the Korean Consumer Electronics Industry, *International Economic Journal*, 1(3), 91 – 104, Autumn.

Kaltofem, B., 1999. *Erschließung neuer, geschlossener Märkte*, Audi AG, Ingolstadt.

Kinra, N., 1995. Strategic dimensions in marketing planning: large versus small/medium companies in the Indian television market, *Marketing Intelligence & Planning*, 13(4), 34-44.

Kluge, S., 2000. Empirically Grounded Construction of Types and Typologies in Qualitative Social Research, *Forum Qualitative Social Research,* 1 (1).

Laux, J.M., 1992. *The European Automobile Industry*. Twayne Publishers, New York.

Moran, T. H., 2000. The product cycle model of foreign direct investment and developing country welfare, *Journal of International Management,* 6, 297-311.

Mukherjee, A., Sastry, T., 1996. *The Automotive Industry in Emerging Economies: A Comparison of Korea, Brazil, China and India*, Indian Institute of Management, Ahmedabad, India.

Petri, H., 2004. *Herausforderungen und Chancen für die Automobilproduktion am Standort Deutschland*, Mercedes Car Group –Produktion, Juni.

Sieben, B., Schwede, C., Song, Y., Wagenitz A., Hellingrath, B., 2008. Development and Simulative Validation of a Typology-Driven Method for the Re-design of International Supply Chains. Submitted to: *Sixth Annual International Symposium on Supply Chain Management*, October, Calgary.

Schieferdecker, R., 2003. *Produktionsplanung und - steuerung bei flexiblen Arbeitszeiten*, Aachen.

Schmitt, P., Thiesse, F., Fleisch, E., 2007. Adoption and Diffusion of RFID Technology in the Automotive Industry, *15th European Conference on Information Systems (ECIS 2007)*, E-DIGIT by SIGADIT, St.Gallen, Switzerland.

Smitka, M. J., 1990. The Invisible Handshake: The Development of the Japanese Automotive Parts Industry, *Business and Economic History*, 19 (2).

Tan, Z. A., 2002. Product cycle theory and telecommunications industry - foreign direct investment, government policy, and indigenous manufacturing in China, *Telecommunications Policy*, 26, 17–30.

Toth, M., Hellingrath, B., Wagenitz, A., Klingebiel, K., 2006. Eine dynamische Bewertungsmethodik für Build-to-Order Strategien, *Themenspezial – Automobilindustrie.*

Toth, M., Hellingrath, B., Wagenitz, A., Klingebiel, K., Spittank, S., 2004. OTD-Sim (Order-To-Delivery). In: Schnell, M.; Laakmann, F. (Hrsg.), ed. *Wege zur innovativen Fabrikorganisation*, Band 3. Verlag Praxiswissen, Dortmund.

Tulder, R.V., Ruigrok, W., 1997. European Cross-National Production Networks in the Auto Industry: Eastern Europe as the Low End of European Car Complex, *Foreign Direct Investment and Trade in Eastern Europe: The Creation of a Unified European Economy* , June 5-6, Vienna.

VDI-Richtlinie 3633, 1993. *Simulation von Logistik-, Materialfluss- und Produktionssystemen*. Beuth, Berlin.

Volpato, G., 2002. Carmakers Internationalisation Strategies: An Overview. In: *La Lettre Du GERPISA*, N° 158, Février – Mars.

Wagnera, S. M., Johnson, J. L., 2004. Configuring and managing strategic supplier portfolios, *Industrial Marketing Management,* 33, 717–730.

Wagenitz, A., 2007. *Modellierungsmethode zur Auftragsabwicklung in der Automobilindustrie.* Available from: https://eldorado.uni-dortmund.de/bitstream/2003/24408/1/Dissertationsschrift_Axel_Wagenitz.pdf. Dissertation, Dortmund.

DECISION SUPPORT SYSTEM FOR UREA SYNTHESIS SYSTEM OF A FERTILIZER PLANT

Sunand Kumar [a], Sanjeev Kumar[b], Puran Chandra Tewari [c]

[a]Department of Mechanical Engineering, National Institute of Technology, Hamirpur Himachal Pradesh (India).
[b]Department of Mechanical Engineering, A.K.G. Engineering College, Ghaziabad, Uttar Pradesh (India).
[c]Department of Mechanical Engineering, National Institute of Technology, Kurukshetra, Haryana (India).

[a]sunand@nitham.ac.in, [b]getsan25@rediffmail.com, [c]pctewari1@rediffmail.com

ABSTRACT

This paper introduces a decision support system for urea synthesis system of a fertilizer plant. It consists of five subsystems arranged in hybrid configurations. Decision support system for urea synthesis system has been designed with the help mathematical formulation using probabilistic approach. For this purpose, differential equations have been developed. Then steady state probabilities have determined. Besides, a pay off matrix is also developed which provides the various availability levels (Aii) for the different combinations of failure and repair rates for each subsystem. The optimum values of failure and repair rates for each subsystem are also determined. This decision model could be useful in a comparative evaluation of alternative maintenance strategies. So, the results of this paper would be highly useful in determining the optimal maintenance strategy, which will ensure the maximum availability of urea synthesis system in a fertilizer plant.

Key Words: Decision Support System, Decision Matrix, Availability.

1. INTRODUCTION

Fertilizer in India has emerged as one of the major import to meet the domestic demand. This gap between production and consumption is widening slowly because of lagging behind the production and increasing of the demand. Continuous efforts are needed to enhance production, which is based upon the development of efficient machines and maintenance strategies for full utilization of all the available resources (Kumar, Singh and Singh, J. 1988; Kumar and Pandey 1993). Production of fertilizer consists of three phases: developing phase (chemicals which are brought to the factory site), technological phase (balancing of equipment and input material, matching of man and machine) and research phase (modification of process parameters to obtain higher volume of fertilizers of specific qualities within the factory constraints (nature of the equipments used and concentration process).

There are physical limitations to develop the new chemicals, simultaneously, most of the Indian fertilizer plants are fairly well set i.e. their equipments are standardized and at a given capital investment, most of the companies are reluctant to add in the existing system unless they observe high benefit cost ratios (Sunand, Dinesh and Mehta 1996; Sunand, Dinesh and Mehta 1999).

A logical extension of the efforts to increase production in the fertilizer industry would be the maximization of fertilizer production at manufacturing stage (maximum utilization of maintenance crew and machine) keeping all other factors of the plant constant. In fertilizer industry a large number of equipments/ components are assembled to a basic skeleton to produce a final product, where the materials flow due to pressure difference from one section to another in some prearranged sequence. At each section a byproduct is separated due to change of pressure and / or temperature. Each section has a peculiar skeleton, each unit has a peculiar nature, and to achieve the goal of maximum production it is necessary to analysis the behavior of each unit of the section to economize operational parameters resulting in improved system availability. The analysis also helps in improving the process design and for reliable operation of the process (Dhillon and Singh 1981). A probabilistic analysis of the system under given operative conditions is helpful in the design modification for minimum failure of the units and to achieve the goal of optimum system availability. The units of the section are subjected to random failure and can be brought back into service by properly planned maintenance (Srinath 1994). Their operating behavior in the existing company environment is difficult to predict without knowing their

interrelationship. Mathematical modeling is a tool to develop these relationships.

The fertilizer plant is a very large and complex engineering unit having continuous production of urea. It comprises of various systems viz. urea synthesis, urea decomposition, urea crystallization and urea prilling systems etc. For smooth running, each of its system should remain in upstate. One of the most important functionary systems is urea synthesis system, the subject of our discussion. In fertilizer plant the CO_2 gas and liquid ammonia NH_3 are fed to the urea synthesis reactor. The reactor is maintained at 190^0C temperature and 250 atm pressure and allows CO_2 and NH_3 to react forming urea in gaseous form. A repairable system is characterized by a large number of interconnected components with their own failure behavior and repair time distributions. System availability in such a case is a complex function of failure and repair time distributions of individual components within the system (Sunand, Dinesh and Mehta 2000, Sunand, Tewari and Rajiv 2007). The decision support system for the plant has been developed on the basis of an actual study conducted in a medium sized urea fertilizer plant situated near Delhi. The failure /repair time data in the plant were studied and based on the model behavioural analysis of the system is carried out. This paper discusses the decision support system of the urea synthesis system of a fertilizer plant.

2. THE UREA SYNTHESIS SYSTEM

The urea synthesis system comprises of a compressor used to compress the carbon dioxide, two reciprocating pumps used to boost the pressure of liquid ammonia and heaters, which are used to heat the ammonia gas. In this process the CO_2 gas and liquid ammonia NH_3 available from ammonia production process are fed to the urea synthesis reactor. In reactor these gases are react to forming urea in gaseous form (Kumar, Tewari and Sanjeev 2007).

2.1 System Description

The urea synthesis system consists of five subsystems arranged in series:

i) Subsystem (A1): CO_2 booster compressor is a single unit arranged in series. Its failure causes the complete failure of the system.

ii) Subsystem (A2): CO_2 compressor arranged in series, Failure of this subsystem causes the complete failure of the system.

iii) Subsystem (A3): Three NH_3 preheaters arranged in series. Failure of any one causes the complete failure of the system

iv) Subsystem (H): It consists of four liquid ammonia feed pumps arranged in parallel.

Two pumps remain operative in parallel and other two in cold standby. Failure of three pumps at a time will cause complete failure of the system.

v) Subsystem (L): It comprises of three recycle solution feed pumps arranged in parallel. Failure of units reduces the capacity of the system but complete failure occurs when all units fail at a time.

2.2 Assumptions

The assumptions used in the performance modeling are (Dhillon and Singh 1981):

i) Failure/repair rates are constant over time and statistically independent.

ii) A repaired unit is as good as new one, far as performance is concerned.

(iii) Each subsystem has separate repair facility; there is no waiting time for repair in system.

(iv) The standby units/subsystems are of the same nature

(v) The system failure/repair follows exponential distribution.

(vi) The service includes repair and/or replacement of the units/subsystems.

(vii) System may work at reduced capacity.

(viii) There are no simultaneous failures.

The transition diagram (Fig.1) of the urea synthesis system shows the various possible states, the system can acquire. Based on the transition diagram, a performance-evaluating model has been developed. The failures and repairs for this purpose have been modeled as birth and death process.

2.3 Nomenclature

The symbols and notations associated with transition diagram of the urea synthesis system.

|	
◯	Indicates the system in operating condition.
▭	Indicates the system in breakdown condition
⬭	Indicates the system in reduced capacity state.
Ai, H,L (i=1,2,3)	Indicate that the subsystems are working at full capacity.
Ai, h, λ	Indicates that all subsystems are in complete failed state.
H1,H2	Indicate that stand-by unit of the sub-system H is in operating state.
L1, L2	Indicate that the subsystem L is working at reduced capacity.
$P_0(t)$	Probability of the system working with full capacity at time 't'.
$P_1(t), P_6(t)$	Probability of the system in cold standby state.
$P_2(t) - P_5(t)$,	Probability of the system in

94
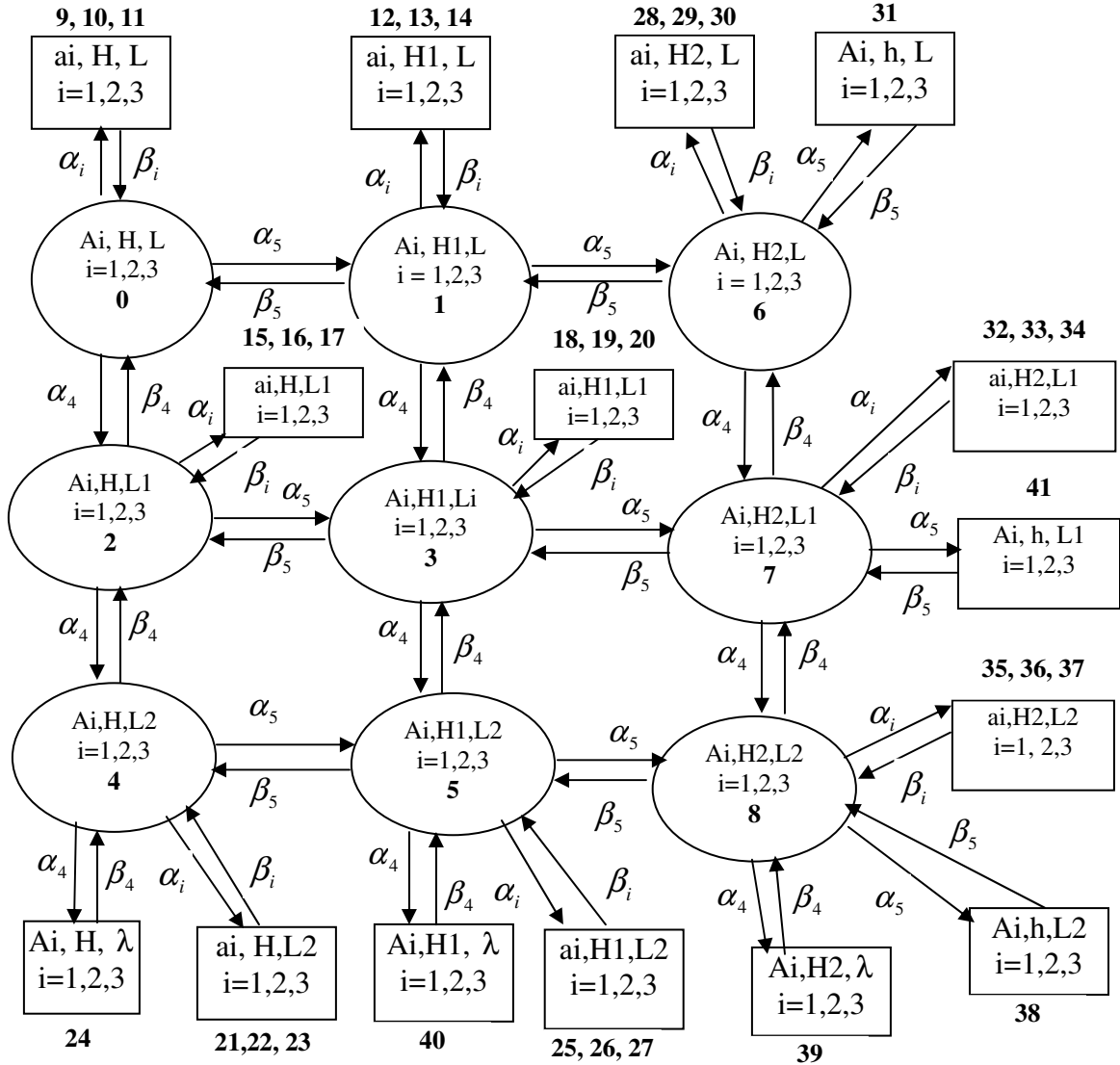
Figure 1: Transition Diagram of the Urea Synthesis System

$P_7(t), P_8(t)$ reduced capacity state.

$P_9(t) - P_{41}(t)$ Probability of the system in failed state.

α_i, i =1,2,3,4,5   Mean failure rate in Ai, H,L

β_i, i=1,2,3,4,5   Mean rate of repairs in Ai, H, L.

d/dt   Represents derivative w.r.t. time 't'

3. MATHEMATICAL MODELING OF THE UREA SYNTHESIS SYSTEM

The mathematical modeling is done using simple probabilistic considerations and differential equations are developed using Markov birth-death process. If the state of the system is probability based, then the model is a Markov probability model. The present availability analysis is concerned with a discrete-state continuous-time model, is also called a Markov process. Markov model is defined by a set of probabilities p_{ij} where p_{ij} is the probability of transition from any state i to any state j. For example, the equipment transits from operable state (i) to failed state (j) with probability P_{ij}. One of the most important features of the Markov process is that the transition probability p_{ij}; depends only on states i and j and is completely independent of all past states except the last one, state i (Srinath 1994).

The objective here is to obtain an expression for the probability of n occurrences in time t. Let the probability of n occurrences in time t be denoted by $P_n(t)$, i.e., Probability $(X = n, t) = P_n(t)$ (n = 0, 1, 2, . .). Then, P_0 (t) represents the probability of zero occurrences in time t. The probability of zero occurrences in time (t + Δt) is given by

$$P_0(t + \Delta t) = (1 - \beta \, \Delta t) P_0(t) \text{ i.e.}$$

The probability of zero occurrences in time (t + Δt) is equal to the probability of zero occurrences in time t multiplied by the probability of no occurrences in time Δt. The probability of no occurrences in time Δt is obviously given by $(1 - \beta \Delta t)$. The probability of one occurrence in time (t + Δt) is composed of two parts, namely, (a) probability of zero occurrences in time t multiplied by the probability of one occurrence in the interval Δt and (b) the probability of one occurrence in time t multiplied by the probability of no occurrences in the interval Δt. Thus,

$$P_1(t + \Delta t) = (\alpha \Delta t)P_0(t) + (1 - \beta \Delta t)P_1(t)$$

$$P_1(t + \Delta t) - P_1(t)/\Delta t = \alpha P_0(t) - \beta P_1(t)$$

Or

$$Lt \ \Delta t \to 0$$

$$d/dt \ P_1(t) + \beta P_1(t) = \alpha P_0(t)$$

$$or \quad (d/dt + \beta)P_1(t) = \alpha P_0(t) \qquad (A)$$

Using the concept in equation (A), and considering constant failures and repair rates the mathematical formulation is done using probabilistic Markov birth-death approach. The various probability considerations give the following differential equations associated with the urea synthesis subsystems and these equations are solved for determining the steady state performance of the urea synthesis system.

$$(d/dt + \sum_{i=0}^{5} \alpha_i)P_0(t) = \sum_{i=1}^{3} \beta_i P_{i+8}(t) + \beta_4 P_2(t) + \beta_5 P_1(t) \qquad (1)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_5)P_1(t) = \sum_{i=1}^{3} \beta_i P_{i+11}(t) + \beta_5 P_6(t)$$
$$+ \beta_4 P_3(t) + \alpha_5 P_0(t) \qquad (2)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i)P_2(t) = \sum_{i=1}^{3} \beta_i P_{i+14}(t) + \beta_4 P_4(t) + \beta_5 P_3(t) \qquad (3)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_5 + \beta_4)P_3(t) = \sum_{i=1}^{3} \beta_i P_{i+17}(t)$$
$$+ \beta_5 P_6(t) + \beta_4 P_5(t) + \alpha_5 P_2(t) + \alpha_4 P_1(t) \qquad (4)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_4)P_4(t) = \sum_{i=1}^{4} \beta_i P_{i+20}(t) + \beta_5 P_5(t) + \alpha_5 P_2(t) \qquad (5)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_5 + \beta_4)P_5(t) = \sum_{i=1}^{3} \beta_i P_{i+25}(t) + \beta_5 P_{27}(t)$$
$$+ \beta_4 P_{40}(t) + \alpha_5 P_4(t) + \alpha_4 P_3(t) + \beta_5 P_8(t) \qquad (6)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_5)P_6(t) = \sum_{i=1}^{3} \beta_i P_{i+27}(t) + \beta_5 P_{31}(t)$$
$$+ \beta_4 P_7(t) + \alpha_5 P_1(t) \qquad (7)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_5 + \beta_4)P_7(t) = \sum_{i=1}^{3} \beta_i P_{i+31}(t)$$
$$+ \beta_4 P_8(t) + \alpha_5 P_3(t) + \alpha_4 P_6(t) + \beta_5 P_{41}(t) \qquad (8)$$

$$(d/dt + \sum_{i=0}^{5} \alpha_i + \beta_5 + \beta_4)P_8(t) = \sum_{i=1}^{5} \beta_i P_{i+34}(t)$$
$$+ \alpha_5 P_5(t) + \alpha_4 P_7(t) \qquad (9)$$

$$(d/dt + \beta_m)P_i(t) = \alpha_m P_j(t) \qquad (10)$$

Where in equation (10), for
m =1: then,
i = 9,j =0; i =12,j =1; i =15,j =2; i =18, j =3;i =21, j =4; i =25,j =5; i =28,j =6; i =32,j =7; i =35,j =8
m =2: then,
i = 10,j =0; i =13,j =1; i =16,j =2; i =19, j =3; i =22, j = 4; i =26, j = 5; i =29,j =6; i =33,j =7; i =36,j =8
m=3: then,
i = 11,j =0; i =14,j =1; i =17,j =2; i =20, j =3;i =23, j =4; i =27, j =5; i =30,j =6; i =34,j =7; i =37, j =8
m=4: then,
i =24,j =4; i =40,j =5; i =39,j =8
m=5: then,
i =31,j =6; i =41,j =7; i =38,j =8

3.1 Solution of Equations
With the initial condition $P_0(0) = 1$, otherwise zero since any urea plant is a process industry where raw material is processed through various subsystems continuously till the final product is obtained. Thus, the long run availability of the urea synthesis system of an urea plant is attained by putting derivative of all probability equal to zero as

$$d/dt = 0 \qquad at \quad t \to \infty$$

into differential equations, one gets,

$$P_i = (\alpha_m / \beta_m)P_j \qquad (11)$$

By putting the values of probabilities from equations 11 in equations 1-9, one can represents

$C1P_0 = \beta_5 P_1 + \beta_4 P_2$	$C1 = \alpha_5 + \alpha_4$	(12)
$C2P_1 = \beta_5 P_6 + \beta_4 P_3 + \alpha_5 P_0$	$C2 = C1 + \beta_5$	(13)
$C3P_2 = \beta_5 P_3 + \alpha_4 P_0 + \beta_4 \ P_4$	$C3 = C1 + \beta_4$	(14)
$C4P_3 = \alpha_4 P_1 + \alpha_5 P_2 + \beta_5 P_7 + \beta_4 P_5$	$C4 = C3 + \beta_5$	(15)
$C5P_4 = \beta_5 P_5 + \alpha_4 P_2$	$C5 = \alpha_5 + \beta_4$	(16)
$C6P_5 = \beta_5 P_8 + \alpha_4 P_3 + \alpha_5 P_4$	$C6 = C5 + \beta_5$	(17)
$C7P_6 = \beta_4 P_7 + \alpha_1 P_1$	$C7 = \alpha_4 + \beta_5.$	(18)
$C8P_7 = \beta_4 P_8 + \alpha_4 P_6 + \alpha_5 P_3$	$C8 = C7 + \beta_4$	(19)
$C9P_8 = \alpha_4 P_7 + \alpha_5 P_5$	$C9 = \beta_5 + \beta_4$	(20)

3.2 Solving these equations recursively

$P_1=C17\ P_0$ $C17= C14+C15+C16 / (1+C12) +$ C11/ (C13/(1+C12)+C10   (21)

$P_2 =C19\ P_0$ $C19= (C1- (\beta_5*C16))/\beta4$ (22)

$P_3=C18\ P_0$ $C18= (C13/(1+C12)*C17)-C16$ (23)

$P_4=C21\ P_0$ $C21= ((C3*C19)-\alpha_4- (\beta_5*C18))/C3$ (24)

$P_5=C22\ P_0$ $C22= ((C5*C21)-(\alpha4*C19))/\beta5$ (25)

$P_6=C20\ P_0$ $C20= ((C2*C17)-\alpha_5-(\beta_4*C18))/\beta_5$ (26)

$P_7=C23\ P_0$ $C23 = ((C7*C20)-(\alpha_5*C17))/\beta_5$ (27)

$P_8=C24\ P_0$ $C24= ((\alpha_4*C23)+(\alpha_5*C22))$ (28)

$C31= (\alpha_4/\ C4) – (\alpha_5*C5)/(C4*\beta4)) +(((C7*C2) + (\alpha5*\ \beta5)) / \beta4) - C5+\alpha4$

$C10=1/(1+C7*\beta_4+((\beta_4*C5)/C3))*C31$

$C32 = ((C7*\alpha_5) / \beta_4) + (\alpha_5*C1)- (((\beta_4* C5*\alpha_4) /C3)-(C5*C1)) /\beta_5$

$C11=1/(1+C7*\beta_4+((\beta_4*C5)/C3))*C32$

$C12= (C8*C7)/ (\alpha_5*\alpha_7)+(\alpha_4*\beta_4) / (\alpha_5*\beta_5)-(\alpha_5*C5) / C3- (\beta_4*\alpha_4) / ((\alpha_5*\ \beta_5*\ C9))) * ((C4+((\beta_4*C5)/C3)))$

$C13= (C2 / (\alpha_5*\ \beta_5)) * (((C8\ *C7*\beta_5) / \beta_4) - (\alpha_4))- ((C8 / \beta4) + ((\alpha_5*\alpha_4) / \beta_4) + (\alpha_4 / C9))$

$C14= (1/ (\beta_4*\beta_5)) * ((C8*C7) + (C1*\alpha5*\alpha4))$

$C15= ((\alpha_4*\beta_4*\beta_4) / (\beta_5*\beta_5*C9*\alpha_5)) * ((\alpha_5*C5) / C3 -C1)$

$C16= (\alpha_4/\beta_5)* (1+ ((\alpha_5*C5) / C3) + (C1/C9))$

The probability of full working capacity, namely, P_0 is determined by using normalizing condition: (i.e. sum of the probabilities of all working states, reduced capacity and failed states is equal to 1)

$$\sum_{i=0}^{41} P_i = 1$$ $P_0*N= 1$

Where N= (1+ C17+ C18+ C19+ C20+ C21+ C22+ C23+ C24)(1+ α_1/β_1+ α_2/β_2+ α_3/β_3)+ α_4/β_4)(1+ C21+ C24 + C22) + α_5/β_5 (1+ C20+ C24+ C23)   (29)

Now, the steady state availability of desulphurization system may be obtained as summation of all working and reduced capacity state probabilities as

Hence   Av. = $\sum_{i=0}^{8} P_i$

     Av. = $P_0 + P_1+ P_2+ P_3+ P_4+ P_5 + P_6+ P_7+ P_8$

Av. = (1+C17+C18+C19+C20+C21+ C22+ C23+ C24)/N   (30)

Therefore, Availability of the system (Av.) represents the performance model of the urea synthesis system. It can be used for developing the decision support system of this operating system of fertilizer plant.

4. DECISION SUPPORT SYSTEM

From maintenance history sheets of the urea synthesis system of fertilizer plant and through the discussions with the plant personnel, appropriate failure and repair rates of all subsystems are taken and decision matrices (availability values) are prepared accordingly by putting these failure and repair rates

values in expression (30) for Availability (Av.). The decision support system deals with the quantitative analysis of all the factors viz. courses of action and states of nature, which influence the maintenance decisions associated with the urea synthesis system of fertilizer plant (Tewari, Kumar and Mehta 2000).

These decision models are developed under the real decision making environment i.e. decision making under risk (probabilistic model) and used to implement the proper maintenance decisions for the urea synthesis system. Table 1, 2, 3, and 4 represent the decision matrices for various subsystems of the urea synthesis system. These matrices simply reveal the various availability levels for different combinations of failure and repair rates/priorities. These availability values obtained in decision matrices for all subsystems are then plotted. Figures 2, 3, 4 and 5 represent the plots for various subsystems of the urea synthesis system, depicting the effect of failure /repair rate of various subsystems on ash handling unit availability. On the basis of analysis made, the best possible combinations (α, β) may be selected.

Table 1: Decision Matrix of subsystem CO_2 Booster Compressor (A1)

$\alpha_2 = 0.005$, $\alpha_3= 0.01$, $\alpha_4= 0.002$, $\alpha_5= 0.004$
$\beta_2 = 0.1$, $\beta_3 = 0.5$, $\beta_4 = 0.1$, $\beta_5 = 0.4$

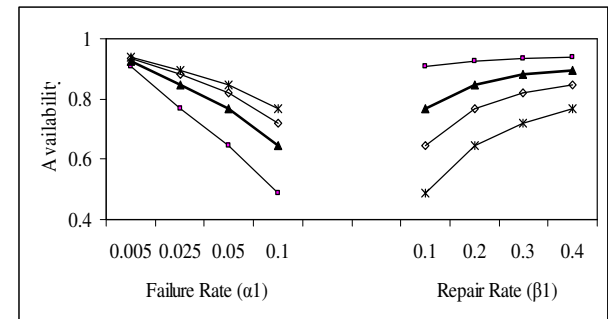| α_1 \ β_1	Availability (Av.)			
	0.1	0.2	0.3	0.4
0.005	0.9066	0.9277	0.9349	0.9336
0.025	0.7675	0.8489	0.8801	0.8965
0.05	0.6439	0.7675	0.8199	0.8489
0.1	0.4971	0.6439	0.7214	0.7675

Figure 2: Effect of Failure and Repair Rates of CO_2 Booster Compressor on the Availability of the Urea Synthesis System.

Table 2: Decision Matrix of subsystem CO_2 Compressor (A2)

$\alpha_1= 0.005$, $\alpha_3= 0.001$, $\alpha_4= 0.002$, $\alpha_5= 0.004$
$\beta_1= 0.1$, $\beta_3= 0.5$, $\beta_4= 0.1$, $\beta_5= 0.4$

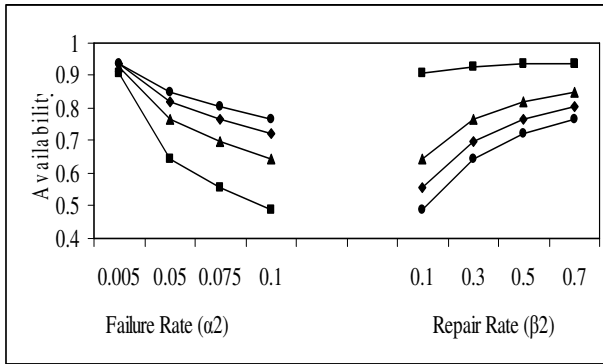| β_2 α_2	Availability (Av.)			
	0.1	0.3	0.5	0.7
0.005	0.9066	0.9277	0.9349	0.9386
0.05	0.6439	0.7675	0.8199	0.8489
0.075	0.5546	0.70	0.7675	0.8062
0.1	0.4871	0.6439	0.7214	0.7675

Figure 3: Effect of Failure and Repair Rates of CO_2 Compressor on the Availability of the Urea Synthesis System.

Table 3: Decision Matrix of subsystem NH_3 Preheater (A3)

$\alpha_1= 0.005$, $\alpha_2= 0.005$, $\alpha_4= 0.002$, $\alpha_5= 0.004$
$\beta_1= 0.1$, $\beta_2= 0.1$, $\beta_4= 0.1$, $\beta_5= 0.4$

β_3 α_3	Availability (Av.)			
	0.1	0.3	0.5	0.7
0.001	0.9000	0.9055	0.9067	0.9072
0.005	0.8688	0.8948	0.90	0.9025
0.009	0.8396	0.8842	0.8937	0.8978
0.013	0.8124	0.8739	0.8874	0.8933
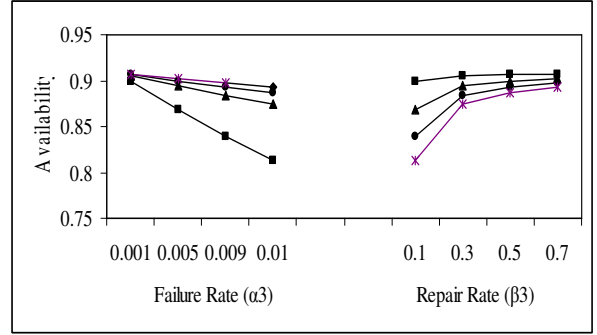
Figure 4: Effect of Failure and Repair Rates of NH_3 Preheater on the Availability of the Urea Synthesis System.

Table 4: Decision Matrix of subsystem NH_3 Feed Pump (H)

$\alpha_1= 0.005$, $\alpha_2= 0.005$, $\alpha_3 = 0.001$, $\alpha_4= 0.002$
$\beta_1= 0.1$, $\beta_2= 0.1$, $\beta_3= 0.5$, $\beta_4= 0.1$

β_5 α_5	Availability (Av.)			
	0.1	0.2	0.3	0.4
0.002	0.9016	0.9068	0.9070	0.9071
0.004	0.8892	0.9061	0.9066	0.9067
0.006	0.8731	0.9051	0.9061	0.9063
0.008	0.8548	0.9039	0.9057	0.9060
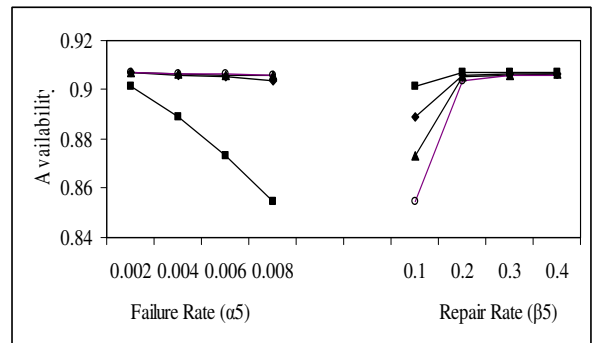
Figure 5: Effect of Failure and Repair Rates of NH_3 Feed Pump on the Availability of the Urea Synthesis System.

5. RESULTS AND DISCUSSION

Table 1 and figure 2 show the effect of failure and repair rates of CO_2 booster compressor on the availability of the urea synthesis system as failure rate of booster compressor (α_1) increases from 0.005 (once in 200 hrs) to 0.1 (once in 10 hrs), the system availability decreases significantly 41 %. Similarly as

the repair rate (β_1) increases from 0.1 (once in 10 hrs) to 0.4 (once in 2.5 hrs), the system availability also increases slightly to 2.7 %.

Table 2 and figure 3 depict the effect of failure and repair rates of CO_2 compressor on the availability of the urea synthesis system as failure rate of compressor (α_2) increases from 0.005 (once in 200 hrs) to 0.1 (once in 10 hrs) ,the system availability decreases very sharply 42 %. Similarly as the repair rate (β_2) increases from 0.1 (once in 10 hrs) to 0.7 (once in 1.43 hrs), the system availability increases by 3.2 %.

Table 3 and figure 4 highlight the effect of failure and repair rates of NH_3 preheater on the availability of the urea synthesis system, as failure rate of preheater (α_3) increases from 0.001 (once in 1000 hrs) to 0.013 (once in 77 hrs), the system availability decreases considerably 8.76 %. Similarly as the repair rate (β_3) increases from 0.1 (once in 10 hrs) to 0.7 (once in 1.43 hrs), the system availability increases hardly by 1 %.

Table 4 and figure 5 explain the effect of failure and repair rates of liquid NH_3 feed pump on the availability of the urea synthesis system as failure rate of heat exchanger (α_5) increases from 0.002 (once in 500 hrs) to 0.008 (once in 125 hrs), the system availability decreases significantly 4.68 %. Similarly as the repair rate (β_5) increases from 0.1(once in 10 hrs) to 0.4 (once in 2.5 hrs), the system availability increases marginally by 1 %.

6. CONCLUSIONS

The availability model and decision support system for the urea synthesis system has been developed with the help of mathematical modeling using probabilistic approach. The decision matrices are also developed. These matrices facilitate the maintenance decisions to be made at critical points where repair priority should be given to some particular subsystem of the urea synthesis system. Decision matrix as given in table 2 clearly shows that the CO_2 compressor is most critical subsystem as far as maintenance is concerned. So, CO_2 compressor subsystem should be given top priority as the effect of its repair rates on the unit availability is much higher than that of CO_2 booster compressor, NH_3 preheater and NH_3 feed pump. Therefore, on the basis of repair rates, the maintenance priority should be gives as per following order:

1. First priority should be given to CO_2 compressor.
2. Second priority should be given to CO_2 booster compressor.
3. Third priority should be given to NH_3 preheater.
4. Fourth priority should be given to NH_3 feed pump.

REFERENCES

Arora, N., Kumar, D., 1997. Availability Analysis of Steam and Power Generation Systems in Thermal Power Plant. *Microelectron Reliability,* 37(5), 795-799.

Dhillon, B.S., Singh, C., 1981. *Engineering Reliability- New techniques and applications.* John willey and sons, New York.

Kumar, D., Singh, I.P., and Singh, J., 1988. Reliability analysis of the Feeding System in the Paper Industry. *Microelectron Reliability,* 28(2), 213-215.

Kumar, D., Singh, I.P., and Singh, J., 1988. Availability of the feeding system in the sugar industry. *Microelectron Reliability,* 28(6), 867-871.

Kumar D., Pandey, P.C., 1993. Maintenance planning and resource allocation in urea fertilizer plant. *International Journal of Quality and Reliability Engineering,* 9, 411-423.

Kumar, S., Tewari, P.C., and Kumar, Sunand., 2007. Performance Modeling and Simulation of Urea Synthesis System of a Fertilizer Plant. *Proceedings of International Conference held at Mumbai (India),* 645-650

Kumar, S., Tewari, P.C., and Kumar, Sanjeev, 2007. Performance Modeling and Simulated Availability of Shell Gasification and Carbon Recovery Unit of Urea Plant. *Proceedings of the 16th IASTED International Conference held at Spain,* 409-413.

Rajiv Khanduja, Tewari, P.C., Dinesh Kumar., 2008. Availability Analysis of Bleaching System of Paper Plant. *Journal of Industrial Engineering, Udyog Pragati, N.I.T.I.E. Mumbai (India),* 32(1), 24-29.

Srinath, L.S., 1994. *Reliability Engineering. 3rd edition,* East-West Press Pvt. Ltd., New Delhi, India.

Shooman, M.L., 1996. Reliability Computation for Systems with Dependents Failures. *Proceedings of IEEE Annual Symposium on Reliability,* 44-56.

Sunand Kumar, Dinesh Kumar, Mehta, N.P., 1996. Behavioral Analysis of Shell Gasification and Carbon Recovery process in Urea Fertilizer Plant. *Microelectron Reliability*, 36(5), 671-673.

Sunand Kumar, Dinesh Kumar, Mehta, N.P., 1999. Maintenance Management for Ammonia Synthesis System in a Urea Fertilizer Plant. *International Journal of Management and System (IJOMAS),* 15(3), 211-214.

Sunand Kumar, Dinesh Kumar, Mehta, N.P., 2000. Probabilistic Analysis of Desulphurization System in Urea Fertilizer Plant. *Journal of Institution of Engineers (India),* 80, 135-139.

Sunand Kumar, Tewari, P.C., Sharma Rajiv. 2007, Simulated Availability of CO_2 Cooling System in a Fertilizer Plant. *Industrial Engineering Journal (Indian Institution of Industrial Engineering, Mumbai),* 36(10), 19-23.

Tewari, P.C., Kumar, D., Mehta, N.P., 2000. Decision Support System of Refining System of Sugar Plant. *Journal of Institution of Engineers (India),* 84, 41-44.

AUTHORS BIOGRAPHY

Sanjeev Kumar is Assistant Professor in the Department of Mechanical Engineering at AKG Engineering College, Ghaziabad, Uttar Pradesh (India). He received his Bachelor of Technology (B.Tech.) from National Institute of Technology, Kurukshetra, Haryana (India) in 1998 and his Master of Technology from National Institute of Technology in 2002. He has 10 years of experience in teaching and research. He has published and presented fifteen papers in International and National Conferences. Presently he is pursuing his Ph.D. in Industrial Engineering from National Institute of Technology, Kurukshetra, Haryana (India).

Dr. P.C. Tewari is Assistant Professor in the Department of Mechanical Engineering at National Institute of Technology, Kurukshetra, Haryana (India). He received his Bachelor of Engineering (B.E.) and Master of Engineering (M.E.) from Indian Institute of Technology, Roorkee (India) and his Ph.D. from Kurukshetra University, Kurukshetra, Haryana (India). He has 20 years of experience in teaching and research. He has published about fifty papers in International and National Journals/ Conferences. His fields of interests are Maintenance Engineering and Quality and Reliability Analysis.

Dr. Sunand Kumar is Professor and Ex-Head in the Department of Mechanical Engineering at National Institute of Technology, Hamirpur, Himachal Pradesh (India). He received his Bachelor of Engineering (B.E.) from Regional Engineering College (Presently known as National Institute of Technology), Calicut (INDIA) and his Master of Engineering from Panjab University, Chandigarh (India) and Ph.D. from Kurukshetra University, Kurukshetra, Haryana (India). He has 24 years of experience in teaching and research. He has published about forty papers in International and National Journals/ Conferences. His fields of interests are Modeling Stochastic Behavior of System in Industry and Maintenance Engineering.

THE INDUSTRIAL INNOVATION PROCESS BY A NETWORK APPROACH

Nelson F. F. Ebecken[a] , Francisco A. Moreira Jr. [b]

[a] COPPE/Federal University of Rio de Janeiro
[b] COPPE/Federal University of Rio de Janeiro

[a]nelson@ntt.ufrj.br , [b]fmoreira@gmail.coml

ABSTRACT

The objective of this paper is to present the methodology adopted by Petrobras, the Brazilian oil company, in e-science collaborative projects to generate ideas of new opportunities obtained by technology convergence for an energy company that searches solutions for the future. The effort is based on contributions of a group of 50 senior scientists from top 5 research institutes working iteratively via the Web and text mining activities to look for new technologies.

Keywords: text mining, collaborative project, convergent technologies

1. INTRODUCTION

The convergence of nanotechnology, biotechnology, information technology, and cognitive science is transforming global society. Technological convergence is beginning to define the way societies interact themselves, the way science is done and the way the global marketplace is organized.

In middle 2006, Petrobras, a Brazilian oil company, established a pioneer multidisciplinary team of senior academics and end users working in partnership to find potential opportunities and benefits of technology convergence.

A scientific advisory board was formed consisting of representatives from top universities as well as Petrobras. The present membership includes major government-funded Brazilian research institutes: Federal University of Rio de Janeiro, Federal University of Rio Grande do Sul, State University of São Paulo, State University of Campinas and the Aeronautic Technology Institute. The function of the advisory board is to channel information from Petrobras to faculty members concerning problems that need research attention and to maintain the focus.

In terms of the technology to enable this kind of online scientific collaboration there is much to choose from: blogs, mailing lists, wikis, forums, consensus Web filters, etc. What is shocking are the cultural and organizational obstacles — not monetary or financial obstacles — to getting individuals and organizations working together.

Many academics have said online collaboration won't work because, to put it in a nutshell, a) publish or perish means that work with no publications equals no motivation and b) who has time for this even if there was motivation? Also most of the scientists are pretty techno-adverse and they view hanging around on science blogs as a time sink.

In fact, most established PIs today were educated in very different circumstances to what undergraduates and postgraduates are today. From our observations most people will use the tools they are familiar with and find change very difficult. Many are unwilling to experiment with computers.

We've had so much push back from our non-geek colleagues on the blog and wiki idea that we started developing something different — a Web-based innovation engine which builds on respect for expertise, the scientific method, evidence-based texts, and, above all, the value of the individual scholar, author, and creator of knowledge.

2. IMPORTANT REMARKS

The start point to design this online collaboration tool was a relatively new breed of website that collaboratively votes on the best links, like Digg, Reddit, NowPublic and NewVines. Each of these popular aggregator sites supplies anonymous readers with pointers to news items that are ranked by other anonymous readers. None of these sites generates news; they only point to it by filtering the links to news items. Readers themselves vote on the importance of linked items suggested by other readers; these votes are then subjected to a complex formula to produce rankings.

That is part of the larger pattern of the appeal of a new online collectivism that is nothing less than a resurgence of the idea that the collective is all-wise. The so-called Digital Maoism is an unholy brew made up of digital utopianism; a desire to avoid individual responsibility; anti-intellectualism — the common disdain for pointy headed professors; and the corporatist "team" mentality (Gorman 2007, Lanier 2006).

The flight from expertise is accompanied by the opposite of expertise — the phenomenon that Andrew Keen has called, in his book of the same name, "the cult of the amateur," a world in which everyone is an expert in a world devoid of expertise.

The argument is that scholarship based on individual expertise resulting in authoritative statements is

somehow passé and that today's younger people think and act differently and prefer collective to individual sources because of their immersion in a digital culture.

If we deviate from the party line — radical epistemic egalitarianism — then we are Elitists, pure and simple, and therefore the enemy.

It's safer to be the aggregator of the collective. You get to include all sorts of material without committing to anything. You can be superficially interesting without having to worry about the possibility of being wrong, except when intelligent thought really matters. In that case the average idea can be quite wrong, and only the best ideas have lasting value. Science is like that.

The structures of scholarship and learning are based on respect for individuality and the authentic expression of individual personalities. The person who creates knowledge or literature matters as much as the knowledge or the literature itself. The manner in which that individual expresses knowledge matters too.

Scientific communities achieve quality through a cooperative process that includes checks and balances, and ultimately rests on a foundation of goodwill and "blind" elitism — blind in the sense that ideally anyone can gain entry, but only on the basis of a meritocracy.

Our online collaboration tool was designed to support the idea that individual scholars matter, not just the process or the collective. The collective is more likely to be smart when it isn't defining its own questions, when the goodness of an answer can be evaluated by a simple result, and when the information system which informs the collective is filtered by a quality control mechanism that relies on individuals to a high degree. Under those circumstances, a collective can be smarter than a person. Break any one of those conditions and the collective becomes unreliable or worse.

3. INNOVATION ENGINE

The vast majority of idea generation methods in existence are a product of Worldview 1 thinking. Worldview 1 adherents think that if you want to have a brilliant idea, you must produce a large number of ideas and the brilliant one will be in there somewhere. Many Worldview 1 methods are derivatives of brainstorming, which is 'quantity leads to quality' thinking in action.

Whereas Worldview 1 is an understanding of reality based on the thinking of Newton and Descartes, first order change and the laws of cause and effect, Worldview 2 represents an altogether different view of reality. Under Worldview 2, problem solving, development, innovation and change come about through small nudges that subtly alter the natural flow of events, such that the needs and interests of all relevant parties are satisfied, quickly and with existing resources. Although causal analysis is still useful for solving simple problems, complex issues are addressed through context manipulation, pattern analysis and constraint removal.

Combining text mining technologies with a unique, collaborative methodology that simultaneously examines user desirability, technical feasibility, and business viability, our innovation engine enables corporate partners to anticipate disruption and provides them with visibility into ideas and technologies that can affect their markets (Figure 1).
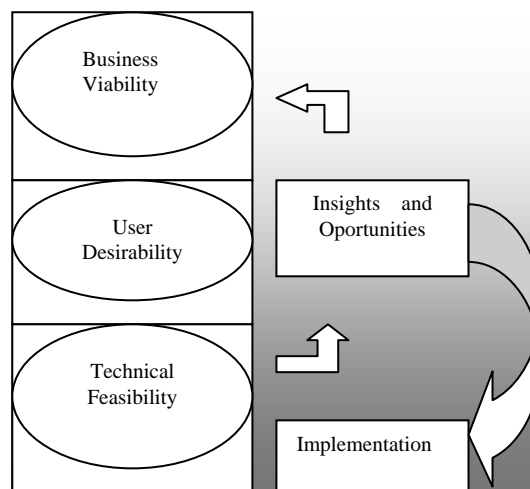

Fig. 1 – Innovation Engine

4. METHODOLOGY

The methodology consists of 5 phases as represented in Figure 2.

4.1. Subject Definition:

Like all good science the success of online collaborative research will be about asking the right questions. Communication and the alignment of goals between researchers are vital ingredients to making a multidisciplinary project fly. But clear detailing of project goals, how each participant will contribute and what he or she will gain also are essential.

4.2. Literature Review:

In this phase, the senior academics select published papers containing the most updated scientific and technological developments relative to the subject under study. This phase is accomplished by Web-based paper abstract posting and voting. All the submitted papers are carefully evaluated by each senior academics under the perspective of opportunity and relevance of the contribution, fostering understanding and development of the subject. The evaluation is registered in the form of votes — "relevant" or "not relevant". If the scientist is not familiar to the paper subject he or she may vote for "abstention". Trying to minimize the effect "Rich Get Richer", partial results are not disclosed during the period of voting.

4.3. Insights and Opportunities:

In this phase all submitted papers are processed and clustered via text mining and the results presented to the scientific advisory board for analysis. The board then selects a set of representative papers to serve as base for speculations, refinements and discussions in the next phase.
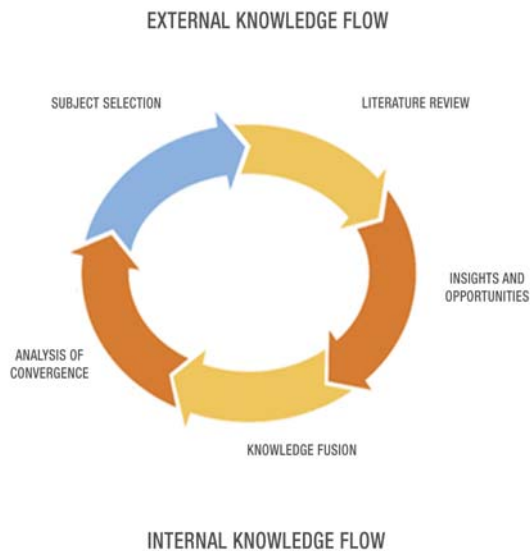
Fig. 2 – Methodology

4.4. Knowledge Fusion:

In this phase the senior academics offer their opinions, ideas and critics, being discussed innovative ideas and solutions posted as comments.

4.5. Analysis of Convergence:

Based on comprehensive text mining analysis, the scientific advisory board may decide to launch a new subject or to ask the senior academics to refine the literature, and/or to continue to discuss the current subject.

Implementation completes the cycle of ideation to drive the concept to its final form.

The possibilities have been evaluated, the prototypes validated and refined. However, innovation isn't something that innovators do — innovation is what customers adopt. It is the market that determines whether something is innovative or not, and whether something is adopted or not. Just because an innovation is adopted doesn't mean that it's a successful business innovation. The challenge for business is not just to get customers and clients to adopt the innovation; it's to get them to pay a premium for the innovation. If the proposed solution is approved by Petrobras, a project team performs detailed design and engineering, chooses manufacturing partners if necessary, and works with the client to perform a timely and successful launch.

5. TEXT MINING IMPLEMENTATION

An alert reader will make connections between seemingly unrelated facts to generate new ideas or hypotheses. However, the growth of published articles means that even the most avid reader cannot hope to keep up with all the reading in a field, let alone adjacent fields. New knowledge is at risk of languishing undiscovered in the literature.
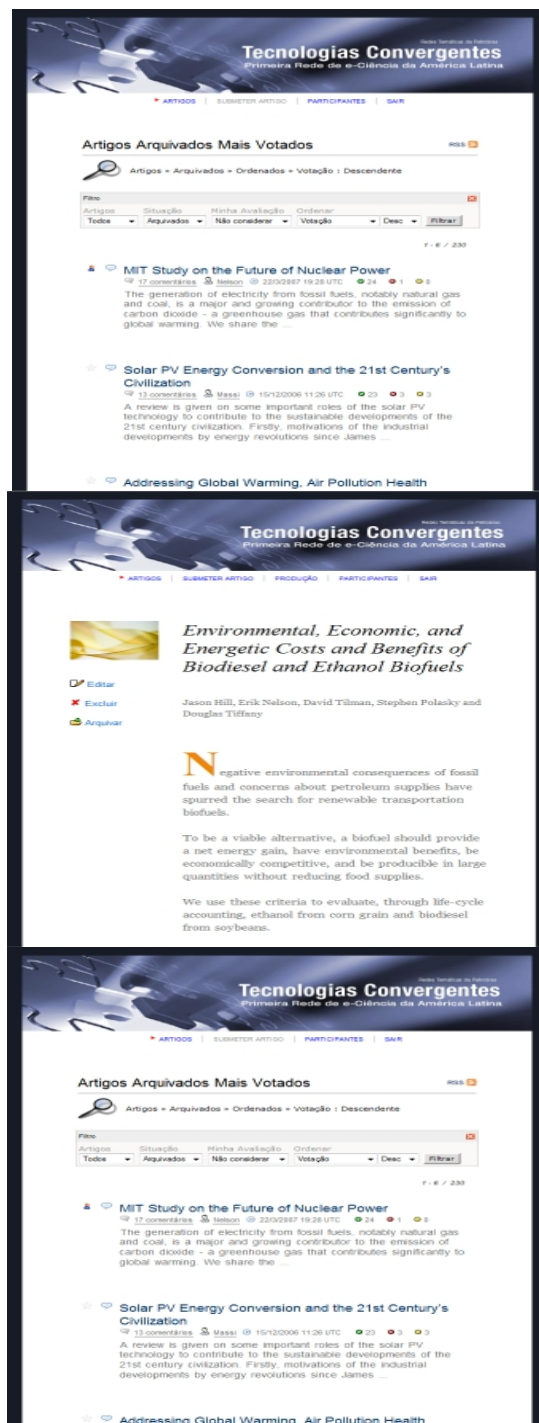
Fig. 3 – Web-based Literature Review

Text mining offers a solution to this problem by supplementing the human reader with automatic systems undeterred by the text explosion. It involves analyzing a large collection of documents to discover previously unknown information. The information might be relationships or patterns that are buried in the document collection and which would otherwise be extremely difficult, if not impossible, to discover. The convergence analysis is based on text mining developments, and used to process the abstracts of the

selected papers to start the discussion and to analyze the comments in the knowledge fusion activity (Camarinha-Matos 2005, Zanasi 2005, Liu 2006, MacLennan and Tang 2005, Chickering and Iver 2004). It is completely integrated to the environment, and the diagram of this process is indicated in Figure 4.

To achieve the tight coupling of Data Mining (DM) techniques in Database Management Systems (DBMS) technology, a number of approaches have been developed in the last years. These approaches include solutions provided by both company and academic research groups.

Toward this objective, the Microsoft (MS) Object Linking and Embedding Database for DM (OLE DB DM) technology provide an industry standard for developing DM algorithms. This technology was first included in the MSSQL 2000 release. The MSSQL Analysis Services (SSAS) has included a DM provider supporting two DM algorithms (one for classification by decision trees and another for clustering) and the DM aggregator feature made possible for developers and researchers to implement new DM algorithms. The MSSQL 2005 version has included more seven algorithms as well as a new way to aggregate new algorithms, using a plug-in approach instead of DM providers. This plug-in approach was adopted to build a new text mining environment, and can support 2 languages: English for the papers and Portuguese for the scientist comments.

Preprocessing includes stop words and stemming processing and the BOW (bag-of-words) toolkit, a library of C functions useful for writing statistical text analysis, language modeling and information retrieval programs.

Clustering is accomplished by the well known K-Means (Zanasi 2005, Liu 2006) algorithm and the search of better clusterization process employs PBM index, a product of three factors, of which the maximization ensures that the partition has a small number of compact clusters with large separation between at least two of them.
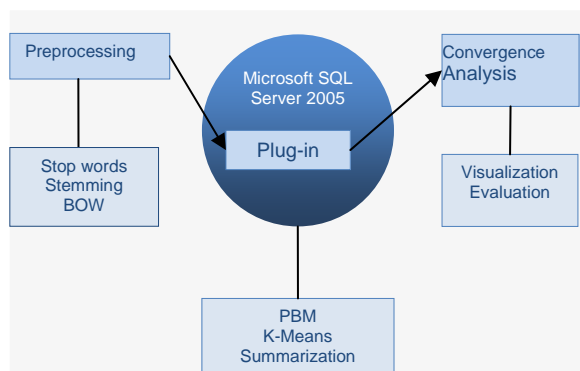

Fig. 4 – Text Mining Implementation

PBM index is calculated by:

$$PBM(K) = \left(\frac{1}{K} \cdot \frac{E_1}{E_k} \cdot D_k \right)^2 \qquad \text{where,} \qquad (1)$$

$$E_k = \sum_{k=1}^{k} E_k$$

is the total intra-cluster distance for K clusters.

The third factor is the biggest inter-centroid distance.

Automatic text summarization is the technique where a text is entered into the computer and a non redundant extract from the original text is returned.

The technique has its roots in the 60's and has been developed during 30 years, but today with the Internet and the Web the technique has become more important. Microsoft Word has since 1997 a summarizer for documents.

In order to generate a summary, we have to identify the most important pieces of information from the document, omitting irrelevant information and minimizing details, and assemble them into a compact coherent report. This however, is easier said than done as it involves some of the main problems of natural language processing. To produce a domain-independent system would require work in natural language understanding, semantic representation, discourse models, world knowledge, and natural language generation. Successes in domain-independent systems are few and limited to identifying key passages and sentences of the document. More successful systems have been produced for limited domain applications.

The used technique for automatic text summarization has been developed for many years: text extraction, which means to extract pieces of an original text on a statistical basis or with heuristic methods and put them together to a new shorter text with as much information as possible preserved. One important task in text extraction is topic identification. There are many methods to perform topic identification. One is word counting at concept level that is more advanced than just simple word counting; another is identification of cue phrases to find the topic (Edmundson 1969).

6. RESULTS

The first Latin America's e-science initiative was tested against reality in their very first iteration. The initial subject was Renewable Energy with Social Responsibility. Using the newly developed online collaboration tool were posted 164 scientific papers, 1,688 evaluations, and 314 comments by 50 senior scientists selected from among the nation's top research institutes. Interestingly, despite the potential for solar and wind electricity generation in Brazil, the multidisciplinary team of academics was not motivated by those alternatives and they started discussing biodiesel production considering microalgae and pollution effects. The innovation engine grouped discussions on 3 clusters: biodiesel, hydrogen and ethanol. At the end of the process 5 research projects were presented to be considered for industrial

development. This was considered as an excellent result, because completely different solutions were proposed.

At this time, as the learning process achieved a good performance, more specific themes are been examined with short time duration cycles.

7. CONCLUSION

The research enterprise increasingly involves multidisciplinary collaborations, sometimes over geographic distance. Technological advances have made these collaborations possible, and the history of past innovations suggests these collaborations are highly desirable. The accumulated body of empiric knowledge and the size of the involved research community provide the important conditions for the foundation of a new scientific organization. This paper introduced a view of an application in this area. As important as collaborative tools, techniques and technologies are, you have to have an environment where there are incentives to collaborate. You have to have an internal economy where there are appropriate rewards and incentives for collaborating, and appropriate disincentives for not collaborating. Interestingly enough, we don't always need to collaborate. You have to have a circumstance where you cannot succeed, cannot get promoted in the organization, cannot get extra organizational resources, unless you are deemed to be somebody who is a creative and cost-effective collaborator across the enterprise. The fact of the matter is that many people have succeeded in organizations not because they're very good collaborators, but they've been successful as leaders on their own, they're not team players at all. But just because you have a common goal doesn't mean you actively and appropriately collaborate with your teammates. It takes shared space to create shared understandings and they realized they couldn't create this value on their own. That's what collaboration has to be about, and it must be recognized explicitly and you must be rewarded explicitly.

ACKNOWLEDGMENTS

REFERENCES

ACM, A new interview with M. Schrage , Available from: http://www.acm.org/ubiquity/interviews/v7i08_schrage.html [accessed 5 July 2008]

ACM, 2006,Service Systems, Service Scientists, SSME, and Innovation, Communications of the ACM, Vol. 49, No.7, 81--84, July .

Camarinha-Matos, L.M. and Afsarmanesh,H., 2005, Collaborative networks: a new scientific discipline, Journal of Intelligent Manufacturing, 16, 439--452.

Chickering M. and Iver, R., 2004, A tutorial for constructing a plug-in algorithm, Microsoft Technical Articles.

Edmundson H. P.,1969, New Methods in Automatic Extraction, Journal of the ACM 16(2) pp. 264--285.

Gorman, M., 2007, Web 2.0: The Sleep of Reason Part II,Available from: http://blogs.britannica.com/blog/main/2007/06/web-20-the-sleep-of-reason-part-ii/ [accessed 5 July 2008]

IAC,2007,Lighting the way: toward a sustainable energy future, InterAcademy Council, October,in: http://www.iafastro.com/index.php?id=498 [accessed 5 July 2008]

Iver R. and Crivat, B., 2004, SQL Server data mining plug-in algorithms, Microsoft Technical Articles.

Lanier, J.,2006, Digital Maoism - The Hazards of the New Online Collectivism, Available from: http://www.edge.org/3rd_culture/lanier06/lanier06_index.html [accessed 5 July 2008]

Lee, P.,2006, Convergent technologies: Future perfect or imperfect?, Available from : http://www.wacc.org.uk/wacc/publications/media_development/2006_2/convergent_technologies_future_perfect_or_imperfect [accessed 5 July 2008]

Liu, B., 2006, Web Data Mining Exploring Hyperlinks, Contents and Usage Data, Springer, December.

MacLennan, J. and Tang, Z.,2005, Data Mining with SQL Server, Wiley Publishing Inc.

Petrobras, 2008, Available from: http://www2.petrobras.com.br/ingles/index.asp [accessed 5 July 2008]

Tang, Z. and MacLennan, J.,2005, Data Mining with SQL Server 2005, Wiley Publishing.

Zanasi, A., 2005, Text Mining and its Applications to Intelligence, CRM and Knowledge Management, WIT Press.

DECISION SUPPORT SYSTEM APPLIED TO COMBINED FREIGHT TRANSPORT

Rigo Nicolas [(a)], Ndiaye Alassane Ballé [(b)]

[(a)]Research Unit ANAST – Naval Architecture and Transport System Analysis, Department ARGENCO, Faculty of Engineering Sciences, University of Liège, Chemin des Chevreuils, 1, 4000 Liège, Belgium
[(b)] QALINC@ - Chair of Innovation for Quality, Logistics and Supply Chain Management, Department BEAMS, Faculty of Engineering Sciences, University of Brussels.

[(a)] nicolas.rigo@ulg.ac.be, [(b)]abndiaye@ulb.ac.be

ABSTRACT

The reasons for choosing a certain way of transporting goods are dependent on numerous factors. Shippers are interested in reliable logistics and low cost, while authorities are in general more concerned with minimizing the environmental impacts, while accident-free transport is in the interest of all parties involved. This paper aims at elaborating an integrated decision aiding tool regarding the selection of combined transport alternatives according to actual road freight transport situations. Based upon a wide range of criteria covering partly the European transport policies, the integrated tool is made up of two aggregation steps. Firstly, aggregation methods integrate the economic, environmental, safety and logistics related performances of transport scenarios into four global scores. Secondly, an outranking method integrates the global performances into a sustainable transport performance index, expressing the overall performance of transport chains regarding the covered European objectives. Finally, the sensitivity of these indexes is tested.

Keywords: Integrated assessment framework, decision support system, combined transport.

1. INTRODUCTION

Across Europe, specific transport chains are researched for which shippers have shown interest in moving their cargo from road to water. It is the task of the research as discussed in this paper to provide an integrated assessment model in order to evaluate the performances of these new transport chains compared to the old ones. Based upon the European transport policies by the year 2010, numerous performance indicators have been elaborated in the fields of logistics, economics, environment and safety. This assessment will result in a final sustainable performance index, aggregating the analyzed impacts and integrating them into a final score.

The performance fields are more or less important according to the actors questioned about the pursued objectives. For commercial parties, the best transport alternative will be the cheapest one that can properly accommodate his logistics related requirements, while for local authorities the safest and least polluting scenario will be the best solution.

Then, the solution will be a compromise between pursued objectives regarding the cost, the logistics, the environment and the safety.

The question that remains to be answered is how well a transport chain that meets these general objectives performs globally. This is the topic of this paper.

The next section proposes a brief overview of the literature review.

Section 3 presents the elaborated integrated decision support tool. The pyramidal structure is presented while the performance indicators and the specific aggregation methods are detailed.

In section 4, the authors review the application of the developed model in a practical case study revolving around the transport of cars, vans as well as trailers from North-Western Europe to Sofia in Bulgaria.

Finally, section 5 presents major conclusions regarding the developed model and its application as well as perspectives of further developments.

2. LITERATURE REVIEW

In the creation of the methodology described in the paper, as well as for the case study that is discussed, a number of publications proved of invaluable use.

For the creation of the integrated assessment framework, the major literature sources are (Roy and Bouyssou 1993) in which the authors detail a wide range of multicriteria decision aiding methods; and (Roy 1985) where the author details the concept of decision process and proposes an overview of multicriteria decision aiding methods as well as some application cases. Moreover, the Directorate-General for Transport presented an inventory of the state of the art for the cost-benefit and multicriteria methods for projects in inland waterways.

After studying these documents, the authors selected the outranking PROMETHEE ("Preference Organisation Method for Enrichment and Evaluation") method because of both efficiency and understanding easiness advantages by the non-mathematical experts.

The analysis of multicriteria decision aiding method applications, highlighted in (Azibi and Vanderpooten 1997, Cescotto, Roubens, Rigo, Gao, Wang, Zhang, Lourenco, Zhou, Xiang and Ferreira 2006, Colson and Mbangala 1998, Rigo, Ndiaye, Dreyer, Zomer, Pinon and Tremeac 2007, Schweigert 1995), helped for elaborating the integrated assessment tool.

In (Brans and Mareshal 2005), the PROMETHEE method is described in detail.

In (Colson 2004, De Bruyn 2002, Ndiaye, Glansdorp, Prunieras and Willems 1993 and Roubens 1991), the authors of this paper found useful information to fine tune the integrated tool.

Moreover, in (Tyworth and Zeng 1998), interesting quantifications of logistics related aspects were found. However, due to a weak availability of data regarding the presented case study these developments were not incorporated in the integrated decision aiding tool.

Since the authors try to reach the objectives pursued by the European Commission; it is interesting to glance through numerous directives from the European Parliament and the Council or other European reports highlighting the need for internalizing external transport effects.

All in all, this review led to the elaboration of the integrated assessment tool presented in the next section.

3. METHODOLOGY

3.1. Introduction
The integrated support system developed within this paper aims at evaluating and aggregating a wide range of internal and external transport impacts in the fields of logistics, economy, environment and safety. The pursued objective is to elaborate a decision helping tool able to compare numerous transport alternatives on the basis of a pyramidal assessment structure.

3.2. The pyramidal scheme
The proposed framework is based on three major steps. Firstly, regarding the four performance domains, lists of pertinent criteria are elaborated. So, the impacts of the transport scenarios can be evaluated on the basis of these evaluation axes. Secondly, the information contained in each field is aggregated into an intermediate global score by using three different aggregation methods. Regarding the economic performances, since various independent costs are used, a sum is relevant to calculate the global cost; the logistics related impacts are aggregated on the basis of an average of normalized values while the safety and environmental impacts are integrated by using the PROMETHEE method.

The second aggregation step aims at evaluating the overall performance of transport alternatives. The PROMETHEE method is used to integrate the four intermediate scores related to the logistics, the environment, the economics and the safety into a final index.

Finally, since the second aggregation is based on specific weights allotted to the macro criteria by the decision makers, a sensitivity analysis is realized to test the robustness of the ranking.

The next section is dedicated to the presentation of the PROMETHEE methodology.
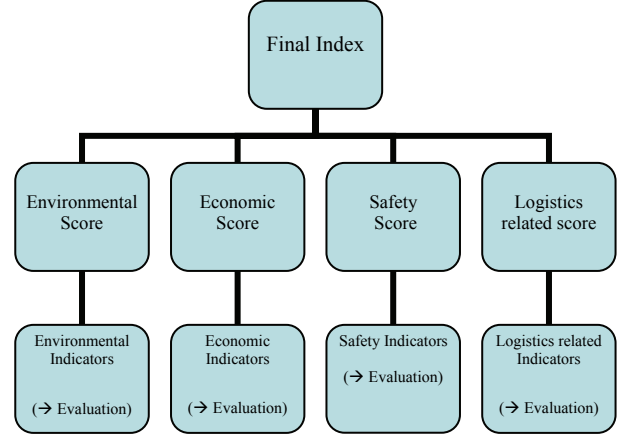
Figure 1: The pyramidal aggregation structure

3.3. The PROMETHEE II complete ranking
Since the PROMETHEE method is used for the aggregation of the environmental and safety impacts as well as for the aggregation of the intermediate scores, the following modeling is recalled.

First, let us consider a set of criteria, $\{g_1(.), g_2(.), g_3(.),...,g_m(.)\}$ and a set of scenarios to compare $A = \{a_1, a_2, a_3,...,a_n\}$. Let us define $g_j(a_i)$ the evaluation of scenario a_i on the axis j.

Let us consider the deviation of impacts of two actions on a criterion:

$$d_j(a,b) = g_j(a) - g_j(b); \quad \forall a,b \in A \quad (1)$$

In order to delete the possible scale effects related to the units of criteria, let us define the following function in the case of a criterion j to maximize,

$$0 < F_j[d_j(a,b)] \le 1; \quad \forall a,b \in A \quad (2)$$

Where:

$$F_j[d_j(a,b)] > 0 \Rightarrow F_j[d_j(b,a)] = 0; \quad \forall a,b \in A \quad (3)$$

If the criterion j has to be minimized, the following relation is considered,

$$0 < F_j[-d_j(a,b)] \le 1; \quad \forall a,b \in A \quad (4)$$

The pair $\{g_j(.); F_j[d_j(a,b)]\}$ is named the generalized criterion associated to the criterion g_j or the preference function related to the criterion g_j.

Various preference functions are available and can be varied to an infinite number of solutions corresponding to the needs of the users. Two examples are described here after:

The usual generalized criterion is defined as follows:

$$F_j[d_j(a,b)] = \begin{cases} 0 & if \quad d_j(a,b) \le 0 \\ 1 & if \quad d_j(a,b) > 0 \end{cases} \quad ; \quad \forall a,b \in A \qquad (5)$$

The U-shape generalized criterion can be defined as follows:

$$F_j[d_j(a,b)] = \begin{cases} 0 & if \quad d_j(a,b) \le q \\ 1 & if \quad d_j(a,b) > q \end{cases} \quad ; \quad \forall a,b \in A \qquad (6)$$

Where q is the preference threshold.
On the basis of these generalized criteria, let us calculate the aggregated preference indices as follow:

$$\begin{cases} \pi(a,b) = \sum_{j=1}^{m} F_j[d_j(a,b)] \times \omega_j \\ \pi(b,a) = \sum_{j=1}^{m} F_j[d_j(b,a)] \times \omega_j \end{cases} \quad ; \quad \forall a,b \in A \qquad (7)$$

Where w_j is the weight allotted to the criterion j. Then, the positive and negative outranking flows are calculated as follows:

$$\phi^+(a) = \frac{1}{n-1} \sum_{x \in A} \pi(a,x) \qquad (8)$$

$$\phi^-(a) = \frac{1}{n-1} \sum_{x \in A} \pi(x,a) \qquad (9)$$

The PROMETHEE II complete ranking is based on the following outranking net flow:

$$\phi(a) = \phi^+(a) - \phi^-(a) \qquad (10)$$

3.4. The criteria and the aggregation methods

3.4.1. The environmental criteria
Regarding the environmental transport impacts, the authors focus on the air pollution. In order to evaluate the effect of transport scenarios on the air pollution, the authors base the approach on the major air pollutants. One of the most important greenhouse gases is CO_2. Another carbon oxide is particularly dangerous: the CO. The emissions of NOx are also considered as well as the Sulphur Oxides SOx. Finally, the particle matters are included in the environmental criteria. All these indicators are expressed in grams per cargo unit.

The environmental indicators are aggregated by using the PROMETHEE method. Therefore, weights have to be allotted to the criteria. Since the societal costs of major air pollutants are well-known, the authors use these values to weight the environmental indicators. Regarding the CO_2 indicator, the related societal cost is about €50,- per ton while the societal costs of the CO, NOx, Sox and PM are respectively about, €100,-/ton, €5000,-/ton, €3000,-/ton and €50000,-/ ton. These figures were found in (NEA, 1999). Actually, these costs are based on averages related to urban and non urban area.

Regarding the preference function, the authors propose to use the usual generalized criterion to conserve all the information.

The actual environmental evaluation of transport alternatives is not detailed in this paper since the authors want to focus on the integrated assessment tool.

As stated before, the global environmental performance is obtained on the basis of the PROMETHEE method, as follows:

$$\begin{aligned} &d_{CO2}(a,b) \ , \ \omega_{CO2} \ , \ F_{CO2}(a,b) \ ; \ \forall a,b \in A \\ &d_{CO}(a,b) \ , \ \omega_{CO} \ , \ F_{CO} \ ; \ \forall a,b \in A \\ &d_{NOx}(a,b) \ , \ \omega_{NOx} \ , \ F_{NOx} \ ; \ \forall a,b \in A \qquad (11) \\ &d_{SOx}(a,b) \ , \ \omega_{SOx} \ , \ F_{SOx} \ ; \ \forall a,b \in A \\ &d_{PM}(a,b) \ , \ \omega_{PM} \ , \ F_{PM} \ ; \ \forall a,b \in A \end{aligned}$$

So that we can obtain $\phi_{Env}(a)$; $\forall a \in A$; the environmental performance of alternative a.

3.4.2. The economic criteria
The economic indicators are the cost of the actual transport legs. Since the authors consider combined transport chains, the possible additional cost of transhipment and intermediate storage are considered. For transhipment and non-waterborne transport, commercial cost as provided by contacted market parties are used, while for waterborne transport a detailed cost breakdown, including but not limited to depreciation interest, fuel cost, crew cost, maintenance and repair and overhead is created in order to be able to establish the interrelation between economic and environmental/safety performance of various devices added to the ship. All these details are not included in this paper and can be found in (Hekkenberg, 2006).

For the calculation of economic performances, we just have to sum the three no-redundant costs.

3.4.3. The safety criteria
Determination of proper safety indicators proved a challenge. Two principally different types of safety can be distinguished external safety and internal safety. External safety is highly dependent on the specifics of the transport route and the population centres surrounding it, which is data that was too time-consuming to process in great detail within this paper. Apart from this, external safety performance of inland waterway transport is generally accepted to be vastly

superior to road transport. As a result, only internal safety was elected as a measure of safety.

The Dutch AVV (Adviesdienst Verkeer en Vervoer, Ministry of Transport) accident database reveals that on Dutch inland waterways, over 13 years, 12 fatalities occurred as a result of accidents with inland ships while 5909 accidents happened. Based on data coming from DG TREN, on average over the last 15 years, an annual transport performance of 37 billion TKM is achieved by inland waterway transport in the Netherlands. Then, the above mentioned figures lead to the following probabilities that the authors propose to use for the safety evaluation in the case study:

- Fatalities per TKM: $2.5*10^{-11}$
- Accidents per TKM: $1.23*10^{-8}$

Regarding the road safety aspects, the next table is elaborated in order to find equivalent probabilities.

Table 1: Safety figures for heavy goods vehicles per country

Country	Number of accidents	Fatalities in HGV	Freight transport (10^6 TKM)
Belgium	47619	28	20390
France	105470	132	169740
Austria	43175	12	12660
The Netherlands	33538	11	30260
Finland	6196	5	28070
Portugal	42114	18	14920
Sweden	16947	13	31840
United Kingdom	228534	63	150920
Ireland	6625	3	10730
Italy	237812	249	151970
Luxembourg	769	3	580
Denmark	7121	5	11060
Greece	16809	23	20000
Germany	362054	-	-
Spain	98433	172	129510

Based on the precedent table, since detailed safety data don't exist for some of the countries involved in the case study, the authors calculate a global EU average.

Table 2: Accident and death probabilities for heavy goods vehicles

COUNTRY	Road (2004)	
	Accidents / 10^6 TKM	Deaths / 10^6 TKM
Belgium	2,34	0,00137
France	0,62	0,00078
Austria	3,41	0,00095
The Netherlands	1,11	0,00036
Finland	0,22	0,00018
Portugal	2,82	0,00121
Sweden	0,53	0,00041
United Kingdom	1,51	0,00042
Ireland	0,62	0,00028
Italy	1,56	0,00164
Luxembourg	1,33	0,00517
Denmark	0,64	0,00045
Greece	0,84	0,00115
Spain	0,76	0,00133
Average	1,307857143	0,001121429
Per TKM	1,30786E-06	1,12143E-09

The safety performances are aggregated into a global safety score on the basis of the PROMETHEE method:

$$d_{Damage}(a,b) \ , \ \omega_{Damage} \ , \ F_{Damage} \ ; \ \forall a,b \in A$$
$$d_{Death}(a,b) \ , \ \omega_{Death} \ , \ F_{Death} \ ; \ \forall a,b \in A \quad (12)$$

On the basis of these factors, we can calculate the following outranking net flows representing the safety performance of the studied scenarios: $\phi_{Saf}(a) \ ; \ \forall a \in A$.

3.4.4. The logistics related criteria

Perhaps the largest challenge in the provision of indicators was for those indicators related to logistics. This had two reasons: First, the market parties involved were unable to put a price tag or other numerical value on the more obvious indicators such as time and speed, nor were they able to provide sufficient background information to allow any quantitative assessment. Second, since the study case was set up in cooperation with these market parties, it was known beforehand that all transport scenarios scored a sufficient mark for logistics related performance.

As a result, the choice was made to set up an own qualitative assessment framework. The ordinal scales are given for each logistics related criterion by a group of experts, in close cooperation with the authors and the persons in charge of the logistics within the CREATING Project. Since the authors use the PROMETHEE method to calculate the final index, combinations of quantitative and qualitative approaches are allowed thanks to the preference functions. However, the authors want to recall that the proposed ordinal scales can be different according to others points of view so that the established results can be debatable and require to take precautions before actual implementation. Finally, it is useful to specify that the results are not really modified when considering that logistics related assessment.

In this, six indicators are defined. These are as follows:

- Logistics related character of the freight,
- Number of border crossings,
- Geographical conditions and traffic density,
- Volume of the transported cargo,
- Number of transshipment and cargo vulnerability,
- Flexibility.

All the following ordinal scales are defined to maximize the criteria.

The logistics related character of the freight means the double consideration of the type of cargo and the "single or long term" specificities of transport scenarios. The type of cargo deals with the possible specific transport conditions required by certain types of goods. Dangerous, cold, heavy, extremely large, fragile goods may require special care, specific rules and regulations making the transport activities more complex from a logistics point of view. A single freight means that the transport task is realized one time occasionally so that no specific interactions exist because the transport task is unique. In the other hand, long term freight means that transport activities are continuously done along a same route. In this case, some problems affecting one transport task can impact on the global continuous service leading to more complex logistics management.

The following table presents the associated qualitative value which will be normalized during the aggregation.

Table 3: The logistics related character of the freight

	Normal Cargo	Special cargo
Single freight	3	2
Long term freight	2	1

At the custom borders, the vehicles must stop and wait for custom control, which sometimes can cause considerable time wasting decreasing the time liability level of the transport service. Crossing a custom border definitely increases the logistics constraints. Table 4 has been elaborated according to the presented case study involving three border crossings at maximum.

Table 4: Border crossings

	Number of border crossings			
Absolute	0	1	2	3
Quantitative value	6	3	2	1

The geographical conditions state for the climatic and topographic characteristics of transport routes. Indeed, seasonal effects such as ice, small water depth can influence the difficulty to fulfill the transport activities and are so of key importance for the evaluation of the logistics related performances. The traffic density of the transport route has great impact on the real transportation time, in other words, on the liability of the transportation. The high traffic density on the transport networks increases the transportation time, the risk of traffic accidents and operational costs,

etc. To avoid all of these effects is impossible, but being prepared to solve them is compulsory for the forwarding companies. Therefore the lower road density of the transport route performs better from a logistics point of view. If inland waterway navigation is involved into the transport chain, the nautical parameters of the shipping route have similar influence onto the transport efficiency, like the traffic density in case of road transport. Putting all these parameters together, the complex subjective evaluation of the given transport task and the proposed solution can be determined as per the following table:

Table 5: Geographical conditions and traffic density

Climatic and geographical conditions	Traffic density	
	Low	High
Easy	4	3
Difficult	3	1

In order to apply correctly this table according to the specificities of the transport route observed in the case study, further analysis has to be done.

In order to quantify the number of kilometers of a transport route with a low traffic density level, we propose to associate it with the "motorways" and "high quality IWW". Indeed, on such sections, we can expect a quite good traffic flow due to the satisfactory circulation conditions. The same reasoning is used for the sections with a high traffic density level. High traffic density networks could be associated to "main roads" and "low quality IWW".

Then we can evaluate the performance of transport chains regarding this indicator.

First let us define the following variables:

- Distance on motorway (km) : Mo
- Distance on Main road (km): Mr
- Distance on road (km): R = Mo + Mr
- Share of Main road from total road (%): SMr = $100*(Mr/R)$
- Distance on high quality IWW (km): Hq
- Distance on low quality IWW (km): Lq
- Distance on IWW (km): I=Hq+Lq
- Share of low quality IWW from total IWW (%): SLq = $100*(Lq/I)$
- Share of IWW from total (%): SI = $100*(I/[I+R])$

Let us define x_l and x_h the evaluations related to table 5: low traffic density and high traffic density. The impact of a scenario on this criterion can be calculated as follows:

$$\frac{\begin{aligned}&[(x_h*(100-SMr)*(100-SI)/10000)\\&+(x_h*(100-SLq)*SI/10000)]\\&+\\&[(x_l*SMr*(100-SI)/10000)\\&+(x_l*(100-SLq)*SI/10000)]\end{aligned}}{4}=LOG_{GC}(a) \quad (13)$$

Where $LOG_{GC}(a)$ represents the normalized impact of a scenario a on this indicator.

The volume of the transported cargo volume is important for the logistics, however it can not be independent from the transportation method itself, since each transport mode "thinks" in different optimal cargo volume.

Table 6: Scale for the cargo volume

Transport mode	The volume of the transported cargo		
	< 100 tons	100 – 1000 tons	> 1000 tons
Road	3*3=9	2*2=4	1*1=1
Inland Waterway	1*3=3	2*2=4	3*1=3

In table 6 the first multiplier of the product considers the transportation mode, while the second one considers the volume of transported cargo. A factor 1 is given to amounts of cargo larger than 1000 tons because, in the absolute, we think it is more difficult to transport 1000 tons than 100 tons of cargo. However, it is interesting to associate another parameter reflecting how the transport modes "think" about these specific amounts of cargo. For example, it is very complex for the road to transport 1000 tons of cargo, so we associate a factor one.

In case of intermodal transport, the "value" of this indicator can be determined as the average of the different transport modes based on their relative lengths inside the total transport route as presented in the following calculations.

Let us define V_r and V_{IWW}, the evaluation of the volume of the transported cargo for the road and the IWW.

Then, we can calculate the impact of the transport chains on this indicator as follows:

$$\frac{(V_R * (100 - SI)) + (V_{IWW} * SI)}{100} \Big/ 9 = LOG_{VC}(a) \qquad (14)$$

Where $LOG_{VC}(a)$ is the normalized impact of a transport scenario a on the transported cargo volume.

The number of transshipment needed in the transport process is an important indicator characterizing the difficulty of the given transport task. Less cargo transshipments mean not only less difficulty but also a smaller risk of cargo damage as well. This ensures higher transport liability level. But the number of transshipments is still not enough to evaluate the difficulty of the entire transport process. We must consider other factors, like the type of the cargo (general, bulk – dry, liquid, dangerous goods) leading to considerations of environmental risks during transshipment, needs for transitional cargo collection or storage, cargo handling specialties, etc. Then, the authors propose the following evaluation table:

Table 7: The number of transshipment

Trans-shipments	Transitional storage and collection							
	No				Yes			
	B	G	C	S	B	G	C	S
0	13	13	13	12	12	11	11	10
1	11	11	11	10	10	9	9	7
2	10	10	10	9	7	5	5	1

B states for bulk, G for general cargo, C for container and S for special type of cargo like RORO, dangerous goods, etc.

From a logistics point of view the flexibility means, that how can a given transport solution react and adjust its output to the rapidly changing demands of the market conditions concerning the delivery time and the cargo volume. Since the economic situations and business conditions are not static in time, this indicator is very important to measure the long term market sensitivity of a given transport solution.

Table 8: The flexibility

Changes in demands		Flexibility			
Cargo volume	Delivery time	easy	fair	difficult	extremely difficult
		4	3	2	1
Growing	steady				
Growing	decreasing				
Steady	decreasing				
Decreasing	steady				
Decreasing	decreasing				

The given transport solution should be analysed and must receive a ranking number. The flexibility of the given transport solution can be considered as the average of the obtained scores.

The aggregation of the logistics related indicators is based on an average of the normalized values as follows:

Let us define:

$$\begin{aligned} &Log_{CF}(a); \forall a \in A \\ &Log_{BC}(a); \forall a \in A \\ &Log_{GC}(a); \forall a \in A \\ &Log_{VC}(a); \forall a \in A \\ &Log_{NT}(a); \forall a \in A \\ &Log_{FL}(a); \forall a \in A \end{aligned} \qquad (15)$$

Where CF states for the logistics related character of the freight, BC the number of border crossing, GC the geographical conditions, VC the volume of the transported cargo, NT the number of transshipment and FL the flexibility. So, the average of these evaluations gives the global performance regarding the logistics as follows:

$LOG\ (a) =$

$Log_{CF}(a)+$

$Log_{BC}(a)+$

$Log_{GC}(a)+$ (16)

$Log_{VC}(a)+$

$Log_{NT}(a)+$

$\dfrac{Log_{FL}(a)}{6}$; $\forall\, a \in A$

3.4.5. Towards the final index

According to the precedent, the PROMETHEE method is used in order to aggregate the four obtained scores into a final global index. Since the macro criteria contain independent information, the authors don't confront redundancy problems. Indeed, the definition of the indicators is done in order to avoid double counting during the calculations.

This aggregation is based on specific allotted weights. So, it is necessary to perform a sensitivity analysis by moving weights while simultaneously analyzing the possible impacts on the indexes. This is the robustness analysis performed in the case study.

4. STUDY CASE

4.1.1. Transport alternatives

The case study revolves around the transport of trailers with goods and/or new cars and vans between Frankfurt am Main in Germany and Sofia in Bulgaria. This transport is currently realized by means of trucks along the road network from North-Western to South-Eastern Europe. The goal is to study the possibility to shift that freight from road to inland ships.

When looking at the geographical and topographical characteristics, it seems that a pertinent intermodal route can be considered: a pre-haulage from Frankfurt am Main to Passau, an unaccompanied "point-to-point" inland navigation using RORO vessels from Passau to Vidin and an end-haulage from Vidin to Sofia. The distance over the pre- and end-haulage is approximately equal to 661 km, the distance over the waterborne section is equal to 1436 km.

Based upon this route, numerous alternatives are considered reviewing various types of ships and onboard technologies.

The first transport scenario is the present situation highlighting the road transport (Ref. Case: Road). The goods are transported by EURO III trucks from Frankfurt to Sofia.

When considering intermodal alternatives, concerning the pre and end-haulage, the same trucks are used while, for the waterborne section, different ships are considered.

The first intermodal alternative proposes to use a small RORO vessel, carrying 29 cargo units with a CCNR II approved engine, sailing at a speed of 16km/h (Ref. Case: SRRV_16).

The reduction of the speed should reduce the emissions by decreasing the fuel consumption. So, the Case 3 proposes to consider the same ship but slowing down to 14km/h (Ref. Case: SRRV_14).

As explained before and without entering in the details, other ships are designed by Project partners: a large RORO and a very large RORO vessel. So, Case 4 proposes to use a large RORO vessel, carrying 63 cargo units and sailing at a speed of 16km/h, equipped with a CCNR II approved engine (Ref. Case: LRRV_16_63).

As Case 3, Case 5 considers Case 4 with a speed reduced to 14km/h (Ref. Case: LRRV_14_63).

In the case of a large RORO vessel, it is proposed to consider the improvement of the hullform which results in a reduction by 5% of the resistance, saving fuel consumption. It is Case 6 considering a speed of 14km/h (Ref. Case: LRRV_14_63_5). This speed is conserved for all the following cases because of the significant fuel consumption reduction.

Case 7 consists in the use of a very large RORO vessel carrying 73 cargo units (Ref. Case: VLRRV_14).

Case 8 reviews Case 6 by increasing the loading capacity to 89 cargo units in optimal conditions (Ref. Case: LRRV_14_89).

Case 7 is reviewed in Case 9 by improving the loading conditions to obtain an optimal loading capacity of 104 cargo units (Ref. Case: VLRRV_14_104).

Case 10 reviews Case 5 by adding an SCR catalyst, a PM-filter and the use of a low sulphur fuel (Ref. Case: LRRV14SP).

4.1.2. Application of the decision support system tool

Firstly, regarding the environmental aspects, the evaluation has been performed in a parallel model (Hekkenberg, 2006), taking into account all the architectural and hydrodynamic properties of the ships. The following table comes from the decision support tool and presents the results of the calculations in gram per cargo unit. In this case, the authors consider the semi-trailer.

Evaluation					
	CO2	CO	NOx	SOx	PM
Road	1480239.6	1435.65	10134.0	0.0	168.9
SRRV_16	2151860.969	7641.341379	17668.24138	1708.212759	522.8413793
SRRV_14	1643441.809	5352.49532	13238.21675	1155.936355	375.1738916
LRRV_16_63	1446644.408	4466.535484	11523.45577	942.1628199	318.0151925
LRRV_14_63	1182246.975	3276.247883	9219.673322	654.9579409	241.2224441
LRRV_14_63_5	1154540.975	3151.518579	8978.261765	624.8619667	233.1753922
VLRRV_14	1125448.786	3020.548806	8724.771883	593.260228	224.7257294
LRRV_14_89	1010128.56	2501.390075	7719.948532	467.9922503	191.2316177
VLRRV_14_104	966759.6712	2306.1482	7342.061033	420.8822754	178.6353678
LRRV14SP	1138165.451	3276.247883	5668.910613	32.46485328	129.2795284

Figure 2: Environmental evaluation table

When running the model, we obtain the following ranking classifying the transport scenarios according to their environmental performances and the allotted weights.

Figure 3: The environmental ranking

N°	Scenario	Final score
1	[LRRV14SP]	0.9858603229196523
2	[Road]	0.6929397152956912
3	[VLRRV_14_104]	0.5754275341549633
4	[LRRV_14_89]	0.35320531193274096
5	[VLRRV_14]	0.1309830897105188
6	[LRRV_14_63_5]	-0.09143020922900552
7	[LRRV_14_63]	-0.3138435081685297
8	[LRRV_16_63]	-0.5553644788382535
9	[SRRV_14]	-0.777777777777777
10	[SRRV_16]	-1.0

Secondly, regarding the logistics, the following properties are used to perform the calculations according to the criteria explained before.

Figure 4: Properties of the road scenario

Figure 5: Properties of the intermodal route

When using the model, the following evaluation table is obtained and summarizes the logistics related impacts. Just to recall, the figures are normalized so that the most the values is close to one the best the scenario performs.

Figure 6: The logistic evaluation table

Then, the model provides the following ranking:

Figure 7: The logistics related ranking

N°	Scenario	Final score
1	[Road]	0.5891921296296295
2	[SRRV_16, SRRV_14, LRRV_16_63, LRRV_1…	0.587124581994302

Now, if we look at the safety performances, table 9 can be elaborated on the basis of 23T cargo units.

Table 9: The safety evaluation

	TKM road	TKM water	A(total)	D(total)
Road	38847	0	0,050806	4,36E-05
SRRV_16	15203	33028	0,02029	1,79E-05
SRRV_14	15203	33028	0,02029	1,79E-05
LRRV_16_63	15203	33028	0,02029	1,79E-05
LRRV_14_63	15203	33028	0,02029	1,79E-05
LRRV_14_63_5	15203	33028	0,02029	1,79E-05
VLRRV_14	15203	33028	0,02029	1,79E-05
LRRV_14_89	15203	33028	0,02029	1,79E-05
VLRRV_14_104	15203	33028	0,02029	1,79E-05
LRRV14SP	15203	33028	0,02029	1,79E-05

Once again, the model aggregates these impacts into one safety performance score and provides the following ranking.

Figure 8: The safety ranking

N°	Scenario	Final score
1	[SRRV_16, SRRV_14, LRRV_16_63, LRRV_1…	0.09090909090909091
2	[Road]	-1.0

Finally, the overall performance can be estimated on the basis of the following evaluation table summarizing the intermediate scores.

	Environment	Logistic	Safety	Economic
Road	0.6929397152956912	0.5891921296296295	-1.0	1681.98
SRRV_16	-1.0	0.587124581994302	0.09090909090909091	1686.04
SRRV_14	-0.7777777777777777	0.587124581994302	0.09090909090909091	1599.95
LRRV_16_63	-0.5553644788382535	0.587124581994302	0.09090909090909091	1311.25
LRRV_14_63	-0.3138435081685297	0.587124581994302	0.09090909090909091	1266.47
LRRV_14_63_5	-0.09143020922900552	0.587124581994302	0.09090909090909091	1261.78
VLRRV_14	0.1309830897105188	0.587124581994302	0.09090909090909091	1184.35
LRRV_14_89	0.35320531193274096	0.587124581994302	0.09090909090909091	1010.13
VLRRV_14_104	0.5754275341549633	0.587124581994302	0.09090909090909091	949.82
LRRV14SP	0.9858603229196523	0.587124581994302	0.09090909090909091	1274.27

Figure 9: The final evaluation table

Then the model performs the underlying calculations and provides the final ranking. This ranking highlights the best scenarios regarding the used indicators and makes it possible to compare the new alternatives with the present situation.

N°	Scenario	Final score
1	[VLRRV_14_104]	0.3888888888888889
2	[LRRV_14_89]	0.2777777777777778
3	[LRRV14SP]	0.2222222222222224
4	[VLRRV_14]	0.16666666666666666
5	[LRRV_14_63_5]	0.05555555555555558
6	[Road]	0.0
7	[LRRV_14_63]	-0.05555555555555558
8	[LRRV_16_63]	-0.2222222222222224
9	[SRRV_14]	-0.3333333333333333
10	[SRRV_16]	-0.5

Figure 10: The final ranking

In such a decision aiding approach, it is very important to fine tune the parameters correctly. Indeed, the weights allotted to the criteria and the preference functions can have an impact on the final ranking and lead to different recommendations.

A difficult problem to solve is to find common weights satisfying each decision maker.

The comparison of the rankings obtained on the basis of different weightings is a way to test the sensitivity of the recommendations. On figure 11, the sensitivity tool is illustrated. The top part highlights the ranking based on new weights shown on the bottom part.
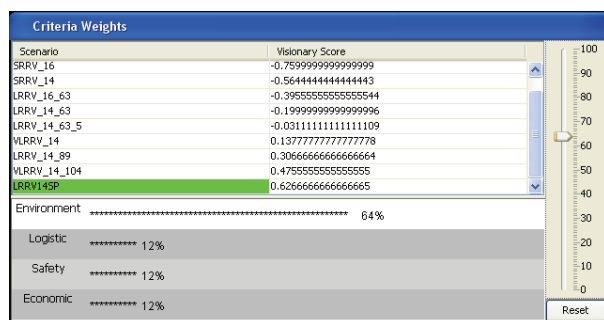

Figure 11: Sensitivity tool

5. CONCLUSION

This paper presented the development of an integrated assessment tool for the evaluation of combined transport chains.

First, the authors developed the integrated assessment framework, highlighting the pertinent indicators and their aggregation in view of the calculation of a final global score expressing the 'overall' performance of the studied transport scenarios.

This new approach has been implemented in a tool which the authors used to demonstrate the applicability of the method to a practical case study.

The authors obtained the four indexes and the final global score providing the final ranking of transport alternatives compared to the present situation.

The authors discussed this ranking by highlighting the importance of the allotted weights and presented a brief example of robustness analysis.

The methodology developed within this paper can be a powerful decision support aid for shippers and shipowners, allowing them to gain better insight into the performance they may expect from their operations.

The performed assessment methodology can be applied to new ship and transport concepts compared to non-optimized concepts in the fields of economy, environment, safety and logistics.

Possible further developments can be identified. Indeed, the internalization of a wider range of transport impacts could be considered: the noise, the congestion, the impacts on the nature and the landscape, the soil and water pollution. The internalization of these external impacts would improve the recommendations.

Finally, it could be interesting to investigate a development of the model in view of the elaboration of socio-cost-benefit ratios providing the economic impacts for the society and the involved actors. Then, we could propose an actual investment helping tool to the decision makers.

REFERENCES

Azibi, R. & Vanderpooten, D. (1997), *Elaboration de critères agrégeant des conséquences dispersées : deux exemples concrets*, in The EURO working group on multiple criteria decision aiding (1999), A-MCD-A ; Multi Criteria Decision Aiding, Colorni A., Paruccini M., Roy B., pp 13-30.

Blaauw, H.G., de Wilde, H., Hekkenberg, R.G.: *Environmental impact in relation to economics of inland navigation*, PIANC conference 2006, Estoril, 2006.

Blaauw, H.G., Thill, C., Zigic, B., Radojcic D., Hekkenberg, R.G.: *The 4 cases of CREATING*, Inland Waterways & Coastal Ships II conference, Royal Institution of Naval Architects, 2006.

Brans, J.P., & Mareschal, B.: PROMETHEE Methods in J.Figuera, S. Greco and M. Ehrgot (editors): */Multiple Criteria Decision Analysis: State of the Art Surveys/*, pages 163-189, Springer-Verlag , Berlin, 2005.

Cescotto, S., Roubens, M., Rigo, N., Gao, S., Wang, X. , Zhang, A., Lourenco, N. , Zhou, J., Xiang, X., Ferreira, JPL. 'Management of Water Pollutants Based on Multi-Criteria Analysis and Fuzzy Logics', *Computational Methods in Engineering and Science EPMESC X, Augustus, 2006, Sanya, Hainan,China,* Tsinghua University Press & Springer-Verlag.

Colson, G. (2004), *Notes du cours d'Aide à la décision – ULg - Faculté d'Economie de Gestion et de Sciences Sociales*, Liège, 2004.

Colson, G. (2004), *Notes du cours des Méthodes d'Aide à la décision – HEC - Liège*, Liège, 2004.

Colson, G. and Mbangala M. (1998), *Une analyse multicritère des performances d'entreprises publiques de chemins de fer africains*, n° 9806, coll. « G.E.M.M.E ».

De Bruyn, C. (2002), *Systems Analysis in Management – ULg - Faculté d'Economie de Gestion et de Sciences Sociales*, Liège, 2002.

Directorate-General for Energy and Transport, 2006, *White Paper, European transport policy for 2010: time to decide*, Brussels, EU Commission

Directorate-General for Energy and Transport, 1994, *Inventory of the state of the art for cost benefit and multicriteria methods for projects in inland waterways,* Brussels, EU Commission

Directorate-General for Energy and Transport, 2007, *Preparation of an impact assessment on the internalisation of external costs*, Brussels, EU Commission

Directorate-General for Energy and Transport, 2006, *EU Energy and Transport in figures, Statistical pocketbook 2006*, Brussels, EU Commission

Haenen, M., Becker, J., Burgess, A., ,CREATING, Deliverable WP2, Final Report, Innovative logistics concepts for Integrated Transport Chains', Brussels, November 2006.

Ndiaye, A.B., Rigo, N., Lundoluka, F., Hekkenberg, R.H., Hadhazi, D., Simongati, Gy., Hargitai, Cs.: 'CREATING, Deliverable WP4, Final Report, Performance Assessment', pp. 46 – 52, Delft, August 2007.

Ndiaye, A.B., Glansdorp, C. Prunieras, J., Willems, C., *A particular approach for the assessment of River Information Services (RIS) in Europe*, Bulletin – International Navigation Association ISSN 1680-1059, n°114, pp79-92, Bruxelles, 1993.

Papageorgiou, M. (2006), *Dynamic Traffic Flow Modelling and Control – Technical University of Crete¸* Chania, 2006.

Pomerol, J.C. et Barba-Romero, S. (1993), *Choix multicritères dans l'entreprise*, Hermès, Paris.

Rigo, N., Ndiaye, A.B., Dreyer, B., Zomer, G., Pinon, H., Tremeac, Y. (2007), 'Elaboration of an integrated assessment framework for the global evaluation of E-maritime operations', *Proceedings of the BIVEC-GIBET Transport Research Day 2007, pp 281-293*.

Rigo, N., Halet, F., Ndiaye, A.B. (2007), 'LODA model, Logistic Optimization and Decision Aiding', Liège.

Rohacs, J., Hadhazi, D., Simongati, Gy., Hargitai, Cs. (June 2005), *Methods for Evaluation of Sustainable Transport Performance*, Budapest.

Roubens, M. (1991), *Méthodes Mathématiques en Economie et en Gestion – ULg - Faculté des Sciences Mathématiques*, Liège, 1991.

Roy, B. et Bouyssou, D. (1993), *Aide multicritère à la décision : Méthodes et cas*, Economica, Paris, coll. « Gestion ».

Roy, B. (1985), *Méthodologie multicritère d'aide à la décision*, Economica, Paris, coll. « Gestion ».

Schweigert, D. (1995), *Proceedings of the 5th Workshop of the DGOR-Working Group – Multicriteria Optimization and Decision Theory*, FB Mathematik Universität, Kaiserslautern, 1995.

Van Mourik, D. 1995, 'Social, ecological and economic criteria and cost-benefit analysis', *Methods of Multicriteria Decision Theory, pp. 133-148.*

Vermeulen, J.P.L., Boon, B.H., van Essen, H.P., Dings, J.M.W., Bruinsma, F.R., Koetse, M.J. (2004), *The price of transport – Overview of the social costs of transport*, Delft, December 2004.

A SIMULATION-BASED DSS FOR FIELD SERVICE DELIVERY OPTIMIZATION

Mario Rapaccini[a], Filippo Visintin[b], Alessandro Sistemi[c]

[a,b,c] University of Florence, Dipartimento di Energetica "Sergio Stecco", Via Cesare Lombroso 6/17 – 50134, Firenze

[a] mario.rapaccini@unifi.it, [b] filippo.visintin@unifi.it, [c] alessandro.sistemi@unifi.it

ABSTRACT

This paper aims at presenting the preliminary results of a research targeted at developing a Decision Support System (DSS) for the Field Service Delivery System (FSDS) design. The paper is organized as follow: firstly, we illustrate what a FSDS is; secondly, we identify the variables to consider in order to design a FSDS and the relationships between them; thirdly, we describe how a DSS supporting the FSDS design should be developed; and finally, we show the result of of a pilot experiment in which a DSS has been developed and applied to a real case study.

Keywords: field service delivery system (FSDS), decision support system (DSS), discrete event simulation, after sales service

1. INTRODUCTION

This paper aims at describing a tool that can support the management of a Field Service Delivery System (FSDS), that is, the company's function devoted to deliver services – such as the product's installation and maintenance – directly at customer's site.

The rationale of the paper lies in the fact that, despite the increasing interest raised by field service in many industries and the large number of applications supporting the field service operations management, there is a lack in models/devices supporting the *design* of the FSDS (Visintin 2007).

A great deal of information is usually available in the Enterprise Information Systems, but generally there are not tools allowing service managers to fully and properly utilize these information, in order to understand the effects that different managerial policies can have on the overall system performance (Agnihotri and Karmarkar 1992).

Moreover, the FSDS design activity is a very complex task. It requires, in fact, to forecast *when* and *where* the service requests will arise, to figure out *what* skills and parts will be required in order to fulfill the service requests, and to decide *what* criteria should be followed to dispatch the technicians.

In addition to that, service managers are asked to achieve increasingly high performances, both in terms of customer satisfaction and costs.

Because of the complexity and the uncertainty characterizing the FSDS design there is a strong need of simulation-based tools supporting service managers.

It is in fact tremendously expensive and time consuming, to modify the FSDS configuration ex-post.

The aim of this paper is thus to discuss how a Decision Support System (DSS) for designing the FSDS could be developed and to illustrate a first example of a DSS.

The paper is therefore organized as follows: firstly, we define and describe a generic field service delivery process; secondly, we identify all the variables that service managers should take into account to design the FSDS; thirdly, we assess if and where the information relevant to these variables can be found in the most common Enterprise Information Systems; fourthly, we describe how the discrete event simulation can be successfully used to create such a DSS; and finally, we show the result of a pilot experiment in which a DSS has been developed and applied to a real case study.

2. THE FIELD SERVICE DELIVERY SYSTEM

A FSDS is made of a set of technicians, each mastering a given set of skills and covering a given geographic area, that are remotely dispatched at the customers' sites to fix the customer's problems upon request (Blumberg 1991; Visintin 2007).

A typical field service process can be subdivided in three main activities:

1. help desk;
2. dispatching;
3. service delivery.

Figure 1 describes a typical field service delivery process.

The Help desk is the company interface with the customer, so it is devoted to receive the incoming service request, to identify the customer and to offer (when it is possible) a remote assistance, that is, to avoid the service delivery on-field.

If the call avoidance does not succeed, then comes the need of selecting and dispatching a field engineer to the customer (Agnihotri and Mishra 2004).

Finally, the selected technician gathers all the needed technical data and actually visits the customer.

In designing and managing this process, service managers need to take into account several variables

116

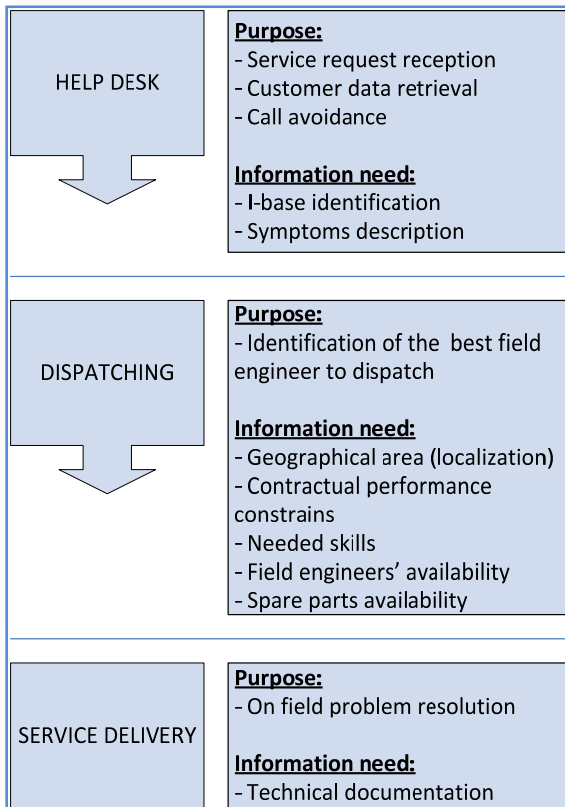(Agnihotri and Chakravarty 2005), that we present in the next paragraph.

Figure 1: Field service delivery process description

3. FSDS DESIGN PROCESS VARIABLES

The design of a FSDS is a decision-making process characterized by several independent, dependent and control variables (Agnihotri, Narasimhan and Pirkul, 1990).

The control variables are those that managers can manipulate in order to obtain, in a given context, a desired outcome. They concern:

- the system *Capacity*, that is the overall number of technicians;

- the technicians' *Dispatching* policy, that is, how a service request is assigned to a field engineer;
- the technicians' *Scheduling* policy, that is, how the amount of service requests is planned to be served by the field engineers throughout the working day;
- the *Districting* policy, that is, the criteria to follow in order to assign each technician a territory;
- the *Cross-training* policy, that is, the criteria to follow in order to assign each technician a given set of skills (Agnihotri, Mishra and Simmons 2003; Upton 1994);
- the *Spare parts management* policy, that is, the way spare parts are stored, reordered and delivered to the customers.

The independent variables, are those out of the service manager's control, while the dependent variables, are functions of the independent and the control variables.

In Table 1 the aforementioned variables are described in detail, while Figure 2 shows the delivery system design process.

As we see, the independent variables represent the *input* for such a process, because they identify the context in which managers have to work and that they cannot control. The values of the independent variables are totally known before taking any system designing decision.

The control variables, instead, represent – to some extent – the *output* of the decision making itself. They are, in fact, the aspects that managers have to decide about, in order to design a FSDS. We can therefore say that the values of the control variables are the results of the decision making process.

Finally, the dependent variables are those measuring the *outcome* and thus the effect of the decision taken by the service manager. We can identify them with the actual system's operational and financial performances (Agnihotri 1989).
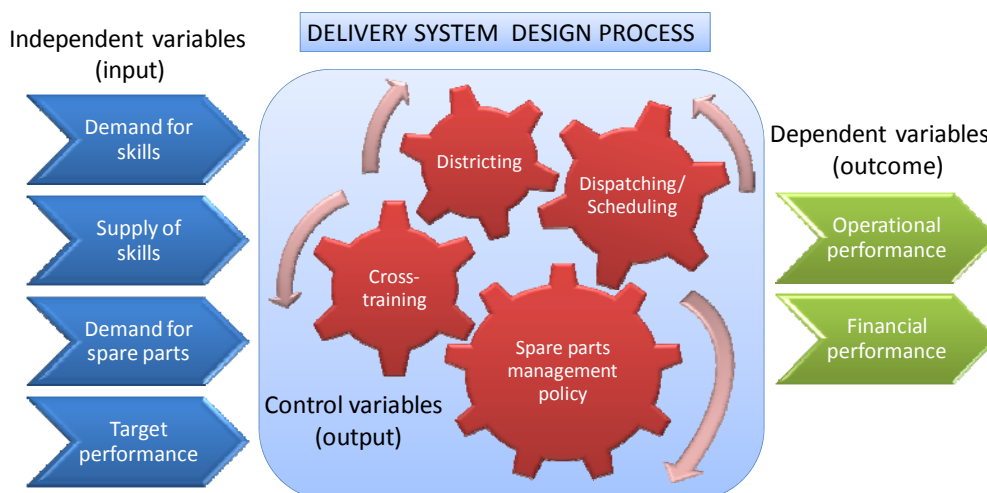

Figure 2: Delivery system design process

117

Table 1: Independent, control and dependent variables in the delivery system design process

Independent variables:	Control variables:	Dependent variables

1. Geographical distribution of the demand for skills. It depends on:
  a. the number of products to serve
  b. the Mean Time Between Failure of each type of product (and in case the different failure modes)
  c. the skills required to fix each type of product (and in case the skill required for each failure mode)
  d. the products' location
 	
2. The current geographical distribution of the supply of skills. It depends on:
  a. the number of field technicians available
  b. the locations (home address) of the field technicians available
  c. the skills mastered by each of field technician
 	
3. Geographical distribution of the demand for spare parts. It depends on:
  a. number of tasks requiring spare parts
  b. spare parts/store location
  c. availability of spare parts
 	
4. The target *operational* and *financial* performance that the delivery system is supposed to meet
	a. *Capacity*: number of field technicians to employ
	
	b. *Cross-training*: skills to impart to each field technician
	
	c. *Districting*: territories to assign to each field technician
	
	d. *Dispatching*: algorithm to use to dispatch field technicians
	
	e. *Scheduling*: algorithm to use to schedule field technicians
	
	f. *Spare parts management policy*: location and quantity of spare parts to keep in inventory
		4. *Operational* performance:
		  a. downtime
		  b. response time
		  c. travel time
		  d. answering time
		  e. first contact resolution rate
		  f. SLA compliance
		  g. resource utilization
		  h. etc…
		
		5. *Financial* performance :
		  a. revenues
		  b. costs
		  c. cash flows
		  d. etc…

4. INFORMATION NEED

As can be noticed the independent variables are a set of data that should be easily retrieved from the Enterprises' Information System (EIS).

As a matter of the fact:

1. all the data regarding the installed-base (i-base), the contracts, the service requests, the technicians and the spare parts are usually available in the companies' Enterprise Resource Planning system (ERP system);
2. the i-base is usually geo-referred with Geographical Information Systems (GIS) and the same instruments allow to define the areas the field technicians are assigned to;

3. the product performance in terms of reliability and availability can be easily obtained through statistical tools, starting from the data available in the ERP system, evaluating in particular the functioning time intervals of the machines as the difference between two consequents service requests regarding the same item.

These three sources should provide all the needed data. The information flows that allow the DSS (represented by a simulation tool) to simulate the real situation are shown in Figure 3.
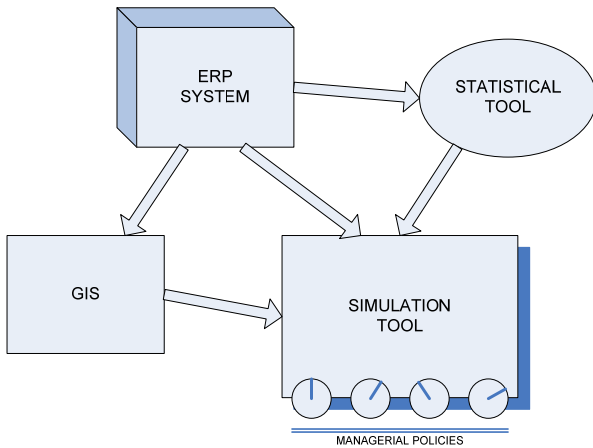
Figure 3: Information flows

5. THE MODEL

A DSS to support the FSDS design should therefore be able to retrieve the required data from the companies' information systems and to perform simulations to test the effectiveness of different managerial policies.

Such a tool has thus to allow to:

- evaluate the variables that are actually relevant in the FSDS designing;
- measure – ex-ante – the effects that the typical decisions taken by service managers could have over the delivery system performance.

Due to the complexity characterizing the field service environment, the methodology adopted for modeling the field service system should be the discrete event simulation (Chung 2003; De Felice 2007; Law and Kelton 2000).

The use of discrete-event simulation, instead of other techniques such as the queuing theory, is fully justified by the number of variables to consider and the randomness of the phenomena to model (Banks 1998; Kelton and Sadowski 2003; Perros 2007).

A representation of the conceptual model that we have developed is shown in Figure 4.

In the model the service requests are entities characterized by several attributes:

- *product type* (and thus the needed technical data);
- *problem type* (and thus the needed skills of the technician);
- *installed base localization* (and thus the needed location of the technician);
- *service level agreement* (and thus the performance constraints);
- *need for spare parts.*

While the process is in progress the data about the problems occurred are computed by the statistical tool and the values of the operating reliability are updated.

Each incoming entity has to be assigned to a technician and could require some spare parts.

Both technicians and spare parts are resources with limited capacity, characterized by attributes such as the geographical location and the service requests they can be used for.
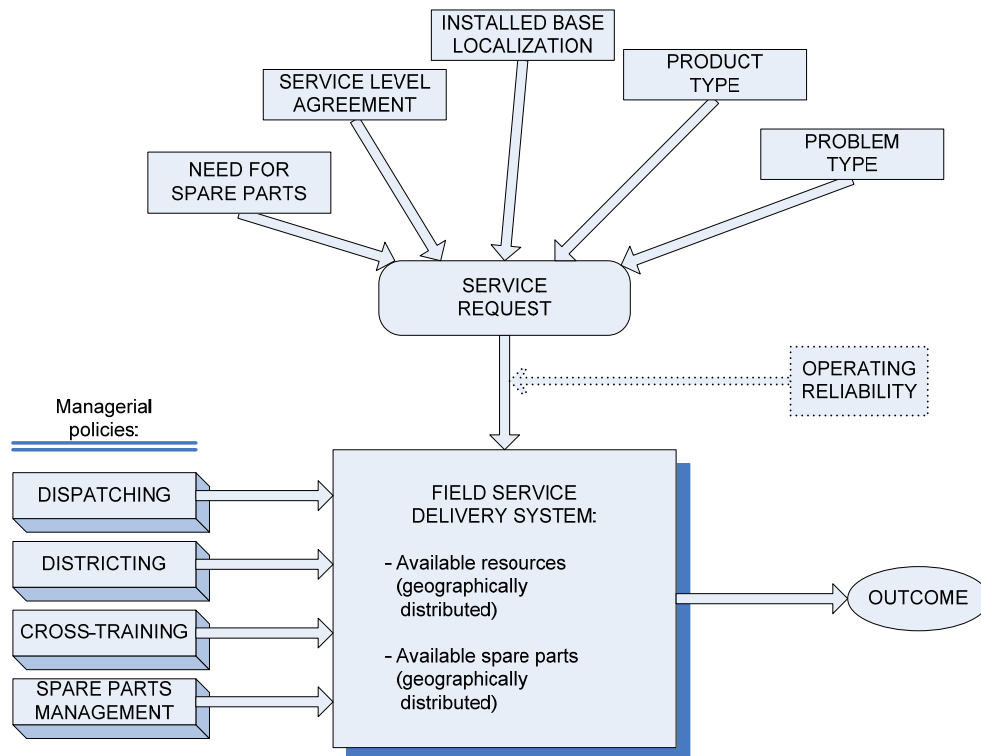

Figure 4: Conceptual model

119

These resources' attributes depends on the chosen managerial policies. Different policies lead to different values of:

- overall number of field technicians;
- dimension of the area the technicians have to cover;
- skills distribution over the technicians;
- spare parts location and quantity.

The model logic, based on the managerial policies, assign the entities to appropriate resources, creating an entities' flow.

Fundamental element of the simulation model is the dispatching algorithm (implemented in the model logic), that is, the working rule of the model.

This algorithm depends on the values of the control variables and regulates the entities' flow, considering resources' attributes and availability.

It's therefore clear that different managerial policies lead to different algorithms and so to different results of the service delivery process.

The outcome of the process can be finally calculated in terms of:

- resources utilization,
- lead time,
- downtime,
- percentage of fixed requests,
- constraints compliance,
- etc…

The DSS we describe has to be suitable for every kind of company that has to cope with field service delivery problems.

Certainly, each specific application needs some efforts to effectively be able of representing the situation. These efforts aim at analyzing the business processes, evaluating the data availability, defining the rules and the algorithms that are suitable for the case (Pidd, 1992; Ross, 1990).

In general, to create a functional DSS, we have to:

1. develop and verify reliable, flexible and parametric models, able to reproduce "in vitro" different field service delivery systems;
2. validate the models making use of the real data;
3. use the validated models to carry-out scenario analysis and experiments;
4. develop and test algorithms for the scheduling, dispatching and districting optimization by means of the simulation models themselves.

6. CASE STUDY

As a pilot experiment we tried to model the FSDS of a big multinational company that manufactures and services office imaging products.

Through the simulation tool (developed making use of the Rockwell Arena © suite) we have assessed the statistical significance of the effects of the control variables over the overall system's performance.

In order to build and apply the model to the case study, we performed:

- business processes analysis,
- conceptual model of the system,
- data collection and manipulation,
- actual creation of the simulation model,
- experimental design.

The field service delivery process we identified totally follows the model we showed above. We have been able of retrieving all the needed data from the company information system.

In particular, the data we considered are the following:

- service request arrival date and time,
- geographical location of the products (i-base localization),
- characteristics of the products requiring support (product type),
- type of the problems that can occur to each type of products (problem type),
- failure modes and possible solutions,
- needed skills (in dependence of the problem type),
- needed spare parts (in dependence of the problem type),
- skills profiles, availability and work areas of the field technicians,
- instantaneous position of the technicians,
- spare parts availability and location,
- average travel time and work time required to fix the problem,
- target operational and financial performance contractually defined (e.g. Service Level Agreement – SLA),
- service contract validity,
- failure rates characterizing the i-base (operating reliability).

The simplified tool we developed is not able of retrieving automatically the required information from the EIS. We therefore manually extracted the data from the ERP system and the GIS trough worksheets, we performed reliability analysis with a statistical tool and we used the obtained data as an input for the simulation model we created with Rockwell Arena.

After having created the simulation model, we found its stability parameters (Guttman, Wilks and Hunter 1971; Montgomery 2002) and then we performed a Design Of Experiment (DOE) (Box and Hunters 1978; Mood, Graybill and Boes 1974) selecting three of the control variables described above: dispatching, cross-training and districting.

What we did with this variables is perfectly extendible to the complete system, it's just a matter of data amount, time and computational power.

A 2^k DOE analysis showed the influence that changes in the control variables (and in their interaction) give to the overall system's response.

We defined 2 different levels for each control variable and we ran the model in each of the 8 different possible combinations showed in Table 3.

Table 2: Control variables considered in the case study

Control variable	"+" level	"-" level
Dispatching	STQ	SNQ
Cross-training	High flexibility	Low flexibility
Districting	4 zones	2 zones

Table 2 shows the levels we defined for the control variables.

For the dispatching, the levels are:

- Shortest Time in Queue (STQ) policy, that is, the entity is assigned to the resource with the minimum expected waiting time;
- Shortest Number in Queue (SNQ) policy, that is, the entity is assigned to the resource with the minimum queue length.

For the cross-training, the levels are:

- high flexibility policy, that is, all the field technicians are able of fixing 3 different problem types;
- low flexibility policy, that is, all the field technicians are able of fixing 2 different problem types.

For the districting, the levels are:

- narrow districting policy, that is, a territorial partition in 2 zones;
- broad districting policy, that is, a territorial partition in 4 zones.

Table 3: The 8 variables combinations of the 2k DOE

Case	Dispatching	Cross-training	Districting
A	+	+	+
B	-	+	+
C	+	-	+
D	-	-	+
E	+	+	-
F	-	+	-
G	+	-	-
H	-	-	-

In Figure 5 and 6 the comparison between the different possibilities is shown, considering the mean downtime of the products to serve as a performance indicator.

A treatment is the variation that is felt while passing from the "-" level to the "+" level in each of the three variables (1 – dispatching, 2 – cross-training and 3 – districting), or contemporaneously in more of them (12, 13, 23, 123) (Rotondi 2005).

The proportional variation is the effect that treatments causes on the overall system response.

Treatment	Proportional variation
1	- 9.9 %
2	+ 3.8 %
3	- 8.2 %
12	+ 5.0 %
13	+ 13.6 %
23	- 5.6 %
123	- 7.6 %

Figure 5: Performance variation due to the treatments
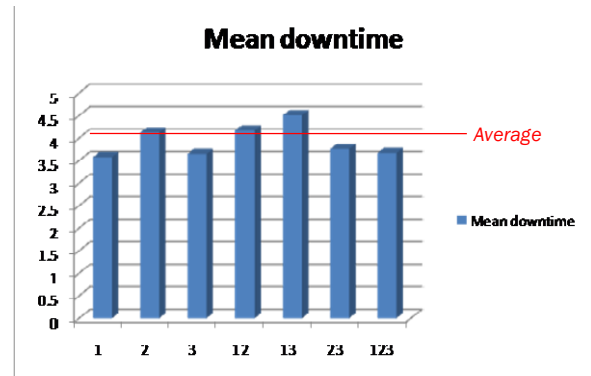
Figure 6: Evaluation of the impact of the considered factors on the performance

The analysis of variance we performed on the simulation model's results, demonstrates that the control variables are statistically significant with a confidence interval of 95%. This says that all the selected factors and their interactions give a significant impact on the performances of the FSDS, so it is correct to analyze them.

In this case the most important element to set seems to be an appropriate dispatching policy (1), followed by a correct districting (3) and then by a proper training and skills spreading policy (2).

We also notice (13) that the performance improvement given by a change in the dispatching policy is much more marked if it is applied together with a broad districting policy (more dispatching alternatives imply higher performance differentials).

REFERENCES

Agnihothri, S., 1989. Interrelationships Between Performance Measures for The Machine Repairman Problem, *Naval Research Logistics*, Vol. 36, No. 3.

Agnihothri S., Chakravarthy S., 2005. Optimal Workforce Mix in Service Systems with Two Types of Customers, *Production and Operations Management*, Vol. 14, No. 2, pp. 218-231.

Agnihothri, S., U. Karmarkar, U., 1992. Performance Evaluation of Service Territories, *Operations Research*, Vol. 40, No. 2.

Agnihothri, S., Mishra, A., Simmons, D., 2003. Workforce Cross-Training Decisions in Field Service Systems with Two Job Types, *Journal of the Operational Research Society*, Vol. 54, No. 4, pp. 410-418.

Agnihothri, S., Mishra, A., 2004. Cross-Training Decisions in Field Services with Three Job Types and Server-Job Mismatch. *Decision Sciences*, vol. 35, n. 2, pp. 239-257.

Agnihothri, S., Narasimhan, S., Pirkul, H., 1990. An Assignment Problem With Queueing Time Cost, *Naval Research Logistics*, Vol. 37, No. 2.

Banks J., 1998. *Principles of Simulation, Handbook of Simulation*. John Wiley & Sons, Inc.

Blumberg, D. F., 1991. *Managing Service as a Strategic Profit Center*. McGraw-Hill.

Box, G. E. P., Hunter, J. S., 1978. *Statistics for Experimenters*. John Wiley & Sons, Inc.

Chung, C. A., 2003. *Simulation Modeling Handbook*. CRC Press.

De Felice, F., 2007. *Applied Simulation and Modelling*. Acta Press.

Guttman, I., Wilks, S. S., Hunter, J. S., 1971. *Introductory Engineering Statistics*. John Wiley & Sons, Inc.

Kelton, W. D., Sadowski, R. P., 2003. *Simulation with Arena*. McGraw-Hill.

Law, A. M., Kelton, W. D., 2000. *Simulation Modeling and Analysis*. Mc Graw-Hill.

Montgomery, D. C., Runger, G. C., 2002. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, Inc.

Mood, A. M., Graybill, F. A., Boes, D. C., 1974. *Introduction to the theory of statistics*. McGraw-Hill.

Perros, H., 2007. *Computer Simulation Techniques: the definitive introduction*. Computer Science Department, NC State University, Raleigh, NC.

Pidd, M., 1992. *Computer Simulation in Management Science*. John Wiley & Sons, Inc.

Ross, S. M., 1990. *A course in Simulation*. Macmillan

Rotondi, A., Pedroni, P., Pievatolo, A., 2005. *Probabilità, Statistica e Simulazione*. Springer Verlag Italia.

Upton, D. M., 1994. The management of manufacturing flexibility, *California Management Review*, vol. 36, pp. 72-89, Winter.

Visintin, F., 2007. *Designing after sales service: strategy, operations and information technology*. Thesis (PhD). University of Florence.

AUTHORS BIOGRAPHY

Mario Rapaccini took his Laurea Degree with honors in Mechanical Engineering (5-yr-course) at Florence University in April 1996. He is a professional engineer since 1996. In may 2000 he achieved Ph.D. (3-yr-course) discussing the thesis "Advanced tool for configuration and impact assessment of Integrated Municipal Solid Wastes Management Systems". Currently, he's assistant professor in SSD ING-IND/35. Research topics covered are: managerial economics and business organisation, ICT, simulation modeling and analysis (SM&A), business process re-engineering (BPR). He's fellow of AiIG, EurOMA, ANIMP, AIRO and ANIPLA.

Filippo Visintin was born in Prato on 1979. In 2003 he graduated with honors in Management and Production Engineering at University of Bologna. In 2004 he was admitted to attend the doctoral program in Industrial Engineering of the Department of Energetics "Sergio Stecco", Florence University. In 2006 he was visiting research scholar at the School of Management of the State University of New York at Binghamton. In 2007 he took a Ph.D in Industrial Engineering defending a thesis titled "Managing after sales services: strategy, operations and information technologies". From December 2007 he is researcher and assistant professor (SSD ING-IND/35) at Florence University, Faculty of Engineering. His current research interests are in the service operation management area. He's fellow of AiIG, POMs and EurOMA.

Alessandro Sistemi was born in Pisa in 1983. He took his Laurea Degree with honors in Management and Production Engineering at Florence University in December 2007 discussing the thesis "Field service delivery optimization for industrial and professional systems". Since February 2008 he is research fellow at the Department of Energetics "Sergio Stecco", Florence University.

ADVANCED MODELS FOR INNOVATIVE DECISION SUPPORT SYSTEMS IN BROADCASTING SCHEDULE PLANNING & MANAGEMENT

Agostino Bruzzone[a], Marina Massei [b], Luca Pierfederici[c]

[a] McLeod Institute of Simulation Science
DIPTEM University of Genoa
Via Opera Pia, 15, Genoa, 16145, ITALY

[b] [c] Liophant Simulation
Via Molinero 1, Savona, 17100, ITALY

[a]agostino@itim.unige.it, [b]marina.massei@liophant.org, [c]liophant@itim.unige.it

ABSTRACT

The paper describes a set of advanced models to support strategic and operative decision in the television schedule planning and to analyze and to control related activities.

The research focus is to define support systems for modeling this complex scenario and its variables, and for permitting the elaboration and implementation of efficient strategies in the related management; in this context this kind of model is a critical to support marketing and advertising space allocation, broadcast planning, share parameter analysis and control, evaluation of the accuracy of the forecast activities and the effectiveness of implemented strategies.

Keywords: innovative decision support systems, broadcasting, schedule planning and management

1. INTRODUCTION

Broadcasting planning is a very complex framework, where simulation can introduce quantitative models able to provide a direct support in planning and controlling activities. In fact it is possible to consider the schedule of broadcasting stations as a bi-dimensional entity where products (programs, movies, news, soap operas, music, etc.) are organized in a time table and assigned to a channel.

Profits principally derive from selling of advertising time spaces during the product broadcasting, and obviously, in this context, profits deriving from the television market are the most significant.

The Rules and policies to maximize turnover derived by selling advertizing buckets need to evaluate many different variables, such as actors, concept and format parameters, channel characteristics etc. This introduce a very complex scenario and currently, quite often, the only tools available to obtain reliable analyses are based on experience and intuitions of decision makers.
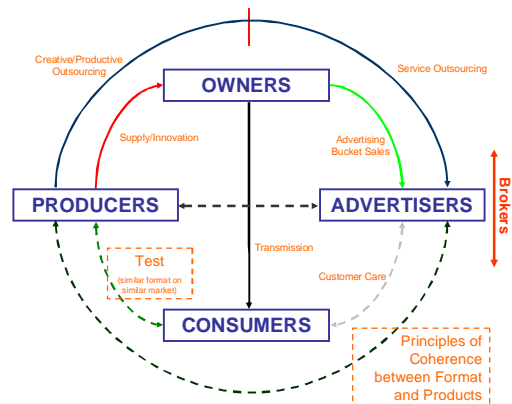

Figure 1: The Actors of the Broadcasting Planning

The actors of this process and their related activities are characterized by many interactions involving supplying, selling, buying and sharing of products and services. It is quite common that owners of broadcasting channels outsource the production of their formats or buy complete finished products in order to guarantee an high performance level of format selling; in fact currently this phenomena is increasing and producers are extending their investigations in order to capture and anticipate advertiser needs.

However buying of outsourced formats introduce a pretty new kind of problem for broadcast enterprises. In fact producers policies for selling of format is currently introducing the concept of stock management in this market sector. Currently formats are sold in a "container", it means that the channel owners in order to acquire a specific desired program have to buy also other products.

This triangulation among owners, producers and advertisers results to be efficient only if it is supported by effective information management; in fact it is required to have deep knowledge and understanding of all the interconnection among programs, advertisings and channels, so it is critical to evaluate the incidence of time buckets during the day not only in term of

audience share, but also in term of specific impact of commercials on the related consumers.

Based on these consideration in order to be competitive it is important to develop an efficient overview of the effects related to the interaction among different products and time-band/channels; by this understanding it becomes necessary to develop models able to define quantitatively the system behavior in order to optimize number of spectators for attracting advertising actors.

Therefore it is fundamental to define customer profile related to each specific period of the day, channel and program type; this is especially important for enterprise controlling multiple channels in order to properly tune their program management policies.

In this view, companies that want to promote their products by commercials or promotions are looking forward for buying spaces during the programs and time-bands where they expect to get more potential customers, in fact their choice is driven also by considering the specific characteristics of each broadcasting channel.

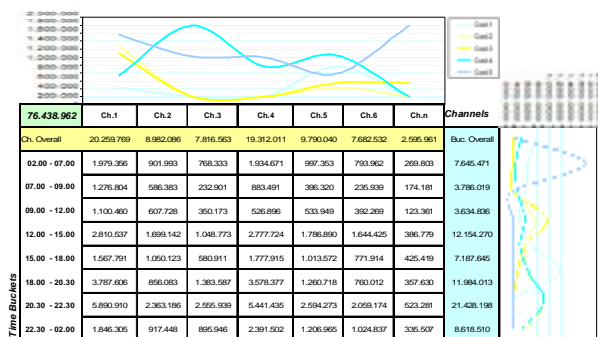| 76.438.962	Ch.1	Ch.2	Ch.3	Ch.4	Ch.5	Ch.6	Ch.n	Channels
Ch. Overall	20.259.769	8.982.086	7.816.563	19.312.011	9.790.040	7.682.532	2.595.961	Buc. Overall
02.00 - 07.00	1.979.356	901.993	768.333	1.934.671	997.353	793.962	269.803	7.645.471
07.00 - 09.00	1.276.804	586.383	232.901	883.491	396.320	235.939	174.181	3.786.019
09.00 - 12.00	1.100.460	607.728	350.173	526.896	533.949	392.269	123.361	3.634.836
12.00 - 15.00	2.810.537	1.699.142	1.048.773	2.777.724	1.786.890	1.644.425	386.779	12.154.270
15.00 - 18.00	1.567.791	1.050.123	580.911	1.777.915	1.013.572	771.994	425.419	7.187.645
18.00 - 20.30	3.787.606	856.083	1.383.587	3.578.377	1.260.718	760.012	357.630	11.984.013
20.30 - 22.30	5.890.910	2.363.186	2.555.939	5.441.435	2.594.273	2.059.174	523.281	21.428.198
22.30 - 02.00	1.846.305	917.448	895.946	2.391.502	1.206.965	1.024.837	335.507	8.618.510

Figure 2: Customer Profiles

In the past the scenario defining audience type and attendance on channels and time-bands was quite well defined among competitors; this situation changed after the growing success of cable tv and satellite stations (sat stations); in fact the development and diffusion of these new technologies opened new areas and policies of competitiveness.

For instance the born of channels aimed on well defined and restricted areas (i.e. one single sport, one kind of music, etc.) guarantees to specific advertisers a strictly selected field of potential customers, even if the absolute share could result to be lower respect other channels where the attendees are not really interested on their specific products; a very good example are channels related to cars for automotive companies, or fish and haunt channels for the companies involved in related production.

The changed in policies in programs production/acquisition and this new competitive framework stress the need for innovative efficient methodologies able to provide accurate analysis of all the system variables. Considering the high stochastic nature of these processes and the strong interrelations among the different factos, it is evident that simulation.

In fact this innovative models require to consider many variables such as to Time Bucket Structure, Customer Profile, Attendance Profile, Format Structure, Channel Characteristics and the additional conditions and correlation among them, in order to provide an efficient and quantitative decision support systems.

2. STATE OF THE ART

Data related to audience profile are collected by various corporations; this information is composed by both public and private data and includes a wide range of data, involving timeframe, channels, customer age, cultural background, spending power, etc.

Several studies and researches have been focusing on channel performance analysis and audience evaluation in order to give an efficient and useful interpretation of these data based on models and technique for correlating them with desired goals.

For instance, A.Vanheuverzwyn and J. Chaskalovic approached the problem of confidence of audience indicators in radio broadcasting applying bootstrap techniques to define accuracy level for each considered indicator, station, day frame and target; they built models based on multiplicative structures for confidence interval, using data mining exploration to investigate potential dependencies of the confide interval regarding the other variables and recently they implemented neural networks to evaluate stability of the models ("Innovation in Estimation of a Reliable Approach for Radio Audience Indicators" , 2007).

Where many data are available and the problem is related to their correlation there were developed other models, implemented in ISL's *CLEMENTINE*, as a data mining toolkit based on AI techniques, that allows to identify patterns and business rules which elude traditional statistical analysis, OLAP (Online analytical processing) or visualization techniques.

Currently, a common main goal of researches is to elaborate efficient models to treat and aggregate massive data collections to extrapolate useful and significant information.

Several researches are currently focusing on identifying the significant variables and dependencies on these systems; by this approach it becomes possible to investigate scenarios and to forecast competitors results.

3. A CASE STUDY

Italy is the reference point for this research; in fact the proposed scenario is characterized by a small number of tv broadcasters collecting the majority of audience share; in this frame it exist an authority that provides reference figures for classification of spectators and data related to presence profiles during time ranges during the day.

Actually in the scenario few main competitors are operating on air channels, while cable tv have a very limited share in term of audience and advertising sales, so these elements are not significant in our case.

Otherwise the field of tv channel includes a small number of main competitors (6 channels cover 70% audience) and satellite televisions are conquering a growing up share of the market.

As anticipated a national society (named Auditel) provides info related to results of the competitors in terms of number and type of spectators distributed for channels and time frames, and collects these data with different time resolutions (i.e. each two minutes).

However these services are losing their incidence in strategic decisions for the higher complexity level of the competition and market, where now the number of spectators loses incidence on interests of advertisers by comparison with aiming of a specific consumer profile. This fact, combined with the change of programs acquisition process, moving to strong outsourcing, determines the needs of innovation in decision support systems, devoted to guarantee more significant data, and more accurate knowledge of customer and format peculiarities; this requires to develop ad hoc simulation models able to allow what-if analyses and experimental analysis for investigating strategic policies changes.

4. SIMULATION MODELS

As anticipated in introduction, the proposed scenario requires to proceed in the identification of the variables and their definition for modeling these processes. Due to the high variability of market and products and to the strong influence of stochastic factors it is critical and very hard to model relations representing the interconnections among the different objects, variables and parameters.

Critical operations in strategies are identified in forecasting activities, format warehouse management, customer targeting and schedule planning.

Forecasts

Forecasting activities related to the estimation of appreciation level of a format by the spectators is currently mostly based on the experiences of decision makers and qualitative data; sometimes it is obtained also based on results of similar products on similar market, but it is currently increasingly the importance to have a good control on decision strategies by providing quantitative measure on these issues. In fact goal is not to provide models that guarantee very precise info about programs (this is a utopist crystal ball model), but it is required to obtain a statistical based support to evaluate policy performances as well as level of coherence between planning decisions and enterprise mission.

Format Warehouse Management

The production outsourcing for television concepts open issues in strategic planning, due to the fact that producers and brokers sell formats not one by one, but in "containers" including different programs. This is leading to the creation of television programs "warehouses" for broadcasters, and to develop stock

management policies, cost and risk analysis and control procedures.

Customer Targeting

The more aggressive competition in this complex scenario requires a much more effective and efficient customer profile identification. This means that new models need to analyze and evaluate the proper target for advertisers in different channels/timeframes.

Schedule Planning

Success in a television program is related to the match of its format peculiarity with customer profile and channel audience target; the key for obtaining this results is the schedule planning. In fact to move a program into a different timeframe (from morning to afternoon) generates significant difference in term of results.

At the same time it is possible to operate otherwise by defining policies for changing customer profile in different time bucket/channel by attracting them with programs that better match their interests and expectations.

It is evident that the complexity of above mentioned phenomena requires development of specific models; in order to satisfy all these needs it is fundamental to reduce the complexity of the problem by subdividing the overall frame in different objects and entities.

The scenario proposed by the authors is concentrating on the 7 main competitors, corresponding to the 6 channels with wider catchment area and a meta-entity including performance of all the sat channels.

4.1. Entities and Variables

As anticipated the authors are presenting a new approach for creating models and methodologies to define, investigate and analyze the above described scenario. In particular entities and variables are proposed as following; it was identified different main classes for objects to create the scenario, and for each class their attributes are defined.

Format Concept Object

- Type of Contents;
- Star Power;
- Budget;
- Innovation Level.

Customer Object

- Customer Static Profile;
- Customer Dynamic Profile;
- Spending Power Profile.

Calendar Object

- Time Table Definition;
- Weather;

Channel Object

- Typology (Sat/by-Air);
- Referred Customers;
- Channel Mission Coherence Rate;
- Continuity Rate Between Time Buckets.

Innovative models are implemented to valorize attributes, identify and evaluate functions that interconnect different classes in order to develop an efficient decision support systems, based on analysis, aggregations and elaborations of the wide range of available historical data.

4.1.1. Format concept object

This represent the tv program object and it is characterized by the following attributes:

Type of Contents Program Concept

A set of variables is considered significant, in particular a coherence rate (0 ÷1) between the program concept and the following topics:

- economy;
- travels;
- culture;
- cooking;
- ecology;
- topicality;
- violence;
- information;
- interactivity;
- emotions;
- national Religion;
- others Religions;
- fear ;
- comedy;
- strain;
- national favorite sport;
- others sports;
- others.

This variable could be elaborated by Data Fusion algorithms to be further developed in next phase of the research.

Star Power

Star power is a score assigned to a program for evaluating number and incidence of actors/frontmen; this attribute is defined by creating fuzzy allocation matrixes (FAM) involving the following variables:

- National assigned awards:
1. Number;
2. Prestige;

- International assigned awards:
1. Number;
2. Prestige;

- Reward;
- Appearance:
1. Number;
2. Prestige.

It is possible to combine the different factors and to apply fuzzy rules for processing the FAMs.

Budget

Budget represent an important attribute of a tv program; it is proposed to define and evaluate this factor in term of amounts invested in production, marketing and distribution by defining fuzzy variables; in fact fixing the fuzzy membership functions and related ranges it is possible to extract meaning from the budget attributes in relation to its specific nature.

Innovation Level

The innovation levels are also to be defined in term of ranges of innovation considering budget invested and scores attributed by Subject Matter Experts (SME); this data are processed by *Fuzzy Logics* for providing an overall parameter estimator.

4.1.2. Customer object

The Customer represent the audience and its characterization it is defined by the following parameters:

Customer Static Profile

This attributes includes static attributes representing the characteristics of the audience (age, social status, etc)

Customer Dynamic Profile

These variables are defining the interest subjects on the different topics and preference on timeslots; in fact this parameters is both an input and output of the system: it is expected that SME will define reference timeslots and preferences and the models will check correlation among these hypotheses and overall results.

Spending Power Profile

This profile allows to correlate customer speding power on the different advertiser areas.

4.1.3. Calendar object

Calendar represent the object defining the timeslots composing the overall time table with the schedule of all programs, but includes also parameters related to weather for correcting historical data: i.e. bad weather during holiday period increase overall audience levels.

4.1.4. Channel object

The channel are defined as the frames where to broadcast the programs and among the attributes it is defined the channel typology as well as parameters provided by experts that define the channel profile;

these parameters (i.e. coherence) have to be estimated and validate by the simulation model.

It is critical to properly define, model and initialize each attribute in order to create an effective knowledge representation of the entities involved the scenario.

5. METHODOLOGIES

To be able to complete analysis it is necessary to identify interactions and dependencies between the objects. The authors are planning to develop ad hoc artificial neural networks to be used to correlate the overall data representing general share on the programs, channels, timeslots for the identification of parameters to be used in the simulation model.

In addition considering the large number of variables and the high quantities of available data it is expected that specific techniques need to be developed for processing the scenario; in this case the authors are looking forward for the combined use of different techniques; for instance Data Fusion methods are promising for correlating different info among time and sources, while Neural Networks methodologies are expected to be very effective to identify clusters of similar data and to correlate parameters; in addition in order to provide an understandable and significant description of so many parameters the variables definition have to take advantage of Fuzzy Logics models.
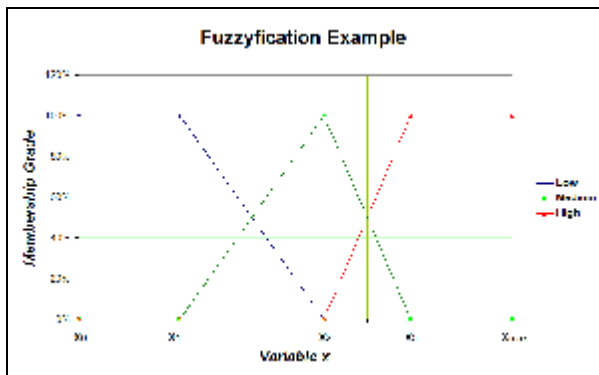

Figure 3: A Fuzzyfication Example

$$MV_{c-1} \equiv CV_c \equiv HV_{c+1} \qquad (1)$$

- MV_c = Minimum Value for the c-Cluster (Membership Grade = 0%)
- CV_c = Central Value for the c-Cluster (Membership Grade = 100%)
- HV_c = Higher Value for the c-Cluster (membership Grade = 0%)

This membership based on triangular functions overlapping 50% is pretty easy, however allows SME to quickly define the three reference values to be used to classify the attribute of a format: for instance the definition of budget for a variety program is fixed by defining these three levels, but the program is not discretely defined, while a membership grade is

computed for each class (inexpensive, regular and very expensive) based on its specific production budget.

The fuzzy allocation matrices allow to combine different parameters in order to estimate their interrelations.

Data fusion algorithms are used to estimate the value of an attribute as combination of the different features measured along the time evolution.
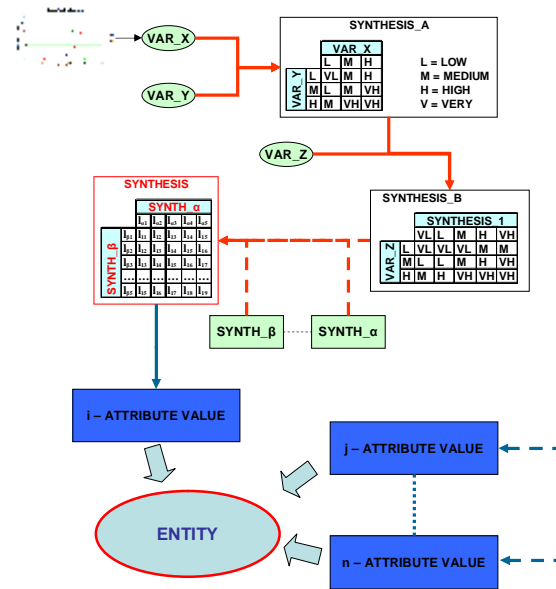

Figure 4: Data Fusion Architecture

In fact broadcasting scenarios are characterized by very short Life Cycles for the products, than is not usual to have enough historical data to elaborate forecasts, so Data Fusion methodologies can be applied to extract info by combining different data. In effect accuracy of forecast cannot be guaranteed due to the high influence of stochastic factors, but the comparison of these results with SME allows to better estimate effectiveness of decisions.
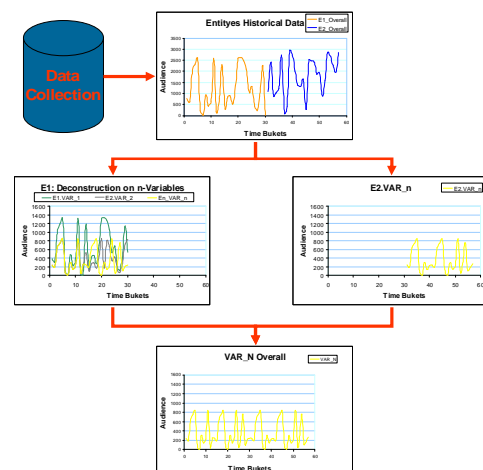

Figure 5: An application of Data Fusion Methodologies

Neural Models allows to check consistencies among elements estimated by SME and historical overall data; for instance it is expected to create the below proposed

127

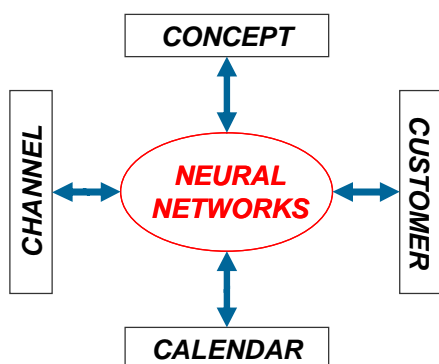scheme for correlating parameters of different objects, where not explicit rules (i.e. FAM) are known.

Figure 6: Correlation for Parameters of Different Objects

6. CONCLUSIONS

The authors proposed a first step forward in creating simulation models for tv broadcasting; the authors formalize the definition of general goals for these models in term of the definition of the schedule and creation/management of program warehouse; for this complex framework it is proposed a case study and the related object definition and attribute identification.

At this stage of the research models have been developed to support definition and measure of the attributes, while the authors are currently working for testing different integrated methods and techniques to be used for creating this new simulator.

The authors have collected data related to the scenario to complete preliminary test, however in the future it will be possible to present experimental results obtained by dynamic verification and validation of these models.

ACKNOWLEDGMENTS

REFERENCES

Mosca R., P.Giribone & A.G.Bruzzone (1993) "Application of Neural Networks to On-Line Simulations of Plant Processes", Proceedings WNN93, San Francisco, November 7-10.

Giribone P. & A.G.Bruzzone (1993) "Stochasticity of Neural Network Applied to Industrial Simulators", ITIM Technical Report, Genova.

Mosca R., P.Giribone, A.G.Bruzzone (1994) "Procedural Approach for a Neural Network Based Analysis of an Industrial Plant: Automatic Bottling Line", International Journal of Flexible Automation and Integrated Manufacturing, Vol.2, no 3. pp. 259-273.

Mosca R., P.Giribone, A.G.Bruzzone (1994) "An Innovative System of Artificial Intelligence based on Neural Network for Technical and Management Analysis of Composting Plants",

Proceedings of FAIM94, Blacksburg, VA (USA), May 8-11.

Fausett L.V. (1994) "Fundamentals of Neural Networks", Prentice Hall.

Giribone P. & A.G.Bruzzone (1995) "Artificial Neural Networks as a System for Production Planning in an Alimentary Plant", Proc. of Simulators International XII, Phoenix, April 9-13.

Giribone P. & A.G. Bruzzone (1995) "Neural Networks And Fuzzy Control For Bulk Terminal Operating Management", Proc. of Summer Simulation '95, Ottawa, July 23-26.

Bruzzone A.G. (1996), "Smart Integrated Interface for Evaluating and Supporting Human Performances", ESPRIT Infromation Technology: Basic Research Working Group 8467 Simulation for the Future: New Concepts, Tools and Application (SiE-WG), January 29, Bruxelles.

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.

Mosca R., Giribone P., Bruzzone A.G. (1996) "Study of the Behaviour of Flexible Production Systems using Neural Nets", International Journal of Production & Control, vol. 7, no.5, 462-470.

Mosca R., P.Giribone, A.G.Bruzzone, A.Orsoni, S.Sadowski (1997) "Evaluation and Analysis by Simulation of a Production Line Model Built with Back-Propagation Neural Networks", International Journal of Modelling and Simulation, Vol.17, no.2, pp.72-77.

Giribone P., Bruzzone A.G. (1997) "Design of a Study to use Neural Networks and Fuzzy Logic in Maintenance Planning", Proc. of Simulators International XIV, SMC'97, Atlanta, Georgia, April.

Giribone P, Bruzzone A.G. M.Antonetti, G.Siciliano(1997) "Innovative Energy Management Techniques In Telecommunication Stations Through The Application Of Neural Models", Proceedings of 1st World Congress on Systems Simulation, Singapore, September 1-4.

Gurney K. (1997) "An Introduction to Neural Networks", CRC.

Bruzzone A.G., Giribone P., Revetria R., Solinas F, Schena F. (1998) "Artificial Neural Networks as a Support for the Forecasts in the Maintenance Planning", Proceedings of Neurap 98, Marseilles, 11-13 March.

Bruzzone A.G., Giribone P. (1998) "Artificial Neural Networks as a Meta-Modelling Techniques for Car Component Industrial production", Proceedings of Neurap 98, Marseilles, 11-13 March.

Bruzzone A.G., Giribone P. (1998) "Quality Service Improvement by Using Human Behaviour Simulation", Proceedings of ESM98, Manchester, UK, June.

Haykin S. (1998) "Neural Networks: A Comprehensive Foundation", Prentice Hall.

Cacciabue P.C. (1998) "Modelling and Simulation of Human Behaviour in System Control (Advances in Industrial Control)", Springer-Verlag Telos.

Giribone P., A.G. Bruzzone (1999) "Artificial Neural Networks as Adaptive Support for the Thermal Control of Industrial Buildings", International Journal of Power and Energy Systems, Vol.19, No.1, 75-78 (Proceedings of MIC'97, Innsbruck, Austria, February 17-19, 1997).

Bruzzone A.G., Revetria R. (1999) "Artificial Neural Networks as Support for Logistics in Super-Market Chains", Proceedings of HMS99, Genoa, September 16-18.

Bruzzone A.G. , C. Burlando, F.Queirolo (2001) "Simulation as General Support for Validating Data Fusion Algorithms", Proceedings of HMS2001, Marseille, October 15-17.

Bosque M. (2002) "Understanding 99% of Artificial Neural Networks: Introduction & Tricks", Writers Club Press.

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., B.C. (2005) "Using Modelling & Simulation As Framework For Testing New Solutions Devoted To Secure Global Flows Of Goods And People", Invited Paper in Proceedings of NATO ARW Data Fusion Technologies for Harbour Protection, Tallinn, Estonia, June 27-July 1.

Bocca E., Massei M., B.E. (2005) "Artificial Neural Networks As Support For Production Control In Aerospace Industry", Proceedings of ESS2005, Marseille, 20-22 October.

Bocca E., Bruzzone A.G., Rocca A. (2006) "Algorithms devoted to Reproduce Human Modifiers in Polyfunctional Agents", Proceedings of SCSC2006, Calgary, Canada, July 30-August 3.

Bruzzone A.G., (2007) "Challenges and Opportunities for Human Behaviour Modelling in Applied Simulation", Keynote Speech at Applied Simulation and Modelling 2007, Palma de Mallorca.

Bruzzone A.G., Massei M., B.M. (2007) "Demonstration For Human Behavior Modeling Within Civil Disorder Scenarios", Proceedings of Summer Computer Simulation Conference 2007, San Diego, July.

Bruzzone A.G., Frydman (2007) "Artificial Intelligence in PIOVRA - Swarm intelligence, Fuzzy logic, Data fusion, KBS", Keynote Speech at Italian PIOVRA Workshop, Rome, Italy, December.

Longo F., Viazzo S., Bocca E. (2005) "Developing Data Fusion Systems Devoted To Security Control In Port Facilities", Proceedings of Wintersim2005, Orlando, December 3-6.

Massei M., B.M., B.C.(2007) "Human Behavior Model for an HLA Federation", Proceedings of EUROSIW2007, Santa Margherita, Italy, June.

Reverberi A. (2006) "Human Behavior Representation & Genetic Algorithms Characteristics Fitting", Invited Presentation on Piovra Workshop, Savona, February 7.

Samarasinghe S. (2006) "Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition", AUERBACH.

Sheppard S. (1998) "Game-theoretic Models of Human Behaviour", Heretical Press.

Taylor B.J. (2005) "Methods and Procedures for the Verification and Validation of Artificial Neural Networks", Springer.

Vanheuverzwyn A., Chaskalovic J. (2007) "Innovation in Estimation of a Reliable Approach for Radio Audience Indicators", Proceedings of VM3 2007, Dublin, Ireland, June.

Viazzo S., B.M., B.C., (2005) "Human Behavioral Models For M&S Applications", Proceedings of ESS2005, Marseille, 20-22 October.

CONSIDERATIONS ON THE PARTICULAR FEATURES FOR PROCESS AND WORKFLOW MODELING

Victoria Iordan [(a)], Alexandru Cicortas [(b)]

[(a)] Department of Computer Science, West University of Timisoara, Romania
[(b)] Department of Computer Science, West University of Timisoara, Romania

[(a)]iordan@info.uvt.ro, [(b)] cico@info.uvt.ro

ABSTRACT
Modeling processes and workflow for complex systems constitutes a challenge for designers. The resources that are used in systems are in limited amounts and these must be shared between the processes and workflow instances activities.

Our considerations refer to the need of a data base in that the workflow constituents: instances and activities must be represented in a data base and also their time evolution. Concerning the resources that are used in the processes and in the workflow instances execution must be adequately represented. Two important classes of information systems, Workflow Management Systems (WfMSs) and Enterprise Resource Planning (ERP) systems, have been used to support e-business process redesign, integration, and management.

We intend to propose a model for resource representation and allocation and as conclusion propose considerations concerning the design of processes, workflows and workflow management systems.

This paper is partially supported by the project PN II 91-047/2007.

Keywords: modeling, workflow, processes, resources, resource management, activity coordination.

1. INTRODUCTION

Real world systems became more and more complex and require a lot of tools for their design and execution. In these systems having many processes is required that these processes: must cooperate; share the same system resources; coordinate and communicate one to other. In many systems there are different organizations that have their own processes that intervene in the fulfilling of some goals. The workflow that is designed to manage the process must be able to satisfy the above requirements.

Between the major problems we consider to be those of the resources and of the failures for both the processes and the workflows. How the ERP also manages the organization resources many considerations will be done also about it, based on (Kearney 2005).

Concerning the concepts, we intend to use those of the WfMC but we do not give the details that can be found in (WfMC 1994; WfMC 1996; WfMC 1999; Kearney 2005). Framework constitutes an indispensable feature in process and workflow design. Frameworks manage complexity by reducing vague or complex topics to a set of simpler questions (Sharp 2001). They organize work, maintain focus during interviews or facilitated sessions, ensure coverage of all aspects of a topic, and provide a standard format to document the findings. One of the most important frameworks is the one we use to consider all dimensions of a process: Workflow design, Information technology, Motivation and measurement, Human resources, Policies and rules, Facilities design.

Ontology or a reference framework (van der Aalst 2002) is a system of defined terms that describe a particular field of knowledge. Some specific aspects like:

- distribution of work (human interaction with processes);
- the relationship between the principal and the contractor;
- organizational structures and the management of processes;
- the role played by information systems in the establishment and management of business processes

must be seriously analyzed.

The paper is organized as follows. The second section treats the process modeling; in the third section, the specificities of workflow design are presented: BP re-engineering, workflow failure recovery, compensation and rollback and fault tolerance. The third section treats the resources. The fifth section analyzes the WfMS and ERP; the sixth section details our considerations concerning the resources, analyzes the failed workflow and their re-engineering. The last section sketch out the future works.

2. MODELING PROCESSES
2.1. General considerations on frameworks
A framework is a set of classes that embodies an abstract design for solutions to a family of related problems. According to (Fayad 1999, Eriksson 2002), the four primary benefits of frameworks can be described as follows:

- Modularity. Frameworks enhance modularity by encapsulating unstable implementation details behind stable interfaces. This modularity helps to improve software quality by localizing the impact of design and implementation changes. The localization reduces the effort required to understand and maintain existing software.
- Reusability. The stable interfaces provided by frameworks enhance reusability by defining generic components that can be reapplied to create new applications. Framework reusability leverages the domain knowledge and prior effort of experienced developers in order to avoid recreating and revalidating common solutions to recurring application requirements and software design challenges. Reuse of framework design components can yield substantial improvements in programmer productivity, as well as enhance the quality, performance, reliability, and interoperability of software.
- Extensibility. A framework enhances extensibility by providing explicit hot spots that allow applications to extend its stable interfaces. These extension points systematically decouple the stable interfaces and behavior of the application domain from the variations required by instantiations of the application in a particular context. Framework extensibility is essential to ensure timely customization of new application services and features.
- Inversion of control. The runtime architecture of a framework is characterized by an inversion of control, often referred as the "Hollywood Principle" (Don't call us, we'll call you). This architecture enables canonical application processing steps to be customized by event handler objects that are invoked via the framework's reactive dispatching mechanism.

The previous benefits of using frameworks allow having many advantages that are significant in design and modeling of systems. The modularity can be viewed in many ways. One is that where the system is service oriented and another is that where the system is component oriented. Permanent systems evolution needs for an adequate extensibility. When events occur, there actions of the system consist in invoking appropriate methods, which perform application-specific processing on the events. Inversion of control allows the framework to determine which set of application-specific methods to invoke in response to external events.

2.2. Details concerning process modeling

The process consists of a number of tasks that need to be carried out and a set of conditions that determine the order of the tasks.

The task is a logical unit of work that is carried out as a single whole by one resource (van der Aalst 2002). There exists another, in our opinion, view due to the fact that one task can be accomplished by more than one resource and in this paper that is the vision about it. The resource is the generic name for a person, machine or group of persons or machines that can perform specific tasks. This does not always mean to say that the resource necessarily carries out the task independently, but that it is responsible for it.

In most organizations there exists a hierarchy under which assignments that people receive can be passed on to people further down the hierarchy. A person who is assigned a task can be considered as a contractor, being also a resource. In many processes two or more tasks must be performed in a strict order. The processes include selection or repetition and as mechanisms are used in process structures: sequence, selection, parallelization, and iteration. The human participation in the process is very complex and in this paper it is detailed only when it implies our model goals.

In many works and products are used the cases for processes and workflow instances (see the van der Aalst works and papers). In most of processes the activity can be seen as a work (contract) that is required by someone (the principal) and executed by another one (the contractor). The communication protocol is illustrated in Figure 1.
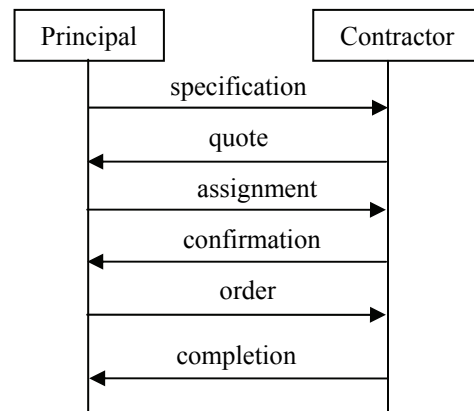

Figure 1: Communications protocol

2.3. Managing Processes

A management system can manage several systems, and in doing so, it ensures the ability of the managed systems to communicate with one another and with the outside world, that is, the managed system at a higher level. Between the management system and the managed system there occurs an exchange of information. This enables the management system to communicate objectives, preconditions and decisions to the managed system and the managed system conversely reports back to the management system.

Based upon these reports, the management system may revise the objectives, preconditions, and decisions. Process management has long been divided into four levels. The distinction between these is based upon the frequency and scope of the decisions to be made. By scope, we mean two things: the period of time over which the decision has an influence and its potential financial impact. The four levels are as follows:

1. Real-time management.
2. Operational management.
3. Tactical management.
4. Strategic management.

Decision making is an important feature of (process) management. The following phases that are always passed through when solving decision problems: Definition (establishing exactly what the problem is and, in particular, within what scope a solution to it must be found), Creation (formulating one or more solutions that fall within the scope defined or satisfy an optimization criterion), Evaluation (assessing different solutions, for instance by multi-criteria analysis) and Selection (selecting one solution that works in order to implement it).

2.4. Information Systems for Business Processes

Workflow management it is one the main trend in our days. It takes the business processes out of the applications. A workflow system manages the workflows and organizes the routing of case data amongst the human resources and through application programs. Just as databases are developed and used with the assistance of a database management system, so workflow management systems (WFMS) can be used to define and use workflow systems.

3. MODELING WORKFLOWS

Workflow Concepts used are: case, activity, and task, work item (van der Aalst 2002). Their interdependencies are shown in Figure 2.
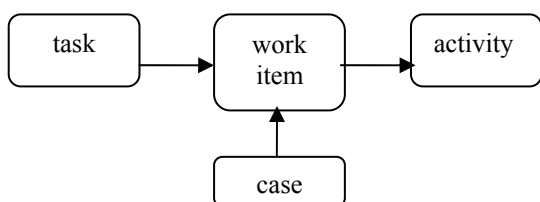

Figure 2: The relationship between the terms task, case, work item and activity

Based on the process definition we can specify:

- what tasks need to be performed for a particular category of case;
- the order in which they must be carried out.

The process definition does not indicate who should do it. The way in which the work items are allocated to resources (people and/or machines) is very important to the efficiency and effectiveness of the workflow.

A workflow system enables an organization to use and manage structured business processes. One important property of workflow systems is that, it becomes easier to change business processes. Exchanging or combining tasks or rearranging resource classes are easy modifications. Improvements influence performance criteria such as completion times, utilization of capacity, level of service and flexibility.

3.1. Business Process Re-engineering

Exist many tools for analyzing workflows that we not intend to detail here, but some specific problems like: soundness, reachability, performance analysis is of major importance (also these can be found in the works and papers of W.M.P. van der Aalst).

We shall consider the relationship between business process re-engineering (BPR) and workflow management. We can define BPR as the fundamental reconsideration of business processes. Its objective is to bring about entirely new business processes which enable drastic improvements to costs, quality and service. In order to achieve this objective, radical changes are often necessary.

Process-oriented thinking is crucial in the use of workflow management systems. Supporting old processes with a workflow system will only deliver a limited amount of improvement. Dramatic improvements are only possible if the old processes are separated from and replaced by new ones.

3.1.1. Workflow Failure Recovery

Workflow Management Systems more and more become the basic technology for organizations to perform their daily business processes and their workflows (Eder 1996). A consistent and reliable execution of such processes is crucial for all organizations. We claim that this can only be achieved by integrating transactional features especially workflow transactions into WFMSs.

Workflow recovery concepts are necessary for the reliable and consistent execution of business processes in the presence of failures and exceptions. In the Workflow Management Systems the workflow engine manages the workflow instance execution.

Many works and papers deal with workflow recovery that has connection with database transactions. Workflow technology targets supporting reliable and scaleable execution, for workflow management systems (WfMS) to support large-scale multi-system applications, involving both humans and legacy systems, in distributed and often heterogeneous environments (Luo 2000).

In case of failures, workflow processes usually need to resume their executions from one of their saved states, called a checkpoint, achieved by saving the states

from time to time persistently. The activity of restoring a checkpoint and resuming the execution from the checkpoint is called rollback. Those techniques have long been used in database systems. A checkpoint is an action consistent checkpoint if it represents a state between complete update operations. A consistent state in the database domain is a state when no update transactions were active. This checkpoint representing a consistent state is a transaction consistent checkpoint. A checkpoint does not need to satisfy any consistency constraints. But recovery after failure must always guarantee that the resultant state is transaction consistent even though any checkpoint used may not be. A checkpoint can be either local or global. A local checkpoint is a checkpoint taken locally, with or without cooperation with any other local check pointing activities at different sites. A local checkpoint can be a fuzzy or consistent checkpoint. During global reconstruction, a set of local checkpoints, usually taken at different site, will be used to find global consistent state. To facilitate the global reconstruction, a global checkpoint, derived from a set of local checkpoints taken at different site, provides a rollback boundary, thus reducing the recovery time.

Failure recovery is essential for transactional workflow management systems. When a failure occurs, compensation is usually used to rollback corresponding processes and ensures semantic atomicity. Previously, many mechanisms have been proposed while performance problem is rarely considered. This is a limitation since compensating activities are rather expensive.

3.1.2. Compensation and rollback

This can be treated as for the data base transactions. The treated way that a transactional scenario responds to failure differs from the way that non transactional scenarios respond to failure. When a non transactional scenario fails, the collaboration simply logs an error and terminates. When a transactional scenario fails, the scenario rolls back to leave data in a consistent state across all of the involved databases. When an error occurs, sub-transactions might have already caused applications to commit work. Therefore, rollback is done through the use of compensation steps, actions that counteract the effects of other actions. A compensation step executes only during rollback.

3.1.3. Fault Tolerance

Fault tolerance is the ability of an application to continue valid operation after the application, or part of it, fails in some way (Dialani 2002). The object data may return (rollback) to previous values if the current values are lost, and processes may return to a state in which a message is re-sent, if the previous attempt apparently failed.

In order to return to a previous consistent state, an application must record a replica of its previous state. The state of a process can be copied using a check pointing mechanism, or only the incremental changes to the process state using a logging mechanism and in every one of those two cases, it is possible to rollback to a previous valid state.

In distributed applications one process may fail without the other processes being aware of the failure. This can lead to the state of the application as a whole (the global state) being inconsistent.

An application is in a globally consistent state if whenever the receipt operation of a message has been recorded in the state of some process, then the send operation of that message must have been recorded also. Is needed a fault tolerance mechanism for distributed applications to keep an application to a consistent global state, or return to the last known consistent state (maximal state) in case of failure.

In figure 3, we present an overview of the fault tolerant system, as applied to the Web Services architecture.
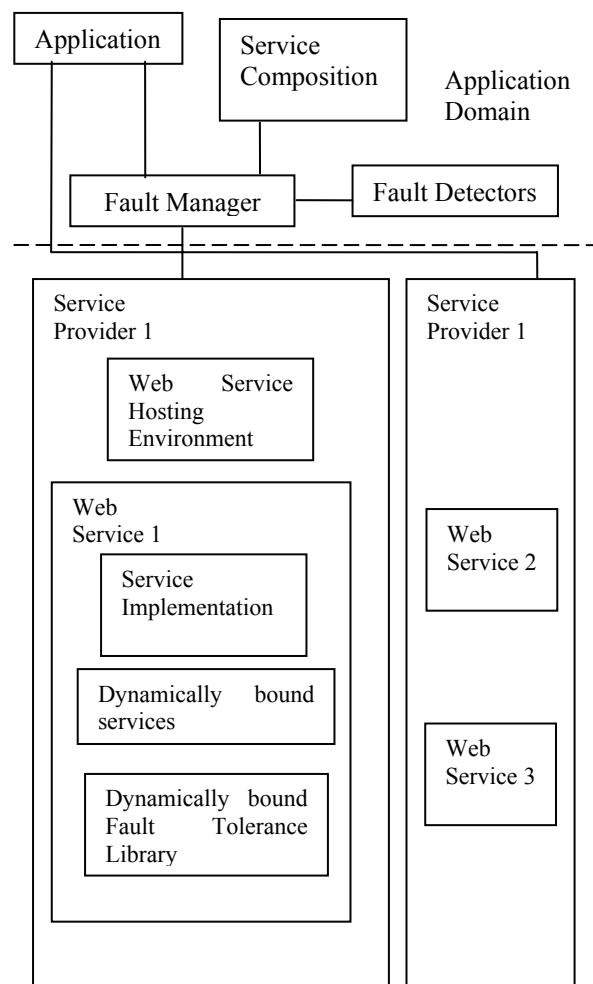

Figure 3: Architecture for Fault Tolerant Web Services

The frameworks for assuring the fault tolerance are various. For example, CORBA employs a two-pronged strategy to recover from a fault, namely the local recovery mechanism and the global recovery mechanism. The context of a local recovery is restricted to recovery of an individual service instance, while

global recovery applies to the entire application. The local recovery mechanism tries to revive the service instance with minimal or no intervention by the global recovery mechanism. A local recovery mechanism escalates the failure notification to the global recovery mechanism in case of its failure to recover the fault locally. The architecture imitates an hour glass model to restrict the dependency between the two layers to a minimal set of interfaces, for co-ordination between the two layers.

4. RESOURCES

The resource management in processes and in workflow was treated in many ways but only some of them will be referred in the following (Lerner 2000).

Many models for resource representation and management have particular specificities. Between those the (Lerner 2000) illustrates an interesting one and we believe a useful model. For resource model designing are needed specifications that they can be used as prescriptions for supporting the resource allocation and scheduling through the application of tools. It is needed to develop and evaluate resource specification formalisms that are sufficiently rigorous that they can be used to reason about complex activities from the processes and how to support these processes most efficiently. The lack of resources causes contention, occasions the need for some tasks to wait for others to complete, and generally slows down accomplishment of larger goals. Often potential delays can be avoided or reduced by using resource analysis to identify ways in which tasks can be made to execute in parallel that avoid resource contention.

In addition to the scheduling concerns, reasoning about resources requires an understanding of the similarities and differences among resources as well as the precise needs of the activities being coordinated. With this understanding it is possible to identify in which situations only a particular resource will suffice and in which situations any of a class of resources may be acceptable for the task. This identification of critical resource needs can also facilitate the identification of likely bottlenecks in the execution of a process or workflow.

The models require capabilities for describing and using major features that will allow emphasizing the resources and their possible interrelations (like similarity) and interactions in order that the execution of the activities can be done without major changes. From the different models we consider that the proposed model from (Lerner 2000) gives a resource model able to emphasize the main requirement for a complex model one. In the model are detailed many important features like:

- similarity between the resources that can be also source of major improvements in real systems;
- resource attributes (like measure, availability, etc.);

- relations and their types.

The resource management component provides information about the types and availability of resources and to track their usage. It also maintains a history of requests for resources that failed, and thus, evaluate the need for more resources of a particular type, is important in speeding up of an overall activity. The resource manager allows acquiring, reserve and releasing the resource under a centralized way.

4.1. Resource Management some Considerations

Resource management problems have almost the same underlying basic structure in common as timetabling problems.

Resource Allocations Modeling the consumption of Resources by Tasks is crucial to resource management and Rules are used in order to do it. The priorities and conflicts are also used widely in the resource management.

Tasks represent the actions to be performed during the execution of a Plan to achieve the set objectives. Tasks usually require certain Resources to be accomplished. The duration of some Tasks even depends on the amount of Resources assigned to them. Since there are Tasks that for example need to be carried out consecutively or concurrently, or with certain time-lags, interdependencies between them regarding their scheduling have to be modeled too.

Rules represent the constraints in models. They can be divided into two categories:

- Active Rules change models to ensure that they are met. Typical examples are the interdependencies between Tasks. An Active Rule could schedule one Task to start when another Task ends.
- Passive Rules are used to verify models and indicate violations, so it is possible to react immediately, like "no Employee should work more than 40 hours per week".

The used Rules are usually very specific to the domain and even to the undertaking to be modeled.

Priorities are used to express the amount of importance assigned to certain elements of models. A Task with a high Priority assigned to it, for example, is more important to be executed than another Task with a lower one. The numbers of used Priority Levels as well as their meanings are usually very specific to the respective domain or even to the undertaking.

The Passive Rules are used to indicate problems. Those indications are represented by Conflicts and need to be managed to be of use.

Additional Functions are obvious for a good resource management:

- Reporting. Throughout the realization of a Plan, the involved parties usually need to be informed about the aspects which are of

interest to them. Teachers, for example should receive a timetable different from the one for a certain class. So the application should support the creation of Reports intended for different audiences like to-do lists for the members of a team, financial reports for the accounting department, et cetera.

- Tracking. The term Tracking refers to the process of constantly collecting data about the actual course of the implementation. This is necessary to make post-implementation analyses possible, which can be a significant part of a project. For example, the profit has to be figured out or somebody needs to be found guilty for the failure of a project.

5. WFMS AND ERP

A WfMS is implemented based on a process specification and execution paradigm. Under a WfMS, a workflow model is first created to specify organizational business processes, and then workflow instances are created to carry out the actual steps described in the workflow model. During the workflow execution, the workflow instances can access legacy systems, databases, applications, and can interact with users.

ERP systems are implemented around the idea of prefabricated applications. To achieve better fit between the prefabricated applications and the needs of the organization, ERP systems must be configured by setting various application parameters. The more parameters an ERP application has, the more flexibility in configuring the business process. However, the workflow model in conventional ERP systems is not explicitly specified because it is embedded in the applications and the parameter tables.

For understanding the differences between WfMS and ERP systems is necessary make a distinction between flow logic and function logic (Cardoso 2004). In ERP systems, flow logic and function logic are both embedded in applications and parameter tables. In contrast, a WfMS separates the two explicitly. Flow logic is captured in a workflow model, usually graphically represented, and function logic is captured in the applications, data, and people the model invokes. Thus, a WfMS enable developers to separate the flows among a system's components (applications, data, and people) from the workflow model. Workflow systems are process-centric, focusing on the management of flow logic. On the other hand, ERP systems are data-centric, focusing on managing function logic via a common homogeneous data infrastructure across the organization to support multiple applications.

One of the problems is that ERP and WfMS technologies are often managed by different groups within an organization.

In (Cardoso 2004) are compared the two technologies along three main dimensions: domain scope (that defines the suitability of a system for a specific type of application or organization),

technological scope (characterizes the systems based on their technological capabilities) and system implementation. WfMS and ERP systems have been developed with distinct sets of technological capabilities. We can highlight these different capabilities by examining the different types of applications each system supports. Business process technology focuses its attention and effort on supporting three different types of applications:

1. workflows involving humans;
2. workflows involving systems and applications;
3. transactional workflows.

ERP systems constitute applications that focus on the integration of data. The objective of ERP applications is to provide an integrated solution to all business areas. The underpinning of shared data structures across many applications eliminates the need to pass data step-by-step among applications by accessing data from a common structure. The focus of ERP systems is mainly on structured data transactions. ERP modules operate directly with common interoperable databases to ensure consistent information for all purposes. This makes the manipulation of data easy. The ERP concept makes the strong assumption that data infrastructures are homogeneous across the organization, that is, the data is stored in interoperable databases and in some cases, the databases used are all from the same vendor.

WfMSs can address all three types of workflows (outlined above); however, they are most suitable for modeling workflows involving humans and software systems, especially if the systems are autonomous and heterogeneous. On the other side, ERP systems are more appropriate to model transactional workflows, which are data oriented.

In the following some considerations on WfMS and ERP are done. Both WfMS and ERP systems will play a major role in organizations' application integration. We presented above how WfMS and ERP systems can be used independently or together to address intra- and inter-organizational application integration.

The current directions in application integration point to the architectures shown in figure 4.

a)   Intra enterprise integration
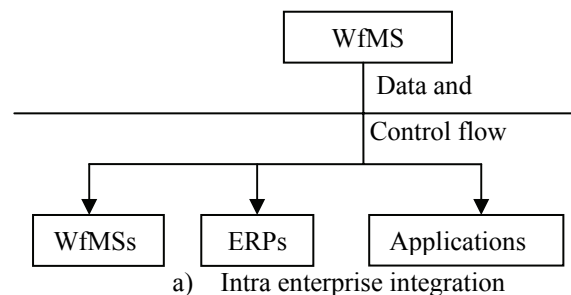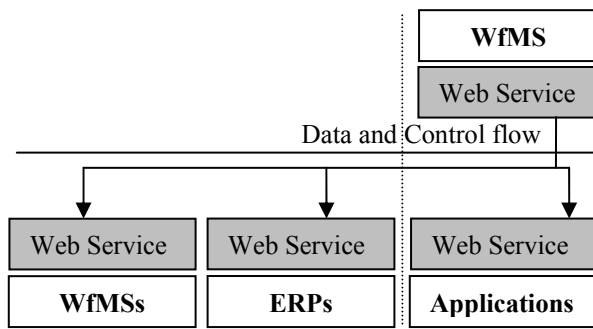
b) Inter-enterprise integration
Figure 4: Intra and inter-enterprise integration using WfMS and ERP systems.

In both intra- and inter-organizational integration, flow logic is being separated out and captured in WfMS with function logic found in ERP systems, other applications or other WfMS. In intra-organizational area, we see this represented in the moves by ERP vendors to integrate WfMS into their ERP architectures and emergence of EAI/BMP workflow based tools.

In some cases, system integration expertise may be needed to manually code and integrate with the WfMS special features such as to link the workflow engine to legacy applications, set transactional properties, define recovery procedures, etc. Additionally, the integration of access control and user rights on both the workflow as well as the applications requires additional effort in a heterogeneous environment (zur Muehlen 2000).

6. OUR CONSIDERATIONS
6.1. Resources
Due to the major importance of the resource and their utilization we intend to propose some particular framework for their representation and a support for their management. For the resource:

- characterization we define its: type (RType), identifier (RId), name (RName), its decomposability (RDe) as logical value, the minimal amount necessary for being usable (effective) (RMin), the current available Amount (RCAv), creation instant time (RCret), similar functionality with (RSF);
- requirement of usage every activity or task that uses it, must specify its: identifier (RId), the necessary amount in the case if it is decomposable (implicitly 1 if it is not) (RNec) and the estimated time interval for its usage (REs);
- used entities concerning their states and localizations in time for every entity that can be used must specify its identifier (RId), state (RSt) that must be allocated or active, the user activity or task that uses it (RUser), and starting time interval in this state (RStaT). The place of the resource is an optionally attribute that can (or not) be used in many specific systems where it is appropriated;

- unsolved requirement list that contains the activity (task) identification, the required amount and the time instant when it was required;
- history that for every entity contains every state in that the entity was;
- history of waiting's that contains the item from unsolved requirement list and the time that it waited.

Resource manager will execute some specific actions:

- whenever when a new resource appears an item in the characterization will be placed;
- when appears a supplementary amount the RCAv is increased and the RCreT is modified;
- every time when an activity or a task asks for a resource and if it exists in the required amount, it will be created an item with the state of the entity and will decrease with the taken amount the current available amount (RCAv) and put the item in the allocated state. If the amount required does not exists the activity (task) enters in the blocked state (waiting for the resource availability) and an item will be placed in the unsolved requirement list;
- whenever when the activity or task starts its execution, the of its appropriate item (RSt) is modified into active;
- every time when a resource is released by its user (activity or task), it must enter in the history an item and increase with the used amount the current available amount (RCAv);
- every time when the RCAv is modified (by increasing it) then the unsolved requirement list is analyzed and if it is possible then the requirement is satisfied and the item is placed in an history list for waiting;
- when an action or a task that required a resource and it was in the waiting state it can cancel its requirement. In this case the Resource Manager will remove it from unsolved requirement list and will put it in to history.

The generic resource manager functions are given in the figure 5. As it can be seen from the above descriptions, the resource description is a complex one. Also it must contain some similarities between the resource functionalities (these are not detailed in the paper). The similar functionalities must be also qualified in performance. Here is not in detail specified the proposed multiagent model for the Generic Resource Manager, that will be given in another future paper. It must be able to have some intelligent agents that are able to learn from the history requirements for the resources.
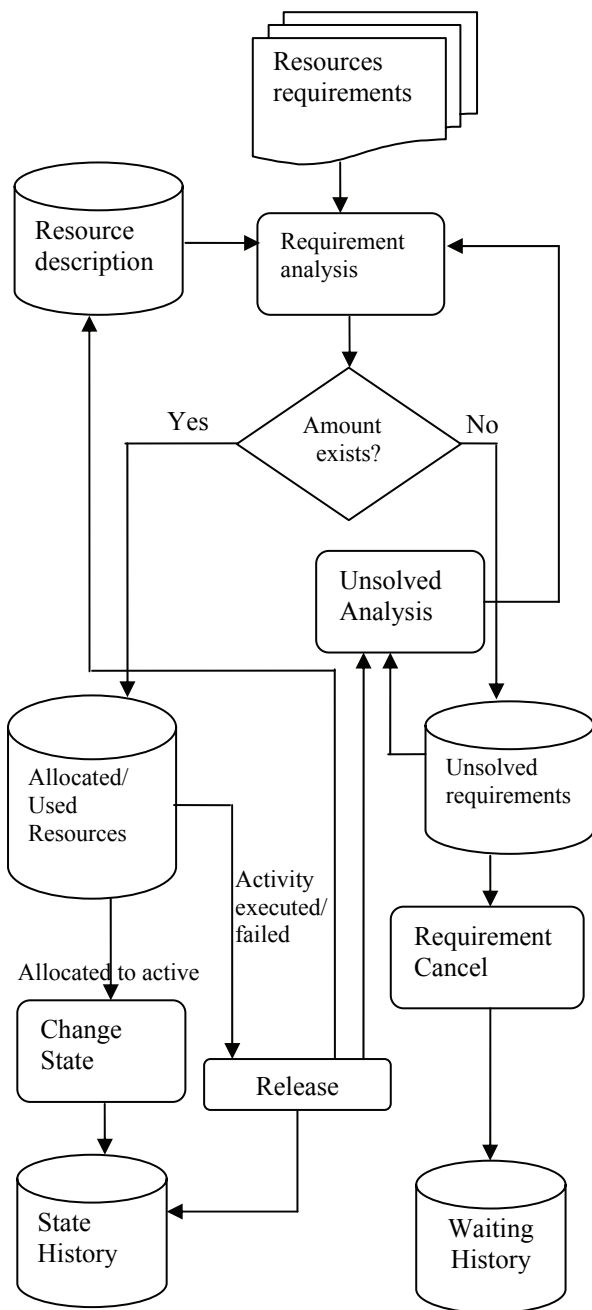
Figure 5: Generic Resource Manager Functions.

6.2. Reengineering failed instance workflows

An instance workflow is executed until the activity that failed and that generated the instance failure. The workflow engines that execute (run) the workflow instances interact with the system, the user application and in many cases with the user (human person). When an activity fails its workflow instance must be analyzed and must detect the reason of the failure.

Many products (like ACTIVE BPEL for BPEL4WS) use compensation and other works propose other solutions like worklets (Adams 2006) in order to solve the problem.

From proposed solutions in many works and papers and due to the fact that the human persons interact with and manage the systems and workflows,

we identified some requirements for achieving a better workflow management:

- (unique) identification of workflows, workflow instances and activities;
- the activities results (after their execution) can be useful or volatile (useless) in the case of the reply the failed workflow instance;
- during the workflow instance execution from its start to the failed activity the state of the real system resources and their states (and also the data bases transactions) were affected. In the case of a reply of the workflow instance the state of the system (and the appropriate data bases) and the resources were affected and all these must be restored to the state just before the failed activity.

In (Cicortas 2006), we proposed a multi-agent system (MAS) that is able to analyze the workflow instance that was blocked and based on its state to redesign the reminded part of the workflow instance that will be executed. This analysis must decide for every activity that its result can be or not can be used in the conditions that allow it, if we dispose for appropriate information. In a workflow, we have some activities whose result is in one of the following categories:

- the activity result can be retained and in future can be used, we can say it is "remanent";
- the activity result is volatile in future, it does not be used, if we want the activity result in the future, we must reply it;
- the activity result has affected the real system state and the system state must be restored in the case if we want to replay this activity we must remove this "effect".

If the workflow fails, then it fails due some activities that it failed. The previous Workflow activities were executed and based on their result as previous analysis was done, we can find the activities whose results are reusable and are remanent.

In the case of the reply of the Workflow we must make an analysis that: restores the activities that were affected in some way the system state, find the activities that have reusable and remanent results, reconstruct a new Workflow with the remaned activities and reusable results. (The problem is that on every parallel branch we must have a reusable result). The actions and their sequencing are the following:

- For every activity the designer must decide if it has remanent result and the way that allow to identify and storing it; also for the other activities it will specify their kind the result is volatile or the effect of the activity or the overall system, in the sense that it must be

discarded (example: updating data bases with the cancellation of the activity effects).

- The Workflow has a description of all its activities the above features.
- During the execution of a workflow instance it will record in a data base, the activities (their results if there is in the case of the remanence) as soon as these are executed.
- When a workflow fails then an analysis is done and the effects of activities on the system state are done; the last activities with remanent results will be retained. Here a mechanism for finding on every branch these activities is required.
- After the previous step when the workflow instance must be replied a new variant of the workflow is constructed: the input, a transition, and the locations of activities with remanent results.
- The new workflow instance can be start.

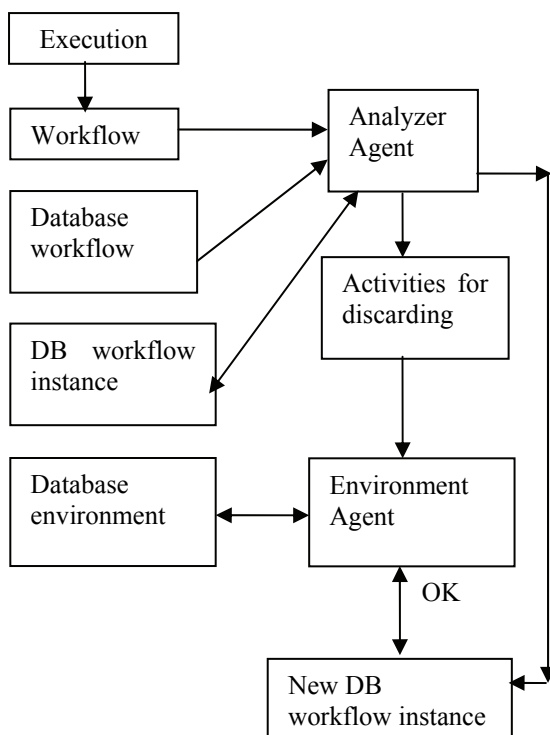The figure sketches out our proposal.

Figure 6: The flow of actions.

The workflow must be described with the above requirements and the system agents act as follows. In the case of a workflow instance failure the Analyzer Agent, make the analysis from that:

- all the activities on the every parallel branch are specified and their results are available;
- for all the activities that affect the overall system state prepares their cancellation actions;

- redesigns the new workflow instance.

The Environmental Agent executes the cancellation activities. The Effective Agent negotiates the conditions for the execution of the new instance of the workflow and launches it in execution. Decision for modeling is based on process complexity, failure of activities that can fail only the activity and "suspend" the process or fail all the process. The problem is to identify the reply of the process and its implications. Done to the process complexity the solution can be:

- restart the process from the beginning or
- reply the process from one of the previous "valid" states before its failure under some conditions.

Needs for replies in a failure case are:

- every activity must have specified its type concerning its result (reusable, affect the environment, insignificant);
- during the execution every activity recalls somewhere (in a data base) its result and this result can be used depending of activity type in the future (if the activity is of reusable type).

7. CONCLUSIONS AND FUTURE WORKS

The considerations form the paper will serve for designing the appropriate models:

- design a resource manager that is able to interact with the system, ERP, and workflow engine" in order that the resource requirements during of the execution of a workflow instance have a real support;
- Re-engineer a failed workflow instance such as the system and the workflow instance are able to (re)start their activities from a coherent and also consistent state.

For that, the communication between the Resource Manager and the real system (or the ERP) and workflow engines will be done using appropriate web services or in the simplest cases messages. The workflows and workflow instances will be described in appropriate and extended data bases so as was stated in (Cicortas 2006).

The failure of a workflow instance must be managed by a multi-agent system that:

- will discard the effects of the failure in the data base (so as in transactional data bases);
- will rebuild the rest of the workflow instance (that must be executed);
- will give to the workflow engine the new (rebuild) workflow instance.

REFERENCES

Adams, M., ter Hofstede, A.H.M., Edmond, D., van der Aalst, 2006. *W.M.P., Implementing Dynamic Flexibility in Workflows using Worklets*, BPM Center Report BPM-06-06, BPMcenter.org, 2006.

Cardoso, J., Bostrom, R.P., Sheth, A., 2004. *Workflow Management Systems and ERP Systems: Differences, Commonalities, and Applications, Information Technology and Management* 5, Kluwer Academic Publishers, pp. 319–338, 2004.

Cicortas, Al., Iordan, V., Fortis, Al., Fortis, F., 2006. Reengineering the Failed Workflows, *Annals of the Tiberiu Popoviciu Seminar, Supplement international Workshop in Collaborative Systems*, Vol. 4, ISSN 1584-4536, pp.57-66.

Dialani, V, Miles, L., De Roure, D., Luck, M., 2002. Transparent Fault Tolerance for Web Services based Architectures, *Eighth International 12 Europar Conference (EURO-PAR'02), Lecture Notes in Computer Science,* Padeborn, Germany, August 2002, Springer-Verlag, 2002.

Eder, J., Liebhart, W., 1996. Workflow recovery, *Proceedings of the First IFCIS International Conference on Cooperative Information Systems* (CoopIS'96) 1996.

Eriksson, B., Baharak, G. Fard , Grubesic, B.,2002. *Frameworks, an Introduction*, 2002.

Fayad, Mohamed E., Schmidt, Douglas C., Johnson, Ralph E., 1999. *Building application frameworks: object-oriented foundations of framework design.*, John Wiley & Sons, Inc., 1999.

Iordan, V., Cicortas, Al., 2007. *Multi-Agent Systems Used for Managing Competences*, The Proceedings of the 8th International Conference on Informatics in Economy, ASE Publishing House, 2007, Informatics in Knowledge Society, ISBN 978-973-594-921-1, Bucharest, pp. 633-638, 2007

Kearney, R.J., 2005. *Enhancing and Extending ERP Performance with an Automated Workflow System*, Workflow Handbook, Lighthouse Point, Florida, ed. Layna Fisher, pp.92-102, 2005.

Lerner, B. S., Ninan, A. G., Osterweil, L. J., Podorozhny, R. M., 2000. *Modeling and Managing Resource Utilization for Process*, Workflow and Activity Coordination, University of Massachusetts, TR: UM-CS-2000-058, 2000.

Luo, Z, 2000. Checkpointing for workflow recovery. *ACM Southeast Regional Conference,* pp. 79-80, 2000.

Sharp, A., McDermott, P., 2001. *Workflow Modeling - Tools for Process improvement and Application development*, Artech House.

van der Aalst, W., and van Hee, K., 2002. *Workflow Management, Models, Methods and Systems,* The MIT Press Cambridge, Massachusetts London, England.

WfMC 1994. *The Workflow Reference Model, Version* 1.1, Standard TC00-1003, Workflow Management Coalition, 1994.

WfMC 1996. *Workflow Management Coalition*, Interface 1: Process Definition Interchange. Draft 5.0, Issued on February 03, 1996, Workflow Management Coalition, 1996.

WfMC 1999. Terminology & Glossary, Standard TC-1011, Issue 3.0, *Workflow Management Coalition*, 1999.

WfMC 2001. *Workflow Handbook 2001*. Workflow Management Coalition, Future Strategies Inc., Lighthouse Point, Florida, USA, 2001.

zur Muehlen, M. , Allen, R., 2000. Embedded vs. autonomous workflow – putting paradigms into perspective, in: Excellence in Practice, Vol. IV, *Innovation and Excellence in Workflow and Knowledge Management*, ed. L. Fischer (Future Strategies, Lighthouse Point, FL, 2000) pp. 49–58.

zur Muehlen, M., 2004. Workflow-based Process Controlling: Foundation, Design and Application of Workflow-driven Process Information, *Advances in Information Systems and Management Science*, Logos Verlag Berlin, 2004.

AUTHORS BIOGRAPHY

Victoria Iordan is Associate Professor Ph.D. at Computer Science Department of Mathematics and Computer Science Faculty, West University of Timisoara. Fields of scientific interest: modeling, simulation, complex systems design, workflow modeling, numerical methods.

Alexandru Cicortas. Actually is Professor Ph.D. at Computer Science Department of Mathematics and Computer Science Faculty, West University of Timisoara. Main of his research and papers concerns: the Discrete Event Systems (DEvS) modeling in distributed architectures; Petri Nets and component modeling are another domain of interest; complex systems design; workflow modeling (he is member of the European VISP project).

FORMAL METHOD OF FUNCTIONAL MODEL BUILDING BASED ON GRAPH TRANSFORMATIONS

Janis Grundspenkis[a], Ieva Zeltmate [b]

[a][b] Riga Technical University, Faculty of Computer Science and Information Technology, Department of Systems Theory and Design

[a] janis.grundspenkis@cs.rtu.lv, [b] ieva.zeltmate@cs.rtu.lv

ABSTRACT
Structural modelling approach has been developed to support systematic, domain model–based knowledge acquisition for model–based diagnosis problem solving. Structural models capture declarative, deep knowledge about system morphology and operation in normal conditions as well under faults. Models of morphological and functional structure are built mainly using knowledge of problem domain experts. Formal methods and algorithms for building model of a morphological structure and its transformation into a model of a functional structure have been developed to reduce the workload of human experts. The paper focuses on model transformation issues and represents new results allowing to extend previously developed method of formal transformations. Analysis of different cases of logic of input and output flows of system components is carried out and corresponding transformations defined. The developed formal method is demonstrated for functional model building of cooling system of internal combustion engine.

Keywords: Structural modelling, reasoning, model transformations, decision making

1. INTRODUCTION
Complex technical systems often have different types of components: software and hardware including electronic, mechanical, hydraulic, pneumatic, electric, and /or even chemical (Kurki 1995). A plethora of different approaches has been developed for modelling of complex technical systems with the purpose of their diagnosis. At the same time many methods and models are not suitable for complex technical systems with physically heterogeneous elements where various physical processes (exchange of information, energy, material flows, etc.) proceed because it is impossible to build mathematical (analytical) models for a system as a whole. Alternative approaches try to overcome this drawback proposing, for example, to build general qualitative models (Fang 1994, Grundspenkis 1997, Harandi and Lange 1990) or to integrate multiple model types (Abu-Hanna and Jansweier 1994, Grundspenkis and Jekabsons 2001). Among these approaches only few support some kind of reasoning. The goal of

structural modelling (Grundspenkis 1997, Grundspenkis 1999) is to develop a framework for knowledge acquisition, representation and processing based on a set of related structural models. The advantages of structural models are their ability to capture both shallow (expert's experience) and deep knowledge (cause–consequence relationships between components), as well as to support inference of a new knowledge for decision making about systems operation (Grundspenkis 2004).

As a rule, knowledge acquisition is a bottleneck when knowledge–based systems are developed. Working with problem domain experts is time consuming and interpretation of obtained results isn't unambiguous. Structural modelling supports systematic domain model based knowledge acquisition. That allows to understand the structure and operations of system under consideration (Grundspenkis 1999). Moreover, already developed formal methods and algorithms allow to build topology of both models, namely a model of a morphological structure (MSM) and a model of a functional structure (FSM). Thus experts only need to define primitives of MSM (details are given in the next section) and a model is generated automatically. Transformation of a MSM into a FSM allows keeping consistency between models which often isn't possible when each model is built from scratch. The main idea of model transformations is based on formal method proposed in graph theory for undirected graphs. The essence of method is as follows: arcs of an initial graph (also called a node graph) are replaced by nodes of so called line graph (Tutte 2001). In structural modelling this method has been modified for directed graphs.

The paper presents a new transformation method in which logic of input and output flows of components of MSM are taken into account. The paper is organized as follows. Due to the fact that the paper pretty much is based on previous works brief descriptions of basic notions of MSM and FSM are given in sections 2 and 3, respectively. The section four is devoted to the analysis of different cases of logic of input and output flows. In result corresponding transformations are defined. An example which demonstrates the proposed

transformation method is given, too. Conclusions and a short outline of future work are given in the last section.

2. THE BASIC NOTIONS OF MSM

Structural modelling approach and its applications are described in details in several papers (Grundspenkis 1997, Grundspenkis 1999, Grundspenkis 2002, Grundspenkis 2004). Due to the focus of the paper in this and in the next section only basic notations of models of morphological and functional structures are given. The first step of model building within the framework of structural modelling is identification of all available knowledge sources (human experts, handbooks, reports, graphics, etc.). The acquired knowledge is captured into small, independent, composable and decomposable units of knowledge which are visualized as objects (see Figure 1). Thus, objects are basic units of abstract model of MSM. These primitives have input and output contacts. If interpreted in an application domain, abstract objects correspond to the components of the given system, and contacts represent their inputs and outputs.
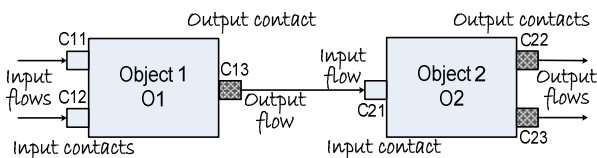

Figure 1: Representation of basic units of MSM

The connection of one object's output to another object's input is the only way how components can interact. Interactions are called flows. Human experts must list all components of a system, define all contacts and related flows. A topology of MSM is determined automatically by the *Automated Structural MOdelling System* (ASMOS) which maintains consistency between different levels of granularity of MSM by applying continuous mapping of models (Grundspenkis 1997).

The MSM supports so called structural reasoning that is based on the analysis of direct and indirect connectivity of objects. A direct connectivity is described by structural equation of the following form:

$flow_1 (IN) LOGOP flow_2 (IN) LOGOP … LOGOP flow_N (IN) = flow_1 (OUT) LOGOP flow_2 (OUT) LOGOP … LOGOP flow_M (OUT).$

Here LOGOP stands for logical operations AND or OR which define logic of flows. At the input side AND means that all connected flows are needed to get output, while OR means that only one flow is needed. At the output side AND means that all flows are produced by a component, but OR means that each output flow is produced in one operation mode. Thus, in case if objects are represented as graph nodes they have AND/AND, OR/AND, OR/OR, and AND/OR logic as it is shown in Figure 2. More complicated cases are discussed later.
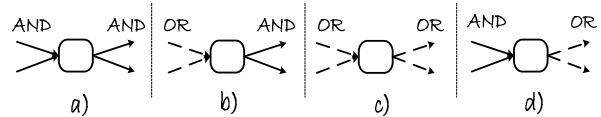

Figure 2: Logic of graph nodes

Composition of structural equations supports exploration of indirect connectivity and made possible conclusions about conditions of flow existence between inputs and outputs of different objects. In case of indirect connectivity the purpose of structural reasoning is twofold. First, it provides knowledge for decision making about requirements for the maintenance of designed operation modes. Second, results of structural reasoning are used to evaluate the degree of structural importance and the rank dispersion that support decision making about functionally overloaded components (Grundspenkis 1999).

3. THE BASIC NOTIONS OF FSM

To support reasoning about processes proceeding in physical components a model of functional structure (FSM) is used which encodes knowledge about functions of normally operating system, i.e., when a system under analysis has no faults. If visualized, the basic units of FSM are displayed as a diagram (see Figure 3).
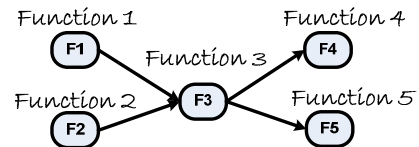

Figure 3: Representation of basic units of FSM

Expert's understanding of how a system works is organized as a representation that describes purpose why particular function is accomplished. More precisely, a function captures the intended purpose of an object (component), that is, a function specifies what a response is to a given stimulus, or, in other words, what is the result of proper behavior (functioning) of a given component. For example, the function of a water pump is "to provide a water flow under pressure". It is the purpose why a water pump is included in the system. This function in common with all others is represented as a node of FSM. Arcs of FSM represent cause–consequence relations between functions. Two notions – ""behavior" and "behavior state" of a contact of a given component are used to represent functions. Behavior is a characteristic of an input/output relation that explains how structure insures the fulfillment of functional specifications. In other words, behavior specifies an action that a component performs upon its "substance", i.e., it specifies how a reaction to a given stimulus is achieved. To provide a water flow under pressure, a water pump's behavior must be "pumping". Behavior state, in its turn, specify how flows "act" in corresponding inlets and outlets of components. Thus, each object described in MSM has its own behavior and

object's contacts are mapped on corresponding behavior states. To conclude the description of basic notions used in a FSM, lets state that the behavior states at input and output contacts of a water pump are described as "water flows trough the inlet of water pump", and "water flows through the outlet of water pump", respectively. Analysis of FSM supports decision making about processes going on in physical components of the system.

4. TRANSFORMATION OF MSM TO A FSM

In the structural modelling framework the basic ideas of transformation of MSM to a FSM have been described in (Grundspenkis 1997). A topology of FSM is derived from a MSM using transformation of a node graph which represents a MSM into a corresponding line graph which represents a FSM. Actually this transformation differs from transformation described in graph theory (Tutte 2001). In structural modelling the following constraint on relationships between nodes of a line graph is defined: there is a relationship between nodes of a line graph if and only if the corresponding arcs of a node graph have the same direction. The difference between transformations defined in graph theory and in structural modelling is shown in Figure 4. A line graph in case *b* corresponds to the graph theory definition, but a line graph in case *c* corresponds to the FSM definition.

a) a node graph   b) a line graph (graph theory case)   c) a line graph (FSM case)
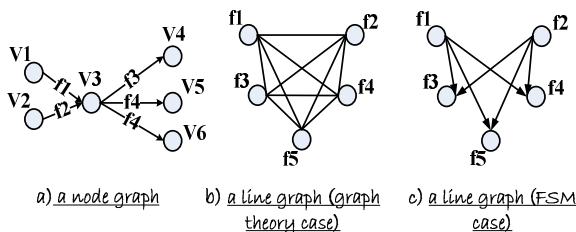
Figure 4: An example of transformation

The defined constraint guarantees correct cause–consequence relationships between nodes of FSM which represent functions. It is worth to stress that automation of the model transformation significantly decreases the workload of experts because they are asked only to define semantics of nodes of FSM, not to build a FSM from scratch.

The rest of the paper is devoted to detailed analysis of all cases of transformations taking into consideration logic of flows of MSM. Specific graphical representations are used to represent the logical relations (see Figure 5).

In our approach square on the flow and symbol ',' atop of the flow are used to display the logical component AND. For example, see the first case in Figure 5 – the square on the flow and symbol ',' (between the members of the expression) atop of the flow specifies that F2 and F3 are connected with the logical component AND. Triangle and symbol ';' are used to display the logical component OR. For example, see the fourth case in the Figure 5. Small line (see

seventh case) and symbols '(' and ')' are used to display the brackets in the logical expression.
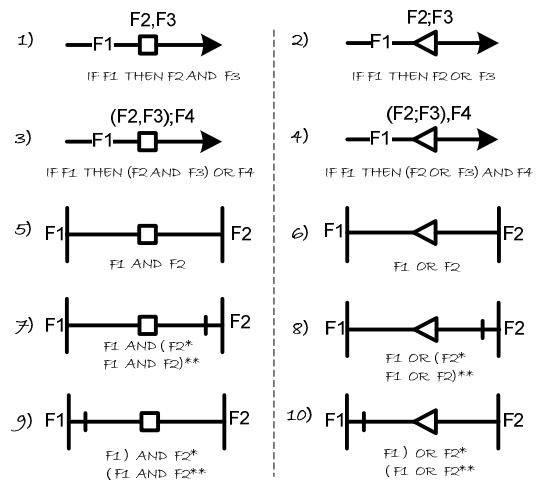
* the case when before or after the viewed function are another function
** the case when before or after the viewed function there aren't another function

Figure 5: Graphical representations of logical relations

First four cases shown in Figure 5 demonstrate situations when one flow at the input side of an object affects many flows at the output. All other cases (from 5 to 10) specify situations when many flows at the input affect one or many flows at the output. The logical expression atop of the flow can be used in all cases.

Theoretically four basic combinations of input and output flows exist:

1. One flow at the object's input side and one at the output side.
2. One flow at the object's input side and many at the output side.
3. Many flows at the object's input side and one at the output side.
4. Many flows at the object's input side and many at the output side.

The first combination is simple and no specific notations are used to understand the logical cause–consequence relationship in a MSM and after transformation (see Figure 6). The logical expression for the case (a) is as follows: *IF F1 THEN F2*. One flow F1 is related to exactly one flow F2.
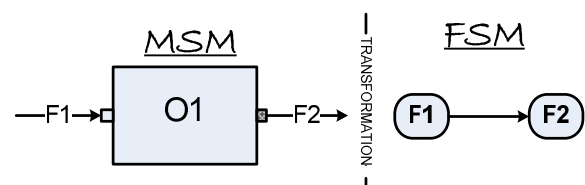

Figure 6: MSM and the corresponding FSM, case (a)

The second combination can be divided in 3 different cases. The first case is when one flow is at the input side and many flows linked with the logical operation AND are at the output side (see Figure 7).

The logical expression for the case (b) is as follows: *IF F1 THEN F2 AND F3*. One flow F1 creates more that one flow simultaneously - F2 and F3.
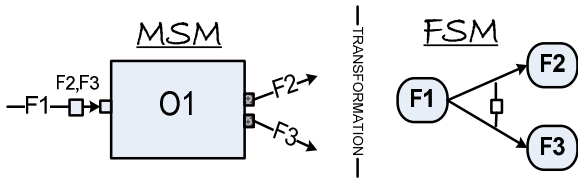

Figure 7: MSM and the corresponding FSM, case (b)

The second case differs from the first one only with the logical operation. One flow is at the input side and many flows linked with the logical operation OR are at the output side (see Figure 8).
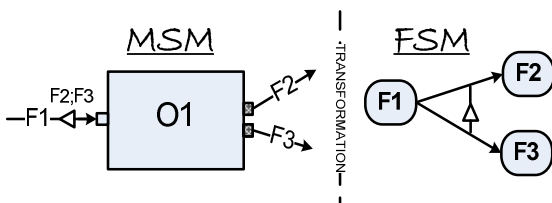

Figure 8: MSM and the corresponding FSM, case (c)

The logical expression for the case (c) is the following: *IF F1 THEN F2 OR F3*. One flow F1 can create one of two possible flows F2 or F3.

The third case is when one flow is at the input side and many flows linked with the logical operation AND/OR combinations are at the output side (see Figure 9). The logical expression for the case (d) is as follows: *IF F1 THEN (F2 OR F3) AND F4*. One flow F1 can create two flows simultaneously: flows F2 and F4 or in other mode flows F3 and F4.
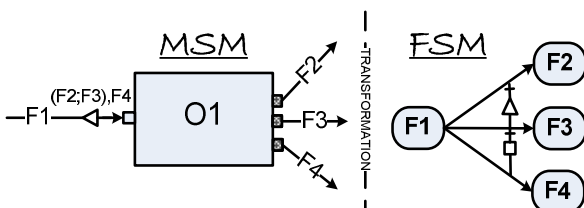

Figure 9: MSM and the corresponding FSM, case (d)

The third combination when many flows at the object input affect one flow at the output also can be divided in three different cases:

1. Many flows linked with the logical operation AND at the input side and one flow at the output side (see Figure 10).
2. Many flows linked with the logical operation OR at the input side and one flow at the output side (see Figure 11).
3. Many flows linked with the logical operation AND/OR combinations at the input side and one flow at the output side (see Figure 12).

The logical expression for the case (e) is the following: *IF F1 AND F2 AND F3 THEN F4*. Three flows F1 and F2 and F3 simultaneously can create exactly one flow F4.
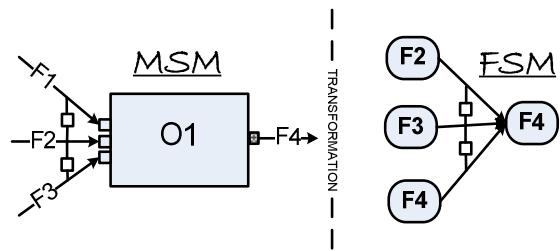

Figure 10: MSM and the corresponding FSM, case (e)

The logical expression for the case (f) is as follows: *IF F1 OR F2 OR F3 THEN F4*. One of the flows F1 or F2 or F3 in a fixed time can create one flow F4.

Figure 11: MSM and the corresponding FSM, case (f)

The logical expression for the case (g) is as follows: *IF F1 AND F2 AND F3 THEN F4*. Two flow combinations F1 and F3 or F2 and F3 can create one flow F4.
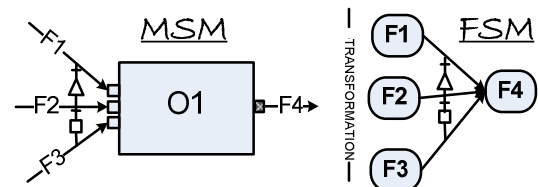
Figure 12: MSM and the corresponding FSM, case (g)

The fourth and the last combination of flows is the most complex and can be divided in many different cases, but we consider only six typical ones. The situation when many flows are linked with some logical operation at the output and all output flows are input of one object is not considered. In this combination the logical expression that can be obtained from a MSM after transformation to a FSM is decomposed in smaller rules. The first case is when many flows are linked with the logical operation AND at the input side and many flows are linked with the logical operation AND at the output side (see Figure 13).

The logical expression for the case (h) in the MSM and the FSM is: *IF F1 AND F2 AND F3 THEN F4 AND F5 AND F6*. Three flows F1 and F2 and F3 simultaneously create three flows F4 and F5 and F6.
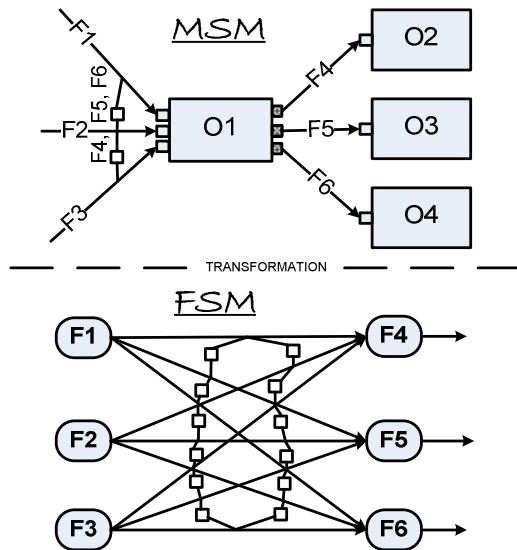
Figure 13: MSM and the corresponding FSM, case (h)

For the second case many flows are linked with the logical operation OR at the input side and many flows are linked with the logical operation OR at the output side (see Figure 14).

Figure 14: MSM and the corresponding FSM, case (i)

The logical expression for the case (i) in the MSM is: *IF F1 OR F2 OR F3 THEN F4 OR F5 OR F6.* One of the three flows F1 or F2 or F3 in a fixed time can create one of the three flows F4 or F5 or F6. The logical expressions in the FSM are:

- IF F1 THEN F4 OR F5 OR F6
- IF F2 THEN F4 OR F5 OR F6
- IF F3 THEN F4 OR F5 OR F6

or equally
- IF F1 OR F2 OR F3 THEN F4
- IF F1 OR F2 OR F3 THEN F5
- IF F1 OR F2 OR F3 THEN F6

The third case is when many flows are linked with the logical operation AND at the input side and many flows are linked with the logical operation OR at the output side (see Figure 15).
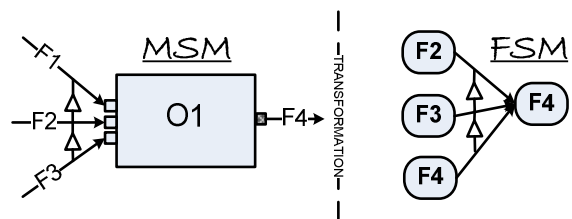

Figure 15: MSM and the corresponding FSM, case (j)

The logical expression for the case (j) in MSM is: *IF F1 AND F2 AND F3 THEN F4 OR F5 OR F6.* Three flows F1 and F2 and F3 simultaneously can create one of the three flows F4 or F5 or F6. The logical expressions in the FSM are:

- IF F1 AND F2 AND F3 THEN F4
- IF F1 AND F2 AND F3 THEN F5
- IF F1 AND F2 AND F3 THEN F6

For the fourth case many flows are linked with the logical operation OR at the input side and many flows are linked with the logical operation AND at the output side (see Figure 16).

Figure 16: MSM and the corresponding FSM, case (k)

The logical expression for the case (k) in the MSM is: *IF F1 OR F2 OR F3 THEN F4 AND F5 AND F6.* One of the three flows F1 or F2 or F3 in a fixed time can create all three flows F4 and F5 and F6. The logical expressions in the FSM are:

- IF F1 THEN F4 AND F5 AND F6
- IF F2 THEN F4 AND F5 AND F6
- IF F3 THEN F4 AND F5 AND F6

The fifth case is when many flows are linked with the logical operation AND/OR combinations at the input side and many flows are linked with the logical operation AND/OR combinations at the output side (see Figure 17).
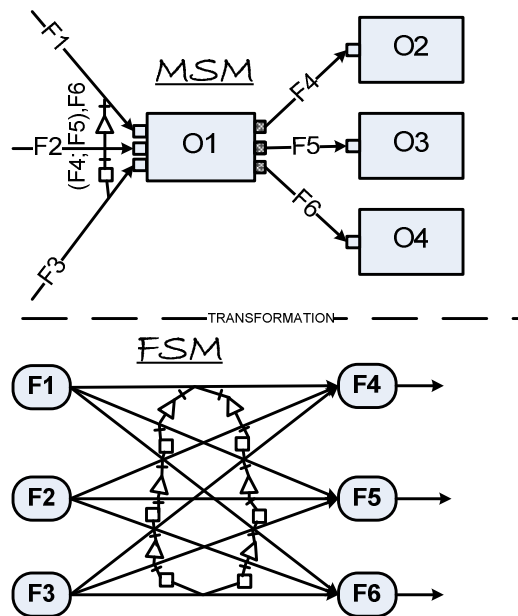
Figure 17: MSM and the corresponding FSM, case (l)

The logical expression for the case (l) in the MSM is: *IF (F1 OR F2) AND F3 THEN (F4 OR F5) AND F6.* Two flow combinations F1 and F3 or F2 and F3 can create one of the two flow combinations F4 AND F6 or F5 AND F6. The logical expressions in the FSM are:

- IF F1 AND F3 THEN (F4 OR F5) AND F6
- IF F2 AND F3 THEN (F4 OR F5) AND F6
- IF (F1 OR F2) AND F3 THEN F4 AND F6
- IF (F1 OR F2) AND F3 THEN F5 AND F6

The sixth case is the combination of the considered cases (from a to l) and for each of them in the MSM one rule can be acquired. As an example (see Figure 18) is displayed combination of cases (a) and (j). Many flows (F1, F2, and F3) are linked with the logical operation AND at the input side and affect many flows that are

linked with the logical operation OR (F4, F5, F6) at the output. Besides, only the flow F1 affects the flow F7.

The logical expressions for the case (m) in the MSM are two. The first is: *IF F1 AND F2 AND F3 THEN F4 OR F5 OR F6.* Three flows F1 and F2 and F3 simultaneou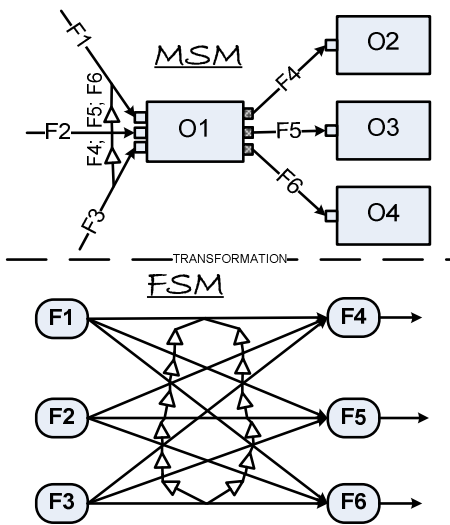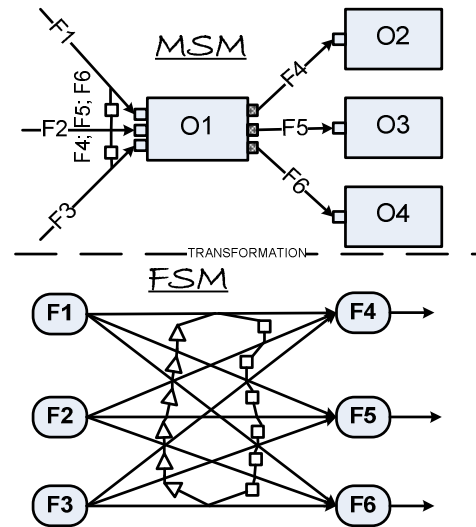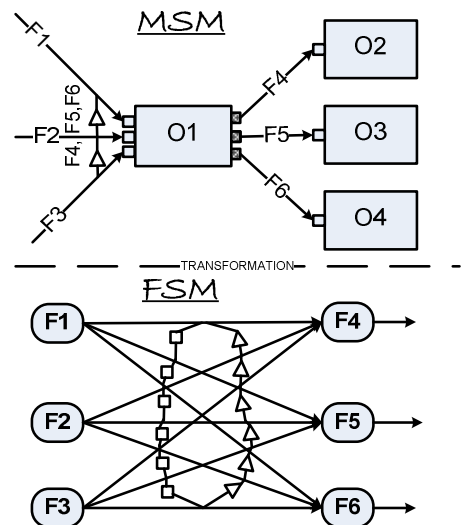sly can create one of the three flows F4 or F5 or F6. The second rule is: *IF F1 THEN F7.* The flow F1 creates also the flow F7. Corresponding logical expressions in the FSM are: IF F1 THEN F7 and all three described expressions that are used for the case (j) within model of functional structure.
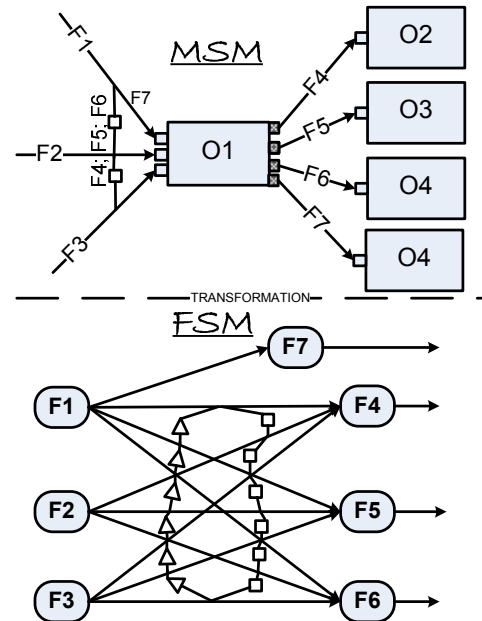
Figure 18: MSM and the corresponding FSM, case (m)

For better understanding of the defined transformations an example of a cooling system of an internal combustion engine (Grundspenkis 1999) is given. First, the model of a morphological structure is represented (see Figure 19).

For model of a morphological structure seven rules that correspond to the listed flow combination cases are defined:

1. IF F1 AND F13 THEN F2 (case e)
2. IF F2 THEN F3 OR F4 (case c)
3. IF F3 THEN F5 (case a)
4. IF F5 AND F8 THEN F6 AND F14 (case h)
5. IF F6 THEN F7 (case a)
6. IF F10 AND F11 AND F12 THEN F8 (case e)
7. IF F9 AND (F7 OR F4) THEN F1 (case g)

After transformations the model of functional structure of a cooling system (see Figure 20) is obtained.
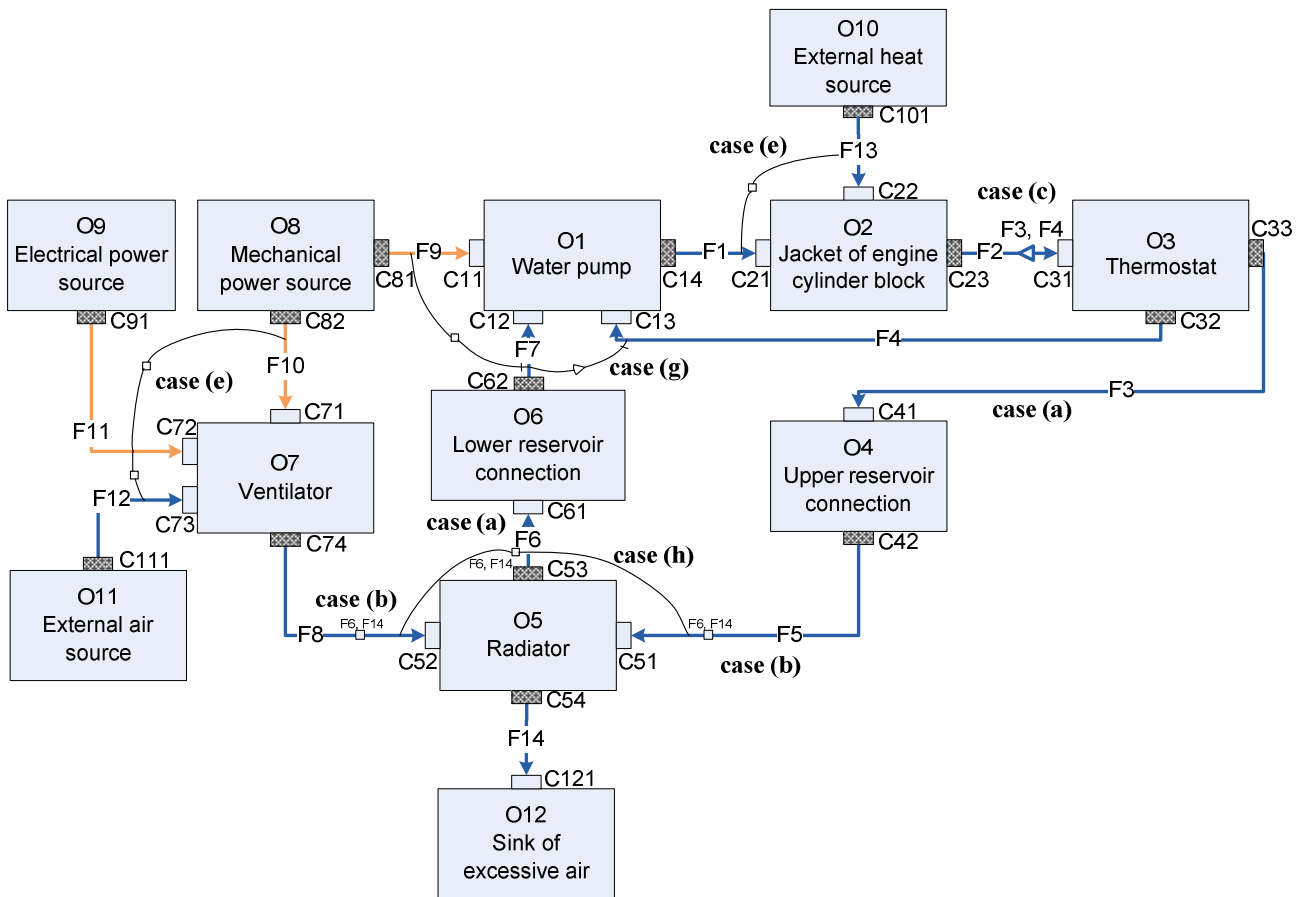
Figure 19: The morphological structure of a cooling system
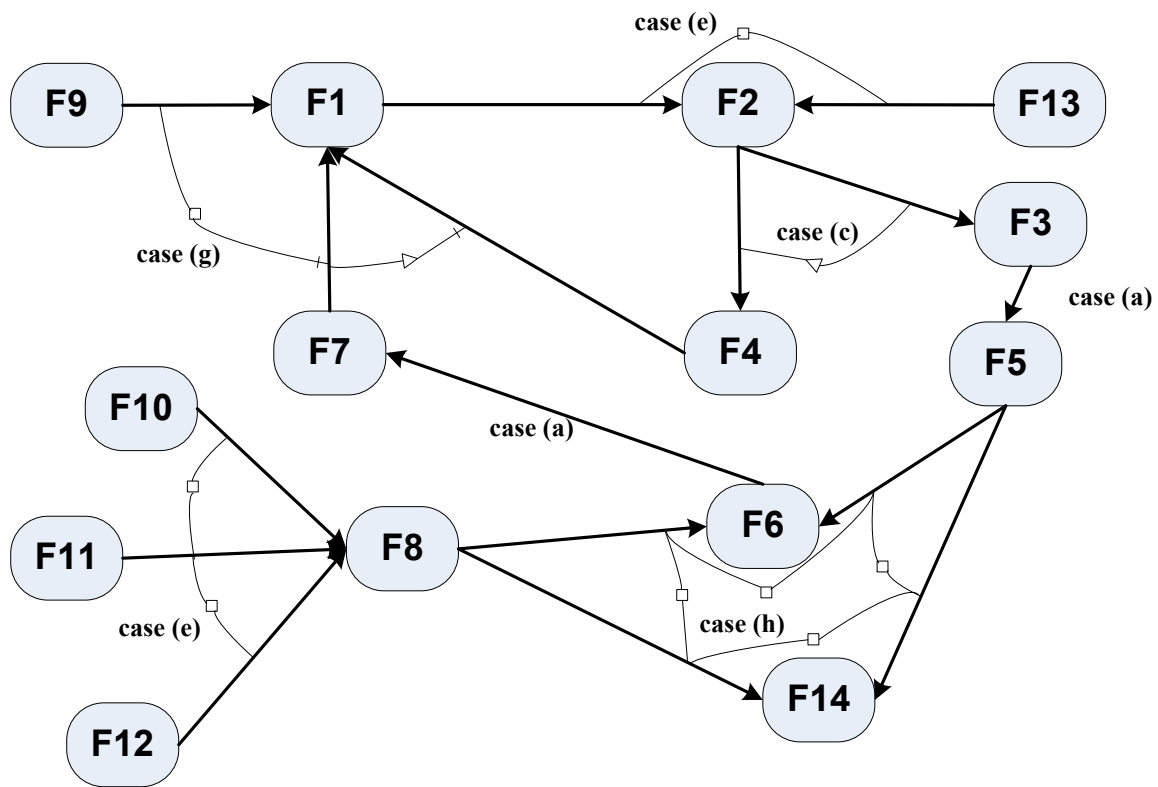
Figure 20: The functional structure of a cooling system

In the represented model of the functional structure seven rules that correspond to the listed flow combination cases can be defined:

1. IF F9 AND (F7 OR F4) THEN F1
2. IF F1 AND F13 THEN F2
3. IF F2 THEN F3 OR F4
4. IF F3 THEN F5
5. IF F5 AND F8 THEN F6 AND F14
6. IF F6 THEN F7
7. IF F10 AND F11 AND F12 THEN F8

It is quite obvious that the defined rules easily can be implemented in some expert system shell providing reasoning about relationships of components and/ or cause–consequence relationships of the system.

5. CONCLUSIONS AND FUTURE WORK

In structural modelling framework significant efforts are made to reduce a workload of human experts during the model building process. For this purpose the methods and algorithms for automatic generation of topologies of structural models have been developed and implemented. The paper describes different cases of logic of input and output flows of system components and defines corresponding transformations. The new results allow extending already developed formal method of transformation of a model of morphological structure into a model of functional structure. For all cases of transformation the corresponding logical expressions given in form of IF … THEN… rules are described. This provides the possibilities to reason and to make decisions about the structure and operation of the system as a whole. The future work is connected with the implementation of the proposed modified algorithm of formal transformations.

REFERENCES

Abu-Hanna A. and Jansweier W. N. H., 1994. Modelling Domain Knowledge Using Explicit Conceptualization. *IEEE-Expert 9(5)*, pp. 53-64.

Fang M., 1994. MFM Model-Based Diagnosis and Implementation. *Report 94–ID–712*, Institute of Automatic Control systems, Technical University of Denmark, Lungby

Grundspenkis J., 1997. Structural Modelling of Complex Technical Systems in Conditions of Incomplete Information: A Review. In: *Modern Aspects of Management Science, No 1*. Riga, Latvia, pp. 111-135.

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. and Jekabsons J., 2001. Model Transformation for Knowledge Base Integration within the Framework of Structural Modelling. In: Databases and Information Systems. *Fourth International Baltic Workshop, Baltic DB&IS 2002,* Vilnius, Lithuania, May 1-5, 2000, Selected Papers, (J. Barzdins and A. Caplinskas, Eds.). Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 261-274.

Grundspenkis J., 2002. Reasoning in Structural Model–Based Diagnosis. In: *Proceedings of 4th International Conference on Quality, Reliability and Maintenance, QRM 2002*, *Oxford,* March 21-22, (McNulty G.J., ed.), *Professional Engineering Publicity,* London, UK., pp. 295-298.

Grundspenkis J., 2004. Automated Transformation of the Functional Model into the Diagnosis Knowledge Base. In: *Proceedings of the 5th International Conference on Quality, Reliability and Maintenance*, *QRM 2004, Oxford,* 1-2 April, 2004, (McNulty G.J., ed.), *Professional Engineering Publishing*, London, UK, 2004, pp. 295-298.

Harandi M. T. and Lange R., 1990. Model-Based Knowledge Acquisition, In: Adeli, H. ed., *Knowledge Engineering, Vol. 1*, McGraw Hill, New York, pp. 103 – 129.

Kurki M., 1995. Model-Based Fault Diagnosis for Mechatronic Systems. Thesis (PhD), Technical Research Centre of Finland, VTT Publications 223, Espoo.

Tutte W.T., 2001. Graph Theory, Cambridge University Press, New York.

AUTHORS BIOGRAPHY

Janis Grundspenkis is a professor at Riga Technical University, the dean of the Faculty of Computer Science and Information Technology, the director of the Institute of Applied Computed Systems, and the head of the Department of Systems Theory and Design. He received his Dr.sc.ing. degree in 1972, Dr.habil.sc.ing. degree in 1993 both from Riga Technical University. His research interests are agent technologies, knowledge engineering and management, and structural modelling. He is a member of Institute of Electrical and Electronics Engineers (IEEE), and Association for Computing Machinery (ACM). He is a full member of Latvian Academy of Science.

Ieva Zeltmate is the system analyst and PhD student at the Department of Systems Theory and Design of Riga Technical University. Her scientific interests are associated with system analysis, design and development of knowledge representation systems and systems for structural modelling.

ANALYSIS OF LOGISTICS IN SUPPORT OF A HUMAN LUNAR OUTPOST

William Cirillo[a], Kevin Earle[b], Kandyce Goodliff[c], J. D. Reeves[d]
Mark Andraschko[e], R. Gabe Merrill[f], Chel Stromgren[g]

[a, b, c, d]NASA Langley Research Center, Mail Stop 462, Hampton, VA 23681
[e, f]Analytical Mechanics Associates, Inc., 303 Butler Farm Drive, Suite 104A, Hampton, VA 23666
[g]Science Applications International Corporation, 1710 SAIC Drive, McLean, VA 22102

[a]william.m.cirillo@nasa.gov, [b]kevin.d.earle@nasa.gov, [c]kandyce.e.goodliff@nasa.gov, [d]john.d.reeves@nasa.gov,
[e]andraschko@ama-inc.com, [f]merrill@ama-inc.com, [g]chel.stromgren@saic.com

ABSTRACT
Strategic level analysis of the integrated behavior of lunar transportation system and lunar surface system architecture options is performed to inform NASA Constellation Program senior management on the benefit, viability, affordability, and robustness of system design choices. This paper presents an overview of the approach used to perform the strategic-level analysis, with an emphasis on the logistics modeling and the impacts of logistics resupply on system behavior. An overview of deterministic and probabilistic analysis approaches is provided, with a discussion of the importance of each approach to understanding the integrated system behavior. The logistics required to support lunar surface habitation are analyzed from both "macro-logistics" and "micro-logistics" perspectives, where macro-logistics focuses on the delivery of goods to a destination and micro-logistics focuses on local handling of re-supply goods at a destination. An example lunar exploration system scenario is provided to tie the theories of strategic analysis to results generation capabilities.

Keywords: strategic analysis, space logistics, lunar outpost, lunar architecture

1. INTRODUCTION
Over the past three years, an interest in establishing an extended human presence on the Moon has been rekindled by the world's space fairing nations. NASA's contribution to this exploration renewal was first formulated in the Exploration Systems Architecture Study (ESAS Final Report 2005) and subsequently refined through ongoing analysis (Cooke 2006; Cooke 2007), which taken together provides the basis of the current NASA Space Exploration Program. The overall value of a human lunar return and subsequent extended duration surface stays will be significantly driven by the logistics requirements, packaging design and re-supply methodology. Transportation and delivery of the resources required to support extended human presence at a lunar outpost is challenging and will involve significantly more risk and cost than delivery of goods to locations currently re-supplied in Earth orbit, i.e. the International Space Station (ISS). Given the constrained payload capability of currently envisioned lunar transportation systems, there is a balance that must be achieved in order to optimize the permissible crew surface stay time at a given location. In addition to delivery of elements for life support and scientific utilization on the lunar surface, logistics must also be delivered in order to support continued habitability and crew needs. These logistics include crew consumables, spares and maintenance equipment, liquids, and gases; all of which must be packaged, transported to the lunar surface, stored for some period of time before use, and finally disposed of in an appropriate way. Determination and optimization of these exploration system drivers required to support extended duration missions forms the basis of the analysis conducted under the Constellation Program by the Exploration Systems Analysis Team (ESAT).

This paper will present an overview of the strategic analysis conducted by the ESAT for utilization by NASA Constellation Program's decision-makers, with a focus on the influence of logistics on strategic-level Figures of Merit (FOMs). An overview of deterministic and probabilistic analysis approaches is provided, with a discussion of the importance of each approach to understanding the integrated system behavior. The logistics required to support lunar surface habitation are analyzed from both "macro-logistics" and "micro-logistics" perspectives, where macro-logistics focuses on the delivery of goods to a destination and micro-logistics focuses on local handling of re-supply goods at a destination. An example lunar exploration system scenario is provided to tie the theories of strategic analysis to results generation capabilities. This scenario presented is notional and is not representative of NASA's official position on lunar exploration.

2. STRATEGIC ANALYSIS OVERVIEW
One of the primary goals of strategic analysis is to provide an integrated assessment of the logistics over the exploration system life-cycle required to support strategic decision making. This integrated analysis encompasses not only performance, but also uncertainty, risk, and affordability, as well as capturing

their associated linkages and feedbacks. Strategic analysis supports decision making by Constellation Program senior management through study of system robustness as well as alternate strategies. This is enabled by assessment of both the planned and expected benefit and cost of exploration systems, which is aggregated into high-level value metrics and FOMs that enable cost-benefit analysis.

The exploration systems analysis methodology is based on resource utilization analysis using predefined element data sets. The exploration system model does not perform element design or sizing. Rather, those data are provided by element experts from their design and sizing tools and analysis. The data are imported into a library for use in the exploration system analysis model.

The overall methodology is designed to simplify the analysis, while still capturing those details that have major impacts on system performance. For example, the exploration system model does not explicitly perform any transportation system analysis. Instead, it focuses on delivery of elements and goods to locations on the lunar surface. Delivery is driven largely by the amount of mass that a crewed or cargo lunar lander is capable of delivering to a given location. These cargo capacities are provided as inputs to the model from transportation system analysts, such that the model does not require the user to set up launches, in-space rendezvous, engine burns, etc and the model is not required to track propellant, delta-velocities, in-space logistics use, etc. In most other cases, some amount of analysis is performed, but the level of detail is limited.
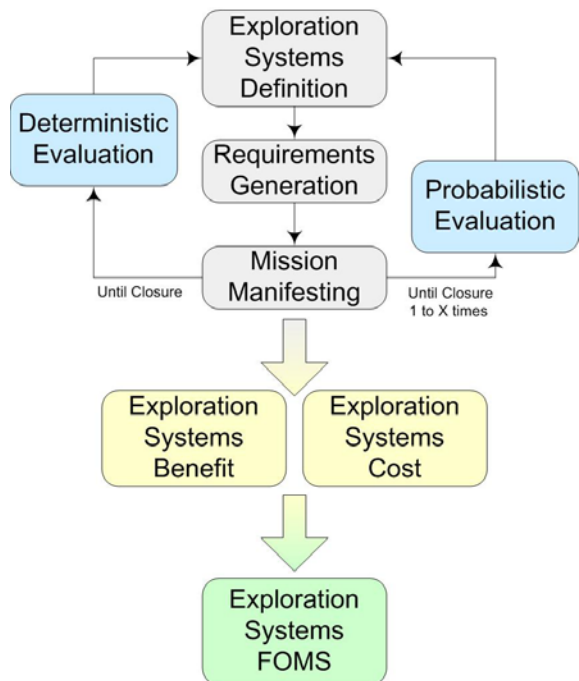

Figure 1: Strategic Analysis Flowchart

Figure 1 provides a high-level overview of the strategic analysis process. Each of these blocks will be briefly described in the following paragraphs (Cirillo, Earle, Goodliff, Reeves, Andraschko, Merrill, and Stromgren 2008).

'Exploration System Definition' is the process in which exploration system architectures and approaches are defined. Flight rates, destinations, transportation system capability, and surface elements specify the exploration system and drive the assumption sets for logistics requirements.

'Requirements Generation' is the calculation of the total mass of required cargo for delivery based on the exploration system definition. Logistics required include: crew resupply, habitat logistics and maintenance, surface element logistics and maintenance, Extra-Vehicular Activity (EVA) consumables, and leakage. The final step is categorization of each logistic by type; either pressurized, unpressurized, gas, or liquid.

'Mission Manifesting' is the optimization of the loading of each mission based on capabilities and requirements from the previous steps and on a set of input loading criteria. Goods are loaded by carrier, accounting for mass and volume limitations.

'Deterministic Evaluation' is the process of evaluating the viability of an exploration system scenario with respect to delivery mass and crew resupply capability. Inputs, such as crew surface durations and scheduling of pressurized logistics container delivery, are varied to maximize exploration system performance, while satisfying the ability to deliver the required exploration system logistics.

'Probabilistic Evaluation' is run after deterministic evaluation and incorporates exploration system risk and evaluates the robustness of the exploration system through Monte Carlo analysis. Exploration systems are adjusted and re-analyzed based on expected loss of mission, crew, rendezvous, and other programmatic risks (as specified).

'Exploration System Benefit' determines which objectives can be satisfied in the given exploration system and then weights can be assigned to determine an overall benefit. Objective weightings are left unassigned by the analyst, so as to allow the decision-maker the freedom to investigate the impact of alternative policy decisions.

'Exploration System Cost' is the calculation of the annual cost of all lunar architecture elements, including Design, Development, Test and Evaluation (DDT&E), Production, and Operations.

'Exploration System FOMs' are high-level Figures of Merit for a given exploration system scenario and are calculated based on scenario performance metrics produced by the exploration system model.

3. DETERMINISTIC AND PROBABALISTIC ANALYSIS

History has shown that complex space exploration systems rarely proceed exactly as planned. Unplanned,

although not always unexpected or unanticipated, events intervene, changing the course of the planned exploration system.

Deterministic analysis alone allows for an evaluation of only the nominal performance of a lunar exploration system. While this is a critical step in the development of the exploration system, using this approach alone neglects the risk and uncertainty associated with human space exploration. Vehicle reliability, technology development risk, budgetary uncertainty, and launch uncertainty all contribute to stochasticity in an exploration system. Strategic analysis that allows for both deterministic and probabilistic modeling will lead to better understanding of the system's range of behaviors due to various modeled uncertainties (Stromgren, Andraschko, Merrill, Cirillo, Earle, and Goodliff 2008).

3.1. Deterministic Analysis

Analysis of the logistics and re-supply methodology of a human lunar outpost/exploration system in a deterministic manner provides an initial assessment of the performance of the exploration system, with the performance being largely driven by logistics resupply constraints for exploration systems supporting extended lunar outpost crewed operations. Sensitivity analysis and trade studies conducted on candidate exploration system scenarios provide insight into the behavior of the nominal exploration system when focused on key system parameters, such as the physical characteristics of the elements, their associated logistics, required crew consumables, and the logistics packaging methodology. Scenarios are defined and analyzed deterministically prior to performing probabilistic assessments.

The deterministic model requires as input an exploration system definition. This definition consists, primarily, of the parameters necessary to describe the set of missions that will constitute the exploration system, such as the number of crew delivered, the length of crewed surface duration, the delivery capacity of the transportation system, and the payloads delivered. Once the exploration system has been defined, the logistics requirements are calculated for each mission based on the mission parameters, the capabilities of the manifested elements, and a set of assumptions about crew consumption, Extra-Vehicular Activity (EVA), logistics, science requirements, and In-Situ Resource Utilization (ISRU). The required logistics are then loaded onto each mission within carriers for delivery prior to their date of use. Any cases in which the logistics could not be loaded due to limited capacity are flagged for further attention. Exploration system definition, logistics requirements calculation, and logistics loading are iteratively performed until the exploration system is performing satisfactorily.

Once the deterministic exploration system has been created, the defined exploration system can then be leveraged as an input into other analysis, to include probabilistic assessments, figures of merit assessment, and sensitivity/trade space analysis.

3.2. Probabilistic Analysis

Methodologies and tools have been developed to provide probabilistic analysis of lunar exploration systems. These probabilistic tools are used to simulate the real-world outcome of exploration systems, based upon the probability of occurrence for non-nominal events, the expected consequence and delays associated with those events, and established contingency operations polices. Using this data, the tools simulate a large number of possible exploration system scenarios, each a possible instantiation of the actual exploration system.

Within each simulated scenario run, the probabilistic exploration system analysis tool performs a mission-by-mission temporal simulation. At each mission step, the tool uses the deterministic exploration system tools to calculate a planned manifest for all remaining missions, including requirements, capacities, and loadings. The outcome of the current mission is then simulated based on probability distributions for all possible non-nominal events and mission event trees. Once the outcome of the mission has been determined, if the mission is successful, the tool tracks the additional material that is delivered to a site on the lunar surface and the amount of material that is consumed. In this manner a running inventory of surface deliverables is maintained. The consumption of material on the lunar surface can also be driven by probabilistic data. Failures of equipment use logistics and crew activity rates can be represented stochastically. If the current mission experiences a failure, then the consequences and resultant delays to the remaining missions are determined, based upon specified contingency operational policy. The remaining flights are reset based upon these consequences.

The tool then moves on to the next flight and repeats the simulation. This flight, and all the flights that follow, are therefore influenced by the events that have occurred cumulatively on all previous flights. After all the flights in a scenario have been simulated, the overall exploration system performance for that case is evaluated. The amount of potential science conducted, the extensibility objectives that are met, additional costs that are incurred, and the risk to the crew are determined.

The probabilistic exploration system tool repeats this process many times, simulating a large number of possible scenario outcomes and collecting performance data for each. The performance data is then integrated into probabilistic distributions for expected exploration system results. These distributions show the likelihood of achieving different levels of exploration system performance based on the current reliability, control policies, and uncertainties within the system. The probability distributions can be compared to the nominal exploration system performance, as predicted in the deterministic exploration system analysis tools, to evaluate the robustness of the given exploration system.

Exploration systems that provide a high level of expected performance across the range of possible probabilistic outcomes are identified as being more robust. That is, they are relatively insensitive to the real-world events that disrupt planned behavior. Exploration systems that exhibit a sharp drop-off in expected performance are less robust.

Based on the results of the probabilistic analysis, revised exploration systems may be developed to provide additional robustness against adverse events and to optimize contingency planning to better ensure a high level of expected exploration system performance. Typically, however, in order to improve the expected performance under probabilistic conditions, it is necessary to sacrifice some level of nominal performance. Nominal performance is typically traded for increased robustness through increased redundancy, contingency deliveries, schedule margin, or other mitigation techniques.

Probabilistic analysis tools allow mitigation techniques to be optimized and can demonstrate the ultimate values of these measures to decision-makers, who otherwise will tend to focus on nominal performance measures. This additional insight into mitigation of critical failures and the implications for the planned exploration system and its associated logistics support necessitate the inclusion of probabilistic analysis when defining an exploration system.

4. MACRO-LOGISTICS

Depending on the overall lunar exploration system, the mass of the logistics and the containers necessary to hold those logistics can account for half to two-thirds of the total mass delivered to the lunar surface by the transportation system. Thus, logistics is a primary driver of overall exploration system performance and must be effectively modeled to reliably predict exploration system performance.

The logistics model (Andraschko, Merrill, and Earle 2008) that is currently incorporated into the deterministic exploration system model tracks the requirements and delivery of logistics that fall into the following seven categories:

1. Pressurized crew consumables – food, clothing, etc.
2. Pressurized spares and maintenance – repair and replacement items for surface elements
3. Unpressurized spares and maintenance
4. Unpressurized science
5. Oxygen
6. Nitrogen
7. Water

The model takes a predefined exploration system, calculates the logistics requirements for each segment, and then manifests carriers and loads logistics onto the landers to ensure that all required logistics are delivered prior to the date they are needed. The model makes some effort to perform the loading efficiently while also accounting for requirements driven by multiple surface locations, element and crew transfers between those locations, and overlapping crew surface periods.

Requirements are calculated for each segment of each mission, by location. Pressurized crew supply requirements are primarily driven by the number of crew and the duration of their stay on the surface. Spares and maintenance requirements are driven by the amount of time each element is active, whether or not crew are present, and total duration on the surface. Science requirements are defined externally on a per mission basis, and incorporated directly into the requirements definition. Oxygen, nitrogen, and water requirements are all based on an Environmental Control & Life Support System (ECLSS) models from subject matter experts at NASA's Johnson Space Center and Marshall Space Flight Center that take as inputs the number of crew, the crew's time on surface, habitat volume, etc. Requirements are calculated for each mission segment and then assigned to the closest lander arrival prior to the start of that segment, to ensure that all required goods will exist at the appropriate location by the time they are needed.

There are additional factors that are currently modeled that affect the requirements calculations. If the ECLSS can electrolyze water, any oxygen requirements are converted to an equivalent water requirement, as water requires less packaging mass and volume to deliver. The model can account for consumables produced by In-Situ Resource Utilization (ISRU) systems, including the buildup of a stockpile over time that is used to reduce requirements on supply delivery. The model also has the capability to allow the crew to extract water from the propellant residuals in the lander descent stage tanks after landing. This value is allowed as a fixed amount per lander and assumes sufficient storage capacity exists and the hardware to convert the propellant residuals to water is in place. To date, exploration systems have utilized both oxygen ISRU and water scavenging techniques to reduce logistics delivery requirements. These options will be further assessed as additional data from actual technology performance evaluations becomes available.

Once the required logistics have been determined and assigned to specific missions, they must be loaded onto those landers or earlier landers traveling to the same location. Logistics must be loaded into logistics carriers, which are then manifested on a lander where space is available. With the exception of the pressurized logistics modules (PLMs), the currently modeled carriers are all derived from the actual carriers used on board the Space Shuttle for delivery to the ISS. Logistics are loaded in these containers up to specified carrier mass and volume limits. The PLM designs are provided by a team of surface habitat designers; however, the packaging for logistics delivered inside the PLMs uses Shuttle & ISS heritage techniques. The pressurized logistics are loaded slightly differently than the other logistics types. They are first loaded into

Cargo Transfer Bags (CTBs), up to the CTB mass and volume limits. The CTBs are then loaded onto the PLMs up to a specified CTB limit, while not violating the PLM mass limit. The manifesting of PLMs on missions is performed by the model user, whereas the manifesting of the unpressurized, oxygen, nitrogen, water carriers and the loading of the PLMs is performed automatically by a logistics loading algorithm.

The loading of logistics into carriers and the carriers onto landers is handled by a loading algorithm that attempts to minimize the unused capacity in each carrier, which therefore minimizes the number of carriers required over the course of the lunar exploration system. This algorithm performs the following set of steps to load the required logistics assigned to each lander, starting with the first mission and progressing to the last:

1. Load required logistics into available space on carriers that are already manifested on any earlier landers at the assigned landing location
2. Load remaining logistics onto the assigned or earlier landers at the assigned landing location, treating already-manifested carriers as if they were filled to capacity, and only manifesting additional carriers if the already-manifested carriers are completely filled
3. Load remaining logistics onto the assigned or earlier landers at the assigned landing location, treating already-manifested carriers as if they are filled to capacity, and manifesting carriers that are not completely filled, as needed
4. Load remaining logistics onto the assigned or earlier landers at the assigned landing location, not treating already-manifested carriers as if they are filled to capacity, and manifesting carriers that are not completely filled, as needed
5. Follow steps 1-4 to load remaining logistics onto landers at OTHER locations if there is a surface element transfer from there to the assigned landing location prior to the assigned landing date
6. If there are additional logistics required that could not be loaded on any previous lander, they are "overloaded" onto the assigned lander (using packaging mass multipliers, rather than actual carrier elements), which will exceed the lander's delivery capacity and cause it to be flagged as "broken"

After the loading has been performed, the user must adjust the element manifest, mission dates and durations, number of crew, or other assumptions and re-run the loading algorithm. This iteration is performed until all required logistics can be loaded into the available space on the landers in the defined exploration system.

5. MICRO-LOGISTICS

The bulk of exploration system manifest analysis has traditionally focused on the delivery of elements and goods to a destination. This focus on macro-logistics captures only a portion of the constraints that will apply to a lunar surface architecture. The local handling of goods at the destination, referred to as "micro-logistics" may also impose severe constraints on architecture operation.

The evaluation of micro-logistics includes a number of areas related to the storage and handling of goods at lunar sites, including: storage requirements for all goods, including system storage requirements for gases and liquids; the movement and storage of cargo carriers; and the collection and disposal of trash (Stromgren, Galan, and Cirillo 2008).

There are several key issues regarding the operation of a lunar outpost that can be analyzed using the micro-logistics models that have been developed for lunar exploration system analysis. Of particular concern is the storage volume required in lunar habitats for all of the consumables that must be accommodated. In addition, the availability of those consumables, particularly critical spares is of significant interest. Other issues include the storage time of goods on the surface, the amount of crew time required to move goods, and the availability of consumables in case of an emergency.

Micro-logistics analysis is conducted using a time-based system dynamics model. This model tracks the location and quantity of all goods at a lunar site over time. Specific items that are tracked include: crew consumables, spares and maintenance items, science equipment and consumables, gases, and liquids. As part of this tracking, the tool models the operation of the ECLSS, simulating the consumption and conversion of gases and liquids.

The system dynamics model is run against a specific set of case results from the exploration system manifest model. Consumption rates, as well as the goods delivery schedule for a specific exploration system are imported. The local storage, movement, and consumption of those goods are then evaluated.

The model simulates how each type of good is moved and used. Consumption rates are dynamic, reflecting real schedules and rates, and accounting for crew timelines and activities. The movement of goods reflects a concept of operations for how each type of good would be stored and positioned and how carriers would be manipulated on the lunar surface. In addition, the model relates crew times to each cargo movement activity simulated in the model and calculates total crew time requirements required to support micro-logistics.

Evaluation of micro-logistics allows analysts to develop logistics plans that can be accommodated using the storage capabilities that are available on the surface and that minimize the crew time required to reposition goods. In addition, this type of analysis provides a prediction for the availability of critical spares and

consumables, which, in turn, can be used to predict the safety and productivity of key surface system elements.

6. EXAMPLE EXPLORATION SYSTEM RESULTS

Over the last decade this strategic analysis methodology has been applied to the Space Shuttle and International Space Station Programs and is now being applied to the development of various options for the planned Constellation Program lunar architecture. The following sections cover the FOMs used to evaluate proposed lunar architectures. Additionally, an example lunar exploration system scenario and sensitivity analysis are presented. Finally, architectural level observations resulting from the analysis of this example scenario are discussed.

6.1. Figures of Merit

Figures of Merit are used to evaluate and compare the relative merits of differing exploration systems, approaches, and executions. The FOMs should be discrete enough to compare relative value expected to be achieved by closely related exploration systems (i.e. capable of evaluating differences in delivered mass, crew time, etc.). For the lunar architecture analyses, a comprehensive set of high-level FOMs were used that covered five major areas: Affordability, Extensibility & Experience, Science & Lunar Survey, Safety & Mission Assurance, and Sustainability.

The Affordability FOMs capture an integrated representation of the ability of a planned budget to cover predicted costs over the life of the exploration system. Affordability results are generated using a combination of deterministic and probabilistic integration and cost estimating tools and models. The scope of affordability integration includes full life cycle costs; conceptual studies, system development, recurring system production, ground & mission operations support, logistics demands, communications infrastructures, prime contractor sustaining engineering, and government oversight costs. The Affordability FOMs consolidate all such information to demonstrate the overages and shortages (cumulative as well as annual) between predicted cost and planned budget profiles. Due to the sensitivity of cost projections and budget implications, only a notional example of Affordability FOM results is included within this paper.

Extensibility & Experience FOMs measure accomplishment in three objective areas: 1) development, testing, and demonstration of relevant technologies, processes, and components for extensibility to future exploration; 2) accumulated experience in living off the Earth, maintaining equipment, and performing useful exploration; and 3) accumulated experience in living on the Moon.

Science & Lunar Survey FOMs measure accomplishments in four objective areas: 1) conduct of fundamental science; 2) science conducted to support future exploration; 3) science/survey conducted to support future lunar exploration; and 4) science/survey

conducted to determine opportunities for commercial endeavors.

Safety & Mission Assurance FOMs measure expected losses of the system. Safety FOMs capture the expected losses that are due to uncertainty or reliability. These include the expected loss of life and expected loss of missions. The primary Safety FOM measures total expected human loss. Mission Assurance FOMs capture expected losses to mission critical elements. FOMs measure probability of loss of these elements. The current risk model utilized was exclusive to transportation system. The surface elements architecture risk model is under development.

The Sustainability FOM measures perceived output of an exploration system and compares that to the minimal acceptable limit. To evaluate Sustainability, a "benchmark event" is established that defines Level of Interest (LOI) required to sustain budget (e.g. Spirit/Opportunity Landing) and a nominal LOI weight is assigned for that event. Next, a LOI weight is assigned to each potential exploration system event based on relative LOI that it will generate. Then, a reasonable "decay rate" is set, where the decay rate is the rate at which interest dissipates. Weights and decay rates are used to calculate a running LOI over the exploration system.

6.2. Example Exploration System Scenario and Sensitivity Analysis

The following section presents an example of strategic analysis. The exploration system scenario presented is notional and is not representative of NASA's official position on lunar exploration. The results of the example scenario are focused on initial outpost buildup and achieving continued human presence. Future analyses will further explore the steady state behavior of the system in more depth. The primary assumptions established for the example scenario include:

- 2019 start date, maximum of 4 missions per year
- Outpost location at Lunar South Pole
- Emphasis on early outpost buildup
- Maximum crewed duration of 180 days
- Current Pressurized Logistics Module (PLM) sizing prioritized to maximize commonality with Core Habitat
- Transportation system performance to Lunar South Pole yields 14.6 t payload for a cargo lander, 0.5 t payload for a crewed sortie lander, and 1.0 t payload for a crewed outpost lander
- Residual propellant in the lander descent stages can be scavenged to generate 400 kg of water per lander

The surface system elements in the exploration system consist of the Core Habitat, power and support units (PSU), mobility chasses (CMC), reusable and disposable pressurized logistics carriers (RPLMs and DPLMs), ISRU oxygen production system (OPS) and

153

tools, small pressurized rovers (SPRs), tri-ATHLETEs (ATHLETE is an acronym for All-Terrain Hex-Legged Extra-Terrestrial Explorer), and lunar communication terminals (LCT). These elements are strategically placed on specific missions to support the emphasis of early outpost buildup. Figure 2 shows the deterministic manifest for the example scenario. Only the surface system elements and pressurized logistics modules are shown in the figure. The unpressurized, gases, and liquid carriers are not shown for clarity. As seen in the figure, the delivery of habitation in 2020 allows for successive crews to stay longer on the lunar surface than a standard sortie mission of 7 days length. The elements are also placed on specific missions to get a balance between the capabilities the elements provide and the logistics required to support the crew for a given number of days. The figure only shows the first seventeen missions since the latter flights would be repeated to sustain continued human presence.

Scenario Description

Example Scenario	Description:	Habitability Emphasis		Date:
		4 Flights per Year		

Lander

Ref. Crewed Capacity	1000 kg
Cargo Lander Capacity	14600 kg
Launch Order	I-V
EDS Loiter	4 Days

Transportation System

Lander Mass	46765 kg
Shroud Diameter	10 m

ECLSS

Water Recovery	yes	93.5%
Brine Recovery	no	-
Solid Waste	no	-
O_2 Generation	yes	
Sabatier	no	
Sabatier+CFR	no	
EVA CO_2 Recovery	no	
Laundry	no	

Outpost Details

1 Core Hab + 2 RPLMs
1 Suitlock
ICP

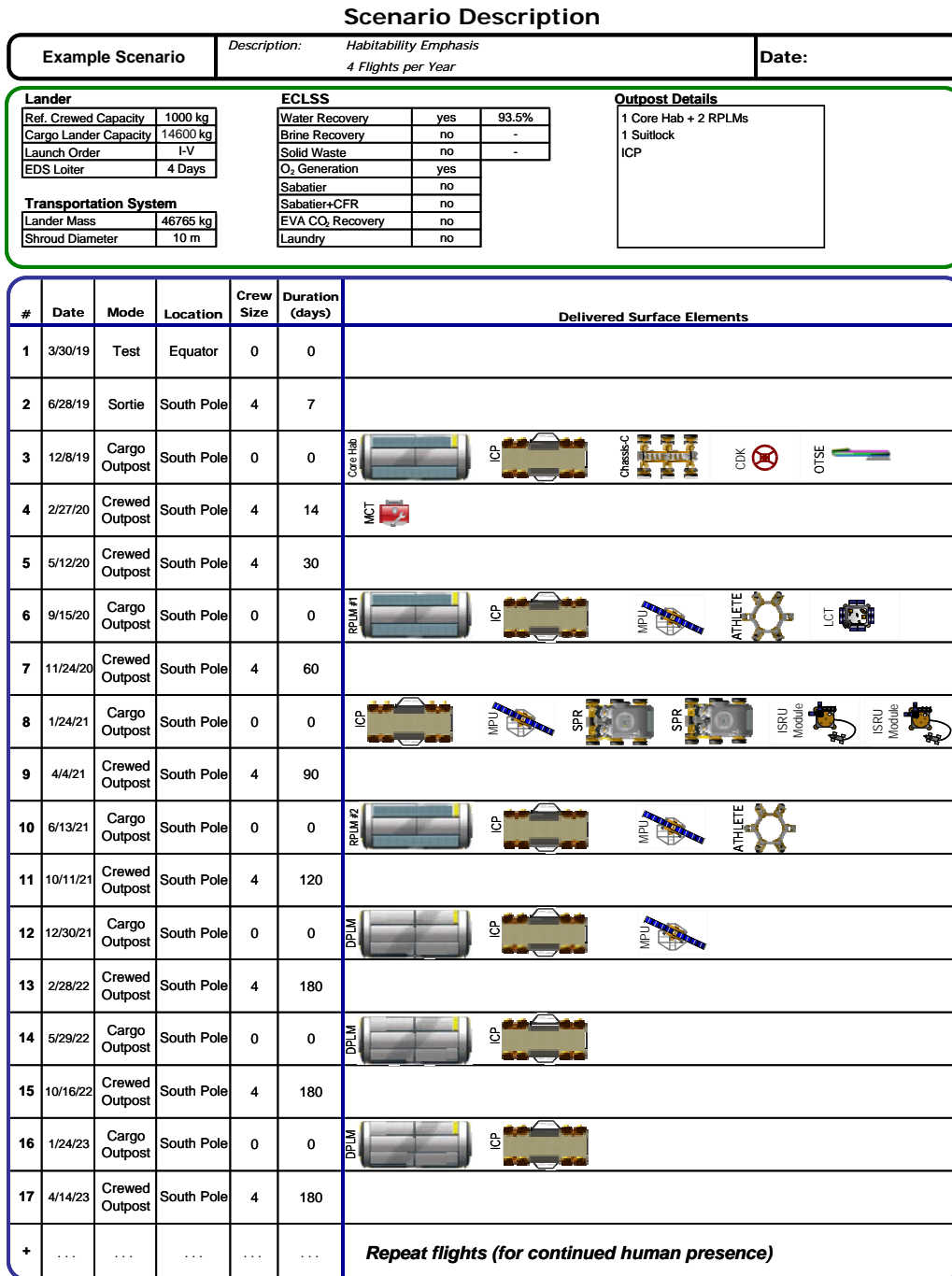| #	Date	Mode	Location	Crew Size	Duration (days)	Delivered Surface Elements
1	3/30/19	Test	Equator	0	0	
2	6/28/19	Sortie	South Pole	4	7	
3	12/8/19	Cargo Outpost	South Pole	0	0	Core Hab, ICP, Chassis-C, CDK, OTSE
4	2/27/20	Crewed Outpost	South Pole	4	14	MCT
5	5/12/20	Crewed Outpost	South Pole	4	30	
6	9/15/20	Cargo Outpost	South Pole	0	0	RPLM #1, ICP, MPU, ATHLETE, LCT
7	11/24/20	Crewed Outpost	South Pole	4	60	
8	1/24/21	Cargo Outpost	South Pole	0	0	ICP, MPU, SPR, SPR, ISRU Module, ISRU Module
9	4/4/21	Crewed Outpost	South Pole	4	90	
10	6/13/21	Cargo Outpost	South Pole	0	0	RPLM #2, ICP, MPU, ATHLETE
11	10/11/21	Crewed Outpost	South Pole	4	120	
12	12/30/21	Cargo Outpost	South Pole	0	0	DPLM, ICP, MPU
13	2/28/22	Crewed Outpost	South Pole	4	180	
14	5/29/22	Cargo Outpost	South Pole	0	0	DPLM, ICP
15	10/16/22	Crewed Outpost	South Pole	4	180	
16	1/24/23	Cargo Outpost	South Pole	0	0	DPLM, ICP
17	4/14/23	Crewed Outpost	South Pole	4	180	
+	…	…	…	…	…	**Repeat flights (for continued human presence)**

Figure 2: Exploration System Description for Example Scenario

The logistics requirements for the example scenario are shown in Figure 3 on a per mission basis (again only through the first seventeen missions). The driving requirements are pressurized goods (i.e. crew consumables and element spares and maintenance mass) followed by unpressurized goods (i.e. element spares and maintenance mass and science). There is no oxygen delivered to the Moon, since the ECLSS has an electrolyzer and water is electrolyzed into hydrogen and oxygen. The water requirement is very close to zero due to the water scavenged from the lander propellant residuals and the ISRU processor producing 1000 kg of oxygen per year. Figure 4 shows how these logistics are delivered on each mission. Logistics are delivered on or before the flight that they are needed to support the crewed missions to the lunar surface. Current transportation system capabilities were primarily driven by a desire to achieve global lunar access. This global access requirement, coupled with current mission design choices that constrain the crew stay on the lunar surface to non eclipse periods, results in an unallocated payload capability on cargo missions during the later part of the scenario. This additional payload capacity could be utilized to send additional elements, science, or other non-pressurized goods.
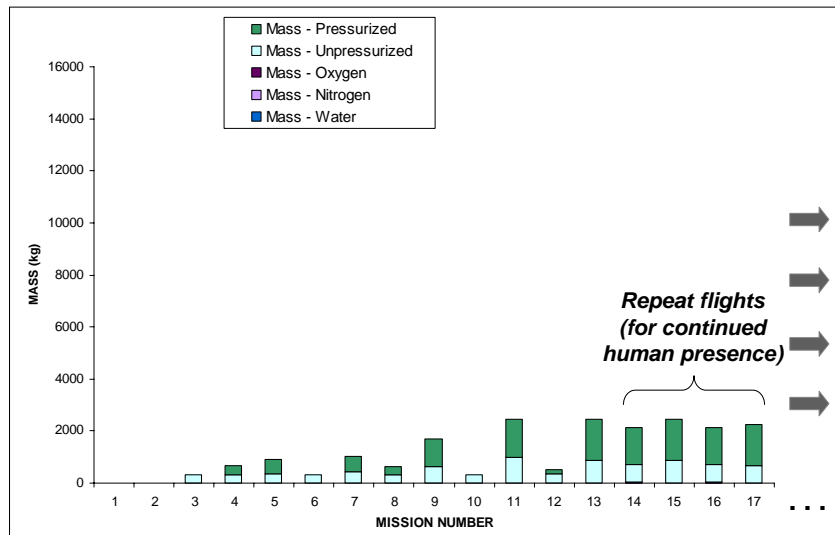

Figure 3: Required Logistics by Mission for Example Exploration System Scenario
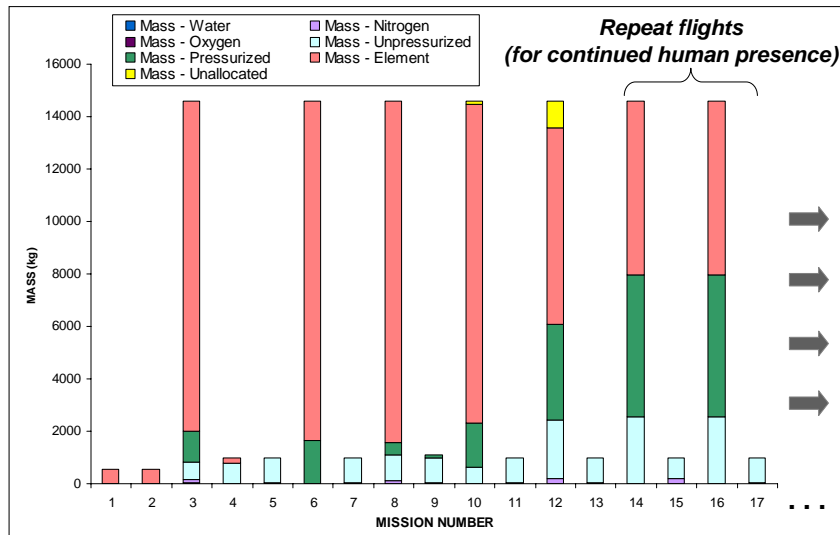

Figure 4: Delivered Mass by Mission for Example Exploration System Scenario

Figure 5 gives the FOMs results for the example scenario. Each of the FOMs gives a comparison of the planned/deterministic scenario and the expected/probabilistic scenario. For multiple exploration systems, the FOMs can be compared side-by-side or cross-plotted to determine the "best" exploration system based on a stakeholders' values and beliefs. For this example scenario, there was no intent to optimize the latter missions in order to improve the Sustainability FOM.

Affordability

Total Budget Delta to 2020, RY$B	TBD
Max Annual Deficit, RY$B	TBD
Max Cumulative Difference, RY$B	TBD

Costs do not include CxP PP&C 65% Confidence Level analysis

Annual Difference Between Budget and Cost

Values are currently TBD

Extensibility/Experience

	Planned	Expected
1st Crewed Mission ▲	2019.5	2020.1
180-Day Capability ◯	2021.5	2023.9

	Planned	Expected
Crewed Missions LOC	19	12.3
Days on Surface	2751	1671
% Extensibility Obj. Met	89%	86%

Crewed Surface Days

- Outpost
- Sortie
- Super Sortie

Science/Lunar Survey

	Nominal	Expected
Science Mass to Surface, kg	9500	5773
Science, IVA/EVA Time, Hours	22667	13486
# Sites Visited X Mobility Radius	1 X 100 km	

Science Crew Time (IVA+EVA) & Surface Mass Devoted to Science

Safety & Mission Success

Prob of Losing 1 or More Crews	31.2%
Prob of Losing 1+ Hab Elements	14.8%
Prob of Losing 2+ Hab Elements	0.6%
Prob of Losing 3+ Hab Elements	0.0%
Prob of Losing > 20% Power	16.8%
Prob of Losing > 50% Power	0.9%

PLOC per Mission

Sustainability

% of Exploration System At or Above Desired Level of Interest	22%

Note: For this example, there was no intent to optimize the latter portion for an increased Level of Interest
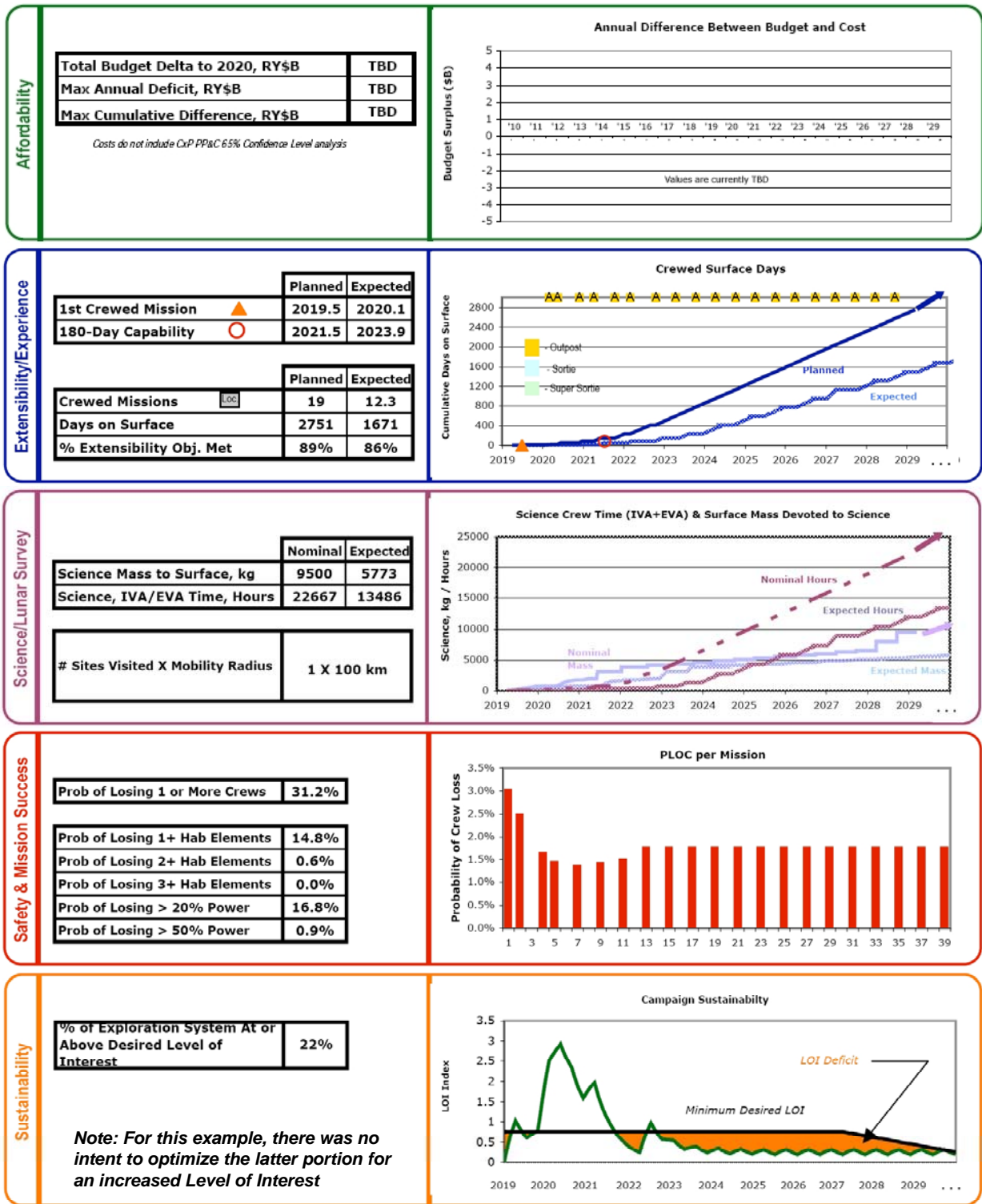
Campaign Sustainabilty

Figure 5: Example Figures of Merit through First Ten Years of Lunar Exploration

As spares and maintenance requirements are a significant driver of exploration system performance, a sensitivity analysis was performed on the example scenario that explored variations in sparing and maintenance mass requirements. For this analysis, sparing and maintenance mass was varied by ±10%, ±25%, and ±50%. The results of the sensitivity analysis are shown in Figure 6. As the figure shows, reduction in spares and maintenance mass required will allow slight increases in crew days, along with significant increases in available mass. Small increases in spares and maintenance requirements lead to slight losses of

crew days and significant reduction in available mass. Large increases in spares and maintenance requirements result in significant loss of crew days and available

mass. Strategic-level analysis when combined with a "bottoms-up" element level assessment is required to yield a more refined spares and maintenance strategy.
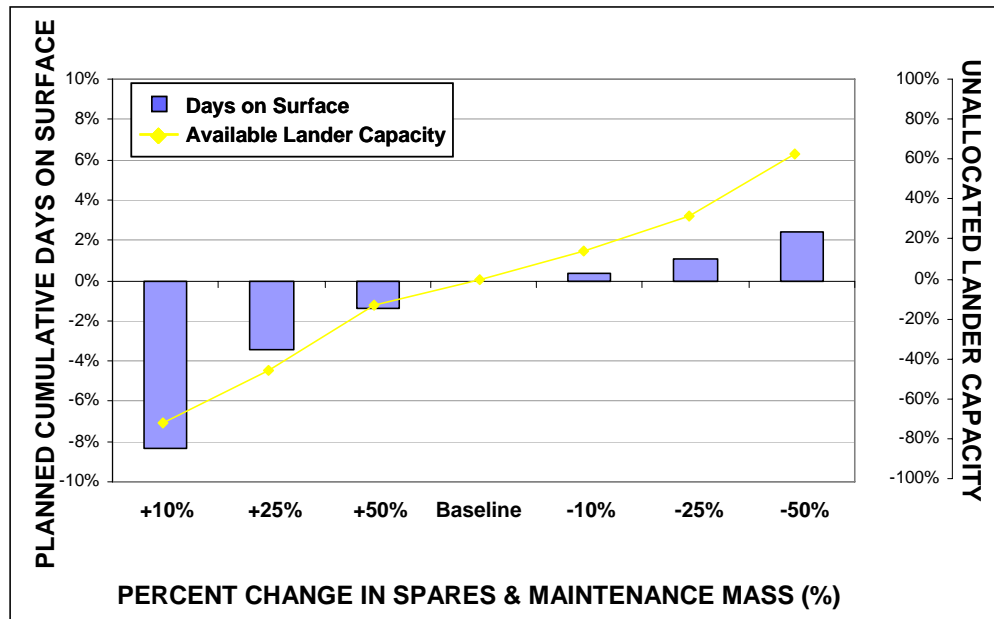

Figure 6: Spares & Maintenance Sensitivity

6.3. Architectural Level Observations

Two key observations were determined as a result of all the exploration systems and sensitivity analyses studied. The first key observation is that a cargo version of the lunar lander enables robustness. The analysis verified that inclusion of a cargo lunar lander is mandatory to enable outpost build-ups that are robust to changes in overall lunar lander performance. The analysis also showed that variations in crewed lunar lander cargo payload performance have secondary impacts on the exploration system behavior when a cargo lunar lander is available to deliver hardware (verified with crew lunar lander cargo payload performance from 0 to 8 metric tons). In addition, variations in cargo lunar lander payload performance have first-order effects on the rate of initial outpost build-up, but less of an impact on long-term exploration system robustness. The second key observation is that logistics are a major exploration system driver. The variability in logistics requirements and strategies remain a first-order driver to exploration system performance.

REFERENCES

Andraschko, M., Merrill, R. G., Earle, K., 2008. Logistics Modeling for Lunar Exploration Systems. *Proceedings of AIAA SPACE 2008 Conference & Exposition*. September 9-11, 2008, San Diego, California, USA.

Cirillo, W., Earle, K., Goodliff, K., Reeves, J.D., Andraschko, M., Merrill, R. G., Stromgren, C., 2008. Strategic Analysis Overview. *Proceedings of AIAA SPACE 2008 Conference & Exposition*.

September 9-11, 2008, San Diego, California, USA.

Cooke, D., 2006. *Exploration Strategy and Architecture*. National Aeronautics and Space Administration. Available from: http://www.nasa.gov/pdf/164310main_2nd_exp_c onf_26_ExplorationStrategyArch_IntegratedLAT_ CookeVolosinLavoie.pdf [Accessed date April 2008].

Cooke, D., 2007. *Lunar Architecture Update*. National Aeronautics and Space Administration. Available from: http://www.nasa.gov/pdf/190083main_AIAA_ES MD_finalSPACE2007.pdf [Accessed date April 2008].

Exploration Systems Architecture Study -- Final Report, NASA-TM-2005-214062, Section 4, November 2005. Available from: http://www.nasa.gov/mission_pages/constellation/ news/ESAS_report.html [Accessed date April 2008]

Stromgren, C., Andraschko, M., Merrill, R. G., Cirillo, W., Earle, K., Goodliff, K., 2008. A Comparison of Probabilistic and Deterministic Strategic Analysis for Human Space Exploration. *Proceedings of AIAA SPACE 2008 Conference & Exposition*. September 9-11, 2008, San Diego, California, USA.

Stromgren, C., Galan, R., Cirillo, W., 2008. Micro-Logistics. *Proceedings of AIAA SPACE 2008 Conference & Exposition*. September 9-11, 2008, San Diego, California, USA.

A DECISION SUPPORT METHODOLOGY FOR PROCESS IN THE LOOP OPTIMISATION

Dan Gladwin [a], Paul Stewart [b], Jill Stewart [c], Rui Chen [d], Edward Winward [e]

[a][b][c] Department of Electronic and Electrical Engineering, University of Sheffield, UK
[d][e] Department of Aeronautical and Automotive Engineering, University of Loughborough, UK

[a]d.gladwin@shef.ac.uk, [b]p.stewart@shef.ac.uk, [c]jill.stewart@shef.ac.uk, [d]r.chen@lboro.ac.uk,
[e]r.e.t.b.winward@lboro.ac.uk

ABSTRACT
Experimental optimisation with hardware-in-the-loop is a common procedure in engineering, particularly in cases where accurate modelling is not possible. A common methodology to support experimental search is to use one of the many gradient descent methods. However, even sophisticated and proven methodologies such as Simulated Annealing (SA) can be significantly challenged in the presence of significant noise. This paper introduces a decision support methodology based upon Response Surfaces (RS), which supplements experimental management based on variable neighbourhood search, and is shown to be highly effective in directing experiments in the presence of significant signal to noise (S-N) ratio and complex combinatorial functions. The methodology is developed on a 3-dimensional surface with multiple local-minima and large basin of attraction, and high S-N ratio. Finally, the method is applied to a real-life automotive experimental application.

Keywords: Experimental decision support, variable neighbourhood search, gradient descent, Simulated Annealing

1. INTRODUCTION

Experimental search with hardware or process in-the-loop is a common procedure in, for example, the engineering and pharmaceutical industries. These procedures are often applied to combinatorial problems during (in the engineering case for example) the development of new hardware systems or control (Stewart, Fleming and MacKenzie 2003, Stewart, Stone and Fleming 2004). The systems under test, can be described as a set of dependent variables which vary according to some functions of independent variables. In this case, there doesn't exist a complete specification of the function relating the variables. This implies that there is no accurate *a-priori* knowledge of the fundamental cause and effect present in the system. Thus, for an example linear function, the values in the co-efficients matrix would be unknown.

Commonly it is required to identify the sets of independent variables which maximise or minimise the dependent variables (Myers and Montgomery 1995). To obtain the necessary information to have a confident estimate of the parameters, it is possible to vary the independent parameters over successive trials (Designed Experiments) and measure the corresponding dependent variables. In order to fully examine this relationship, a large number of trials is often required to identify the location of the desired response. However, real-world problems are difficult to solve by this methodology for a number of reasons (Michalewicz and Fogel 2000):

- The number of possible solutions in the experimental space is so large as to preclude an exhaustive search for the best (or acceptable) and answer (Michalewicz and Fogel 2000).
- The evaluation function that describes the solutions is extremely noisy and/or complex.
- The cost of conducting an experiment at large numbers of points in the search space may be prohibitive in terms of time taken and/or resources used.

These constraints motivate the use of gradient descent methods in order to provide the decision support to direct the search and minimise the number of experiments conducted. Other meta-heuristics such as Genetic Algorithms (Narayana Naik, Gopalakrishnan and Ganguli, 2008) are applicable to this class of problem, but are relatively difficult to implement and tune due to the number of parameters associated with this technique as compared to gradient descent methods. These methods are based upon the statistics of the neighbourhood around a given point, thus relying on local information at each step. However, basic gradient descent methods only provide locally optimum solutions whose values depend on selection of the starting point (Mladenovic and Hansen, 1997). There have been many meta-heuristic methods developed to increase the efficiency of the experimental search, such as simulated annealing (Burke and Kendal 1999), tabu search (Glover and Hanfi 2002), genetic algorithms (Stewart, Stone and Fleming 2004) and variable neighbourhood search (Toksari and Guner 2007).

As the nature of the experimental surface is an unknown, it is important to utilise methodologies which require the minimum number of parameters to be 'tuned' in order to conduct effective search. With this *caveat* in mind, simple gradient descent, simulated annealing (SA) and variable neighbourhood search (VNS) will be considered in this paper, since in most of their varieties,

implementation is simple, and basic tuning rules are readily available.

It has been noted in the literature that the performance of meta-heuristics such as simulated annealing are to some extent compromised when directing search over significantly noisy surfaces (Michalewicz Fogel 2000). This paper describes the implementation of a *weighted stochastic decision support operator* based on RS which guides experimental process to predicted areas of interest in the search space. Basic Gradient Descent, SA and VNS are supplemented by this methodology, and performance is compared to the basic form of the meta-heuristics. The supplemented meta-heuristics are shown to have significantly improved performance when searching noisy environments.

1.1. Scope of this paper

What this paper is about:

- Experimentation on real engineering problems
- Inherently noisy data
- Decision support to reduce experimentation
- Applying a decision support operator to common search methodologies

What this paper is not about:

- Performance comparison of heuristics and meta-heuristics
- Meta-heuristic tuning methodologies
- 'Toy' problems and surfaces

This paper concerns itself with the problem of finding a result in an unseen, noisy search space, where every individual evaluation of a point in the search space is expensive in terms of time and/or resources.

2. SEARCH METHODOLOGIES

A comprehensive description of gradient descent based methods can be found in (Michalewicz and Fogel, 2000). In this section of the paper, the implementations of the algorithms are described. The class of problems addressed in this paper are in general of a *minimisation* type (Baldick 2006). That is, it is desired to minimise a function $f(x)$ over choices of x which lie in the feasible set S such that;

$$f^* = \min_{x \in s} f(x) \tag{1}$$

thus; $\exists x^* \in S$, such that: $f^* = f(x^*)$ and $x \in S \Rightarrow f(x^*) \leq f(x)$. Set S is the constraint set, f^* is the minimum of the problem, while x^* is the minimiser. We will refer to minimisation problems in this paper, however a conversion to maximisation is a trivial task.

2.1. Gradient Descent

A sequence of intermediate values are successively generated by the algorithm. We begin with an initial random guess and successively improve it. In general, none of the iterates exactly solve the problem, therefore we include a termination criteria which when satisfied will cause the algorithm to terminate with a suitable approximation to the exact solution. This is particularly applicable to real-world, noisy surfaces. Iterative hill descent can be described with the general form of recursion;

$$x^{v+1} = x^v + \alpha^v \Delta x^v, v = 0,1,2... \tag{2}$$

where;

- x^0 is the initial guess
- v is the iteration counter
- x^v is the value of the iterate at the vth iteration
- $\alpha^v \in \Re_+$ is the step size
- $\Delta x^v \in \Re^n$ is the step direction
- $\alpha^v \Delta x^v$ is the update to add to the current iterate x^v to obtain new iterate x^{v+1}

In the case of minimisation, the step direction is chosen as to reduce the objective f^*. If we let $\hat{x} \in \Re$, then the vector $\Delta x \in \Re$ is called a descent direction for f at $\hat{x}$ if $\exists \overline{\alpha} \Re_{++}$ such that;

$$0 < \alpha \leq \overline{\alpha} \Rightarrow f(\hat{x} + \alpha \Delta x) < f(\hat{x}), \tag{3}$$

and Δx is a descent direction for f at $\hat{x}$ if the objective is smaller than $f(\hat{x})$ at points along the line segment $\hat{x} + \alpha \Delta x$ for $\alpha > 0$ and $\alpha \leq \overline{\alpha}$. There are some *caveats* associated with gradient descent methods:

- The methods usually terminate at solutions which are only locally optimal
- No information is apparent as to how the discovered local optima deviates from the global minima or other local minima
- The optimum obtained depends on the original configuration
- In general it is not possible to calculate an upper bound for computation time

Gradient Descent thus exploits the best opportunities for improvements, but neglects to explore a large search space. In contrast, random search where points are sampled from S with equal probability explores thoroughly, but forgoes local exploitation. Thus most gradient descent methods execute a random 'jump' at local minima, to balance exploration with exploitation

2.2. Variable Neighbourhood Search

The VNS algorithm implemented in this paper systematically exploits the idea of neighbourhood change in the descent to minima (Burke and Kendall 2005), and attempts to balance local exploitation with global exploration. It is simply an implementation of the basic gradient descent method described in the previous section, however, in this case the step length α^v is variable rather than fixed. A number of variations have been reported, with both lengthening step length (Mladenovic and Hansen 1997) and reducing step length (Toksari and Guner 2006). In this case, reducing step length is implemented, by a static schedule (Kirkpatrick 1983), using a step decementation function given by;

$$C_{k+1} = \alpha \cdot c_k, k = 0,1,2... \tag{4}$$

The initial step length is usually chosen to be significant with respect to the search variable ranges, where α is chosen to be a positive constant > 1. The final value is fixed, generally related to the smallest feasible measurement or control increment of the variables. As with gradient descent, a random 'jump' is implemented to escape local minima.

2.3. Simulated Annealing

The implementation of Simulated Annealing used in this paper is based again on gradient descent, accepting improvements in cost in traversing the search space, however depending on a control parameter c it will accept deteriorations to a limited extent to escape local minima. Initially, at large values of c, large deteriorations will be accepted, as c decreases, smaller deteriorations are accepted, and finally, as the value of c approaches 0, no deteriorations are accepted. The probability of accepting deteriorations is achieved by comparing the value of $\exp\big((f(i) - f(j))/c\big)$ with a random number generated from a uniform distribution on the interval $[0,1]$. In this case, the rate of decrease of the parameter c is achieved by implementing the VNS decrementation function.

3. WEIGHTED STOCHASTIC DECISION SUPPORT OPERATOR (WSDS)

Local search methods such as gradient descent execute a random jump at local optima or other pre-defined termination metric based upon the implementation. SA-type methodologies typically execute a random jump at the termination of the cooling schedule if the global minimum or some upper bound of acceptable performance hasn't been reached. Obviously with unknown experimental functions, the exact value of the global maximum will not be known, however it is common for the designer/experimenter to have an "acceptable" performance metric in mind when starting the experimental procedure which will act as a termination criteria.

Tabu Search has been shown to be a particularly effective meta-heuristic by directing the experimental search 'jumps' away from areas which have been found to be unproductive. However, this doesn't take advantage of the previously gathered data with respect to the possibility of 'predicting' promising areas of search. The RS methodology has been shown to be an effective tool in approximating complex and noisy functions for real time control (Stewart and Fleming 2002, 2004), and thus would appear to be a useful tool to direct experimentation based upon past results.

The response surface methodology is a technique which was initially developed to optimise process control and experimentation by the application of designed experiments in order to characterise the relationship between the system variables and outputs (Myers and Montgomery 1995). The relationship between the response variable of interest y, and the predictor variables $(\xi_1, \xi_2, ..., \xi_k)$ comprise a description of the system of the form

$$y = g(\xi_1, \xi_2, ..., \xi_k) + \varepsilon \tag{5}$$

where ε represents the model error, and includes measurement error, and other variability such as background noise. The error will be assumed to have a normal distribution with zero mean and variance σ^2. In general, the experimenter approximates the system function g with an empirical model of the form

$$y = f(\xi_1, \xi_2, ..., \xi_k) + \varepsilon \tag{6}$$

where f is a polynomial of arbitrary order (generally first or second order in the process control industry). This is the empirical or response surface model. The variables are known as *natural variables* since they are expressed in physical units of measurement. The natural variables are transformed into *coded variables* $x_1, x_2, ... x_k$ which are dimensionless, zero mean, and the same standard deviation. The response function now becomes

$$\eta = f(x_1, x_2, ... x_k) \tag{7}$$

The successful application of RS relies on the identification of a suitable approximation for f. This will often be a first order model of the form

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \tag{8}$$

or a second order model of the form

$$\eta = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x^2_j + \sum_{i<j} \sum \beta_{ij} x_i x_j$$

(9)

It may also be necessary to employ an approximating function greater than an order of two, based on the standard Taylor series expansion. The set of parameters can be estimated by regression analysis based upon the experimental data. The method of least squares is typically used to estimate the regression coefficients. With $n < k$ on the response variable available, giving $y_1, y_2, ... y_n$, each observed response will have an observation on each regression variable, with x_{ij} denoting the ith observation of variable x_j. Assuming that the error term ε has $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$ and the ε_i are uncorrelated random variables. The model can now be expressed in terms of the observations

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$$
$$i = 1, 2, ... n$$

(10)

The β coefficients in equation (10) are chosen such that the sum of the squares of the errors ε_i are minimised via the least squares function

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2$$

(11)

The model can be more usefully expressed in matrix form as

$$y = X\beta + \varepsilon$$

(12)

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{1k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

(13)

It is now necessary to find a vector of least squares estimators b which minimises the expression

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

(14)

and yields the least squares estimator of β which is

$$b = (X'X)^{-1} X'y$$

(15)

and finally, the fitted regression model is

$$\hat{y} = Xb, e = y - \hat{y}$$

(16)

where e is the vector of residual errors of the model. The Response Surface Method can thus be implemented upon either simulated or actual experimental results to derive a polynomial expression describing the relationship between the causal inputs and resulting outputs of the dynamic systems under consideration.

As data from the experimental results are gathered under the direction of the meta-heuristics, it is possible to generate a surface approximation for the system under consideration.

3.1. WSDS Method

The WSDS method in its basic form is encapsulated in the following pseudo-code.

1. UNTIL
2. run meta-heuristic to global minimum (or acceptable value)
3. END
4. ELSE
5. add *new path data* to *old path data*
6. fit *Normalised Response Surface* to *old path data*
7. generate WSDS surface
8. perform weighted jump
9. END

At present, the method uses a fixed 2nd order response surface, however in future work, it is anticipated that this will be developed into an adaptive system. For the development of this methodology, a realistic noisy surface with multiple local minima, plateaus and one global minimum was considered for algorithm

development. The standard Matlab 'Peaks' surface (figure 1) is corrupted with significant measurement and process noise, for example (figure 2).
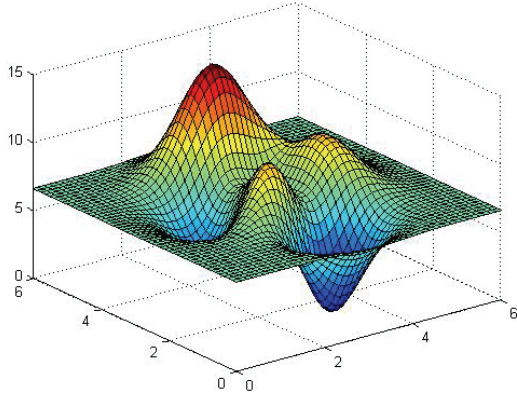

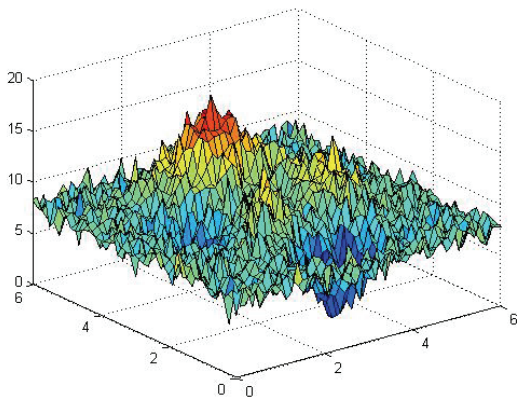Figure 1: Noise-free Algorithm Development Fitness Landscape - Peaks 0

Figure 2: Noisy Algorithm Development Fitness Landscape - Peaks 3

The development surfaces are designated *Peaks 0 - Peaks 3* with increasing levels of noise imposed on the clean *Peaks 0* surface according to;

- Peaks 1 - mean 0.1189, variance 0.0836
- Peaks 2 - mean 0.2842, variance 0.3705
- Peaks 3 - mean 1.7277, variance 0.7648

Using the VNS method as an example, a typical unsuccessful search is shown in figure(3); although unsuccessful, the path does reveal some information about the system, and hence prior to the next "jump" of the meta-heuristic, the most recent path information is added to the path database to generate an up-to date WSDS surface. A fifth order example resulting from the search shown in figure(3), is shown in figure(4).

The contours represent the predicted probability of future successful searches based upon past data.
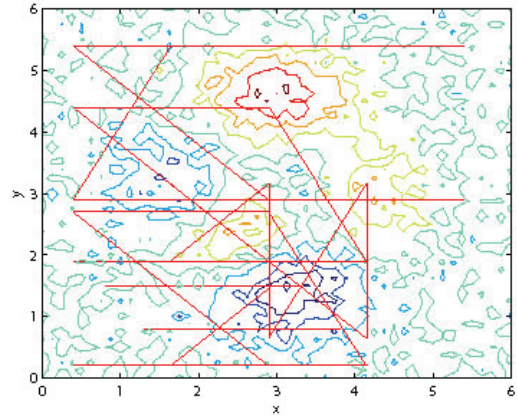

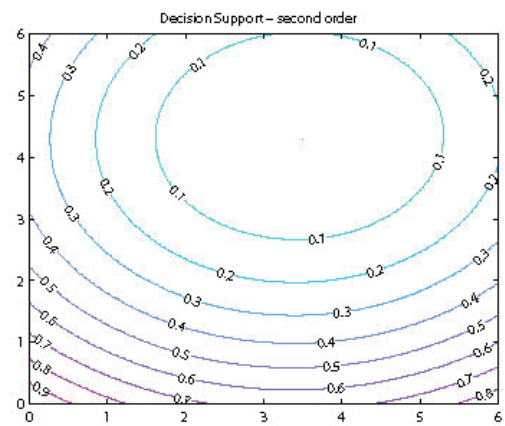Figure 3: Example VNS Decision Support Path

Figure 4: Example Second Order WSDS

The method has successfully identified the lower region of the experimental space as being the most productive in terms of identifying the required acceptable minimum. The parameters for the three search methods are given in table 1.

Table 1: Search Parameters

Method	Stop-point	Stepsize	Schedule
GD	min+0.2	10	none
VNS	min+0.2	$300 \rightarrow 10$	0.5
SA	min+0.2	10	$1 \rightarrow 0.01(0.01)$

4. EXPERIMENTAL RESULTS AND CONCLUSION

Three methodologies (simple gradient descent, variable neighbourhood search, simulated annealing) were applied to a surface of experimental data taken from an automotive application. A technical description of the experimental data is beyond the scope of this paper, however it relates peak cylinder pressures in a single cylinder IC engine with two control parameters which relate to mean and controlled piston trajectories on an experimental hybrid vehicle drivetrain (figure 5). The objective was to find the control parameters which are associated with a predicted peak cylinder pressure.
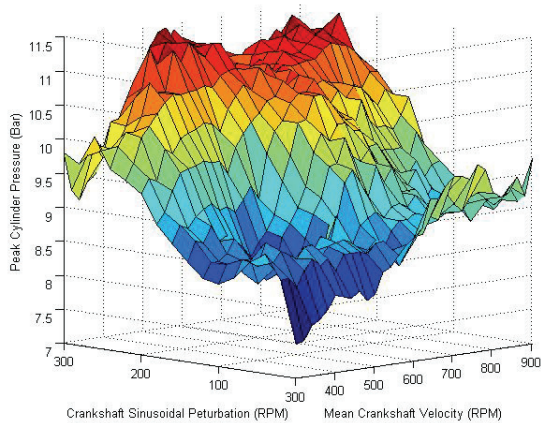
Figure 5: Experimental Engine Response Surface

A comparison of the number of experiments required to find the global maximum by the three search methods is given in table 2, comparing the performance of the methods both with and without the decision support operator. As the methodologies are stochastic by nature, the figures are the averaged results from 50 runs for each application.

Table 2: Support Vs No Support Performance Metrics

Metric	Support?	GD	VNS	SA
Comp	no	9.6098e+3	2.6262e+3	7.257e+3
	yes	757.9	1.8033e+3	879.82
Std	no	8.3786e+3	2.5261e+3	8.1887e+3
	yes	715.9	1.5786e+3	564.2
Max	no	28930	1207	46090
	yes	2945	2440	2776

The decision support method gives an approximate order of magnitude improvement in performance when compared to the base methodologies. Further work will include a study of the effectiveness of higher order surfaces, and the inclusion of this study into adaptive schemes which will further increase the performance of this support method.

REFERENCES

Baldick R., 2002. Applied optimisation, formulation. *Applied Mathematics*, Vol. 119, No. 1, pp. 3-36.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by Simulated Annealing. *Science*, Vol. 220, pp. 671-680.

Michalewicz, Z. and Fogel, D.B., 2006. *How to solve and algorithms for engineering systems.* Cambridge: Cambridge University Press.

Burke, E., Kendall, G., 1990. Comparison of meta-heuristic algorithms for clustering rectangles. *Journal of Computers and Industrial Engineering*, Vol. 37, No. 1, pp. 383-386.

Burke, E., and Kendall, G., 2005. *Search Methodologies: Introductory tutorials in optimisation and decision support techniques*, New York: Springer.

Glover, F., Hanfi, S., 2002. Tabu search and finite convergence. *Journal of Discrete it: modern heuristics,* Berlin: Springer-Verlag.

Mladenovic, M., Hansen, P., 1997. Variable neighbourhood search. *Journal of Computers and Operations Research,* Vol. 24, No. 11, pp. 1097-1100.

Myers, R.H., and Montgomery, D.C., 1995. *Response surface methodology: process and product optimization using designed experiments.* USA: John Wiley and Sons Inc.

Narayana Naik, G., Gopalakrishnan, S., Ganguli, R., 2008. Design optimization of composites using genetic algorithms and failure mechanism based failure criterion. *Composite Structures*, Vol. 83, No. 4, pp. 354-367.

Stewart, P., and Fleming, P.J., 2002. The Response Surface Methodology for rapid prototyping of real-time control systems. *2002 IEEE American Control Conference*, pp. 3343-3348, May 8-10, Anchorage (Alaska, USA).

Stewart, P., Fleming, P.J., MacKenzie, S.A., 2003. Real Time Simulation and Control Systems Design by the Response Surface Methodology and Designed Experiments. *International Journal of Systems Science,* Vol. 34, No. 14-15, pp. 837-850.

Stewart, P., Stone, D.A., Fleming, P.J., 2004. Design of robust fuzzy-logic control systems by multi-objective evolutionary methods with hardware in the loop. *IFAC Journal of Engineering Applications of Artificial Intelligence*, Vol.70, no.3, pp. 275-284.

Stewart, P., Fleming, P.J., 2004. Drive by wire control of automotive driveline oscillations by response surface methodology. *IEEE Transactions on Control Systems Technology*, Vol.12(5), pp. 737-741.

Toksari, M.D., Guner, E., 2007. Solving the unconstrained optimisation problem by a variable neighbourhood search. *Journal of Mathematical Analysis and Applications*, No. 328, pp. 1178-1187.

Organisational Analysis of Different Modalities in Drug Administration

Lucio Buffoni (a), Libero Ciuffreda (a) , Antonio Di Leva (b), Salvatore Femiano (b)

(a) Centro Oncologico Ematologico Subalpino, A.S.O. San Giovanni Battista di Torino
C.so Bramante 88, 10126 Torino (Italy)
(b) Dipartimento di Informatica, Università di Torino - Corso Svizzera 185, 10149 Torino (Italy)

lciuffreda@molinette.piemonte.it, bufluc@libero.it, (dileva|femiano)@di.unito.it

ABSTRACT
The aim of the study is the organisational analysis of the care pathway for patients in the Oncology Division of a large hospital, in order to evaluate pros and cons of different drug administration modalities (the oral modality compared with the endovenous one). The study was focused on patients treated with the chemotherapic drug Navelbina in a day hospital setting, using a process analysis methodology. Starting from the pathway specification and from quantitative parameters of the workload, the efficiency of the patient pathway in terms of cycle time and resources exploitation has been assessed through simulations. This was done in order to highlight the real positive effect of the oral drug administration compared with the traditional one and also to point out solutions to improve the efficiency and to optimise the resources management. The study findings have confirmed the advantages of the oral type modality.

Keywords: business process modeling, software engineering

1. INTRODUCTION

Since a few years, hospital health care units are undergoing management reorganization in order to improve the service and to be more competitive with other health service structures; particular attention is given to budget problems.

It is a very complex task to manage the development of an integrated care department. Managers are facing a relevant number of favorable and unfavorable conditions. To deal with these conditions, they need methods that provide qualitative and quantitative information about care processes. Service quality is strictly dependent on the ability to coordinate different skills and duties, even more than on the ability to optimize each of them. In this scenario, new management tools and ideas are needed in order to conciliate service quality and budged optimization.

This paper presents a practical approach for a quantitative analysis of the Oncology Division (Centro Oncologico Ematologico Subalpino - COES) of the largest hospital in Torino, the *Azienda Ospedaliera San Giovanni Battista*. Aim of the project is to analyse the patient care pathway in the DH (Day Hospital).

A patient care pathway is the "route" that a patient will take from their first contact with a member of the COES staff, usually a nurse of the COES-DH, to the completion of their treatment. You can think of it as a process, on which every activity relating to treatment can be described.

Activities such as consultations, diagnosis, treatment, medication, assessment, and preparing for discharge from the COES-DH can all be mapped on this timeline. Thus the pathway gives an outline of what is likely to happen on the patient's journey, i.e. it details the essential steps in the care of patients with a specific clinical problem and describe the expected progress of the patient (Campbell, Hotchkiss, Bradshaw, Porteous 1998).

Specifically, our analysis is focused on the advantages of prescribing the drug *Navelbina* through the oral way and not by direct injection into the blood stream. The drug *Navelbina* is a very good study subject as it is available in both oral and injecting form and is widely used in clinical treatment. Two different reorganizations of this process, both aiming at increasing the service efficiency, are discussed below.

The first one introduces the oral way of drug administration with no need to modify the present organizational structure. The second one, in addition to what included in the first solution, introduces a change in the infrastructure and technology of the oncology division.

In the literature, there is a strong support for the reengineering of healthcare organizations. Enterprise methodologies originally developed for manufacturing processes are now used to improve the operations and competitiveness of hospitals (Lin., Jomon, 2005) (Detlev Smaltz 2005). Both qualitative (e.g. SWOT - Strengths, Weaknesses, Opportunities and Threats) analysis and quantitative (e.g. process evaluation based on discrete event simulation) analysis have been exploited in real life applications (Harding, Paul, Gillis and Kaye 1993) (Abu-Own A, et al 1999) (Bell CM, et al 2000).

This paper is structured as follows. The second section presents M*-COMPLEX, a general-purpose

open methodology that has been developed to study complex systems (Berio, Di Leva, Giolito, and Vernadat 1995) (Berio, Di Leva 2000) , which includes a business process analysis and reengineering phase.

The third section illustrates the case study, which aims to improve the efficiency and to optimise the resources management of the target organization: the oncological division of a large hospital in our city. Finally the fourth section presents some preliminary conclusions of our analysis.

2. METHODOLOGY

M*-COMPLEX is a structured framework which provides a step-by-step strategy ensuring consistent results. Itanalyses functional, behavioural, and organizational aspects of the object organization, and it strongly enforces an event-driven process-based approach as opposed to traditional function-based approaches for analyzing and designing computer-supported integrated engineering environments.

At the organization level, M*-COMPLEX views the world from two orthogonal points of view. First of all, an enterprise can be analyzed in terms of organization elements that can be classified as organization units (units for short), which control other units at a subordinated level, and so on. Units at the bottom level are called work centres.

Units define areas of responsibilities and authorities and must be analyzed in order to identify their functions, i.e. things to be done and services to be provided. Top-level functions are decomposed at different levels of detail, until the bottom level in which activities are carried on by work centres.

From the other point of view, activities are executed by resources, processing or producing different objects (pure information or material objects). They are subject to scheduling or planning and can be coordinated into organization processes. Thus, an enterprise can be seen as a collection of concurrent processes that define the flow of actions and are triggered by stimuli called events.

Each process specifies the complex control flow between enterprise activities: it shows which activities should be performed at a time for achieving process objectives. The Organization Analysis phase of M*-COMPLEX is structured into two major steps: As-Is analysis and To-Be analysis.

The aim of the As-Is analysis step is to provide managers and engineers with an accurate model of the enterprise as it stands, out of which they can make a good assessment of its current status. The As-Is analysis step encompasses the following sequential tasks: Structural Analysis, Functional Analysis, Process Reconstruction, and Validation.

Other than modelling activities, those tasks also suggest how to report current problems concerning the represented enterprise components, new requirements, and how to discover and report potential and unknown problems.

The To-Be analysis provides guidelines for transforming schema developed during the As-Is analysis and it encompasses three major sequential tasks: Diagnosis, Restructuring, and Validation. The Diagnosis task should point out the potential causes of the current problems reported during the As-Is step.

A matrix cause/solution suggests the guidelines to perform the Restructuring task that modifies models issued from the As-Is step. Finally, adopted solutions are validated against current problems and new requirements during the Validation task.

The analysis is supported by a set of modelling languages, i.e. a set of concepts and constructs which need to be used and shared both by analysts and business users. The integrated model that is developed in the As-Is step consists of a functional model and a process model. These models are based, respectively, on the IDEF0 and the BPMN languages.

The IDEF0 language, due to the simplicity and intuitive appeal of its graphical notations, represents the most widespread formalism for the functional modelling and analysis of enterprises (*IEEE Std 1320.1-1998*).

Complying with business process standards, the BPMN (Business Process Modelling Notation) language has been selected for the description of the process model. BPMN is a graphical notation that has been specifically designed to coordinate the sequence of processes and the messages that flow between different process participants in a related set of activities (BPMN.org 2006).

Moreover, BPMN specifications can be simulated by means of discrete event simulation tools, nowadays available on the market (like iGrafxProcess (igrafx.com). Through simulation, the block designer can manipulate building blocks to check their semantic correctness and to see where inefficiencies lie.

It is also important to remember that the simulation allows an effective "what-if" analysis, checking hypothetical business scenarios, and highlight workloads, resources (in terms of costs and scheduling), and activities (durations, costs, resource consumption).

At last, BPMN objects can be mapped in BPEL, that is the Business Process Execution Language for Web Services (Business Process Execution Language for Web Services). For instance, iGrafxProcess, the tool that has been used in our research, is able to convert BPMN diagrams into BPEL files that specify the sequence of Web Services to be executed.

3. CASE STUDY: THE PATIENT CARE PROCESS IN THE COES-DH ONCOLOGY CENTRE

The COES is a multi-disciplinary research and health care structure, and is one of the most important centres in Italy and Europe. COES research activities take place among many medical specialities (oncology, haematology, endocrinology), diagnostics specialities (molecular biology, tutor immunology, cytogenetic), and radiant treatment. In this study we have paid

attention to the Day Hospital (COES-DH), where usually antiblastic therapies for the care of all solid tumours are administrated.

3.3 Process analysis

During the meetings with medical and nursing staff, several problems concerning the patient care management inside the COES-DH have been pointed out. Among them, infrastructure lacks as absence of integration of tools, scarce automation, logistics problems and hardware resources lack if compared with the workload can be noticed. To face these problems, it is important to describe and analyze the patient care process in the COES-DH.

Patients, after having been accepted, wait for blood test. Then, the patient waits again the results of haematic and chemist exams. Patients that have to be pre-hydrated are immediately hydrated after their blood test. At the same time, blood testtubes are sent to the Laboratory by auxiliary staff. The doctor, during the wait, visits the patients and prepares a draft of the chemotherapy for each of them in order to optimize waiting times.

When the doctor receives the results of the exams, he goes on with the study of results with the aim of customising and fix the therapy. If the results show some problems (toxicity, fever, few neutrophil,) the doctor could decide to prescribe a support therapy and defer the chemotherapy in the next week, otherwise the doctor prints the request of the therapy and sends it to the internal Pharmacy by fax.

The waiting times for the drug preparation and results of the blood exams are long. As a consequence, the cycle time of a patient in the COES-DH is very long.

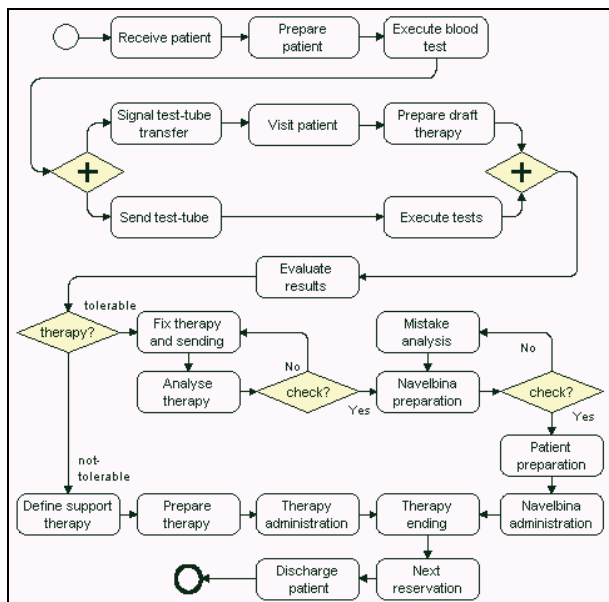In Fig. 3.1 the "As-Is" process for the intravenous administration of Navelbina is shown.

Fig.3.1 – "As-Is Analysis" – the "intravenous administration" process

3.3 Simulation

After the mapping of the process, we prepared an observation chart where, for patients that used the

Navelbina drug, times related to all activities of the process can be collected.

To evaluate the real workload, the activities have been observed for a week. For each activity we measured the starting/ending point, the arrival time of patients, sending fax time, and results and drugs arrival times.

The process uses the following resources (5 nurses and 3 doctors with different schedules):

Table 1

Resource	n.of	schedule
Doctor	1	from 8 am to 3 pm
Doctors	2	from 9.15 am to 4.30 pm
Nurses	2	from 8 am to 3 pm
Nurses	2	from 8.30 am to 4.30
nurse	1	from 3 pm to 11 pm

The activities'duration of the model have been deduced from the experimental measures. A normal distribution has been used in the event that the Kolmogorov-Smirnov & Shapiro-Wilk test returns positive values. Otherwise, a triangular distribution based on *min*, *max*, and *mode* values of the sample has been selected. Results of this analysis are displayed in Table 1.

Activity name	Resources	Time (min)
Receive patient	1 nurse	UnifDist(1;2)
Prepare patient (blood test preparation)	1 nurse	TriangleDist(35;41;38)
Execute blood test	1 nurse	UnifDist(2;3)
Signal test-tube transfer	1 nurse	UnifDist(1;3)
Visit patient	1 doctor	TriangleDist(5;8;7)
Prepare draft therapy	1 doctor	UnifDist(5;7)
Send test-tube (send test-tube to Laboratory)	Staff	UnifDist(26;35)
Execute tests	Laboratori	TriangleDist(37;63;47)
Evaluate results (evaluation of exams and patient examination)	1 doctor	UnifDist(3;5)
Define support therapy	1 doctor	TriangleDist(3;7;5)
Prepare therapy (prepare support therapy)	1 nurse	TriangleDist(2;5;3)
Therapy administration	1 nurse	TriangleDist(10;20;12)
Therapy ending	1 nurse	TriangleDist(3;5;4)
Fix therapy and sending (the therapy is fixed and then will be sent by fax to the Pharmacy)	1 doctor	UnifDist(7;12)
Analyse therapy	Pharmacy	UnifDist(4;7)
Navelbina preparation	Pharmacy	NormDist(52;14)

Mistake analysis (the doctor settles any problems in the drug preparation)	1 doctor	TriangleDist(5;7;6)
Patient preparation	1 nurse	TriangleDist(2;5;3)
Navelbina administration	1 nurse	UnifDist(10;15)
Therapy ending	1 nurse	TriangleDist(3;5;4)
Next reservation	1 doctor	UnifDist(3;5)
Discharge patient	1 nurse	UnifDist(6;8)

By means of a simulation of this "As-Is process", it is possible to obtain some significant parameters as cycle time (range of time that a patient spends in the COES-DH) and resource utilization of the more critical resources (doctors and nurses). In our case, the following results have been obtained:

Resource utilization:
- doctor 57%
- nurse 53%

Cycle time: 205 minutes.

It must be pointed out that the resource utilization applies to the particular care process we have studied and not to the whole activity executed in the COES-DH. Indeed if we insert in the process any other kind of chemotherapy, all resources turn out to be heavily used.
In analysing simulation results it must be pointed out that main difficulties are related to the long waiting times to obtain exams results of analysis from the laboratory and drugs from the Pharmacy. Let us analyse these problems separately.

LABORATORY
Waiting time to receive results from the Laboratory depends on three factors:
- - Test-tube labelling.
- - Test-tube transport from COES-DH to Laboratory.
- - Test result availability notification.

Test-tube labelling is a process that influences the waiting time to obtain exam results. Indeed, bad printing of the label or its wrong positioning on a test-tube results in the arrest of the analysis automated line.

This requires intervention by a technician to resume the line. In order to prevent this event, robots have been developed to produce test-tubes in which the labels are correctly printed and positioned.

Test-tubes are currently transported from COES-DH to Laboratory by auxiliary staff. This process is time consuming (it requires about 30 minutes) and this is a relevant part of the total waiting time. A good solution to this problem would be the employment of a Pneumatic Mail tube system in substitution of the auxiliary staff; this would result in a considerable save in transport time.

Regarding test result notification, at present doctors, in order to know test results, have to repeatedly check the result availability with queries to a software application. A possible solution would be the use of

acoustic and visual signals to let the doctors know as soon as the test results are ready. This way, waiting time would be reduced from the current 26- 35 minutes to about 4-5 minutes.

PHARMACY
Waiting time to receive drugs depends on two factors:
- Transmission of therapy requests.
- Drug preparation and transport.

At present, doctors have to insert a therapy request into the LIS (Local Information System), print it and then send it by fax to the Pharmacy. It might happen that the doctor decides to make some changes to a therapy on the base of the patient's condition. Some times the doctor introduces these changes by writing on paper and not on LIS. Since the paper form is the only request form officially accepted in the Pharmacy, the pharmacist has to add further effort to his job, and introduces new manual activities in the procedure. This also introduces in the process an element of risk. A possible solution would be the use of a new certified computerized procedure in place of the fax procedure.

The second factor which influences the waiting time depends on the time which is necessary to prepare the drug and to transport it by means of an auxiliary staff. We have measured it takes about 52 minutes to obtain the drug. A possible corrective action would be the use of the same drug, but administered by oral way instead of intravenous way.

This solution would allow eliminating inquiry, preparation and delivery times if just the COES-DH can manage the oral chemotherapy in a local warehouse.
Computer and telematic resources computer and telematic resources of the COES-DH department are insufficient. Indeed, simulation shows that three PC and one telephone line, used for both fax and voice, are inadequate for the workload.

A solution could be the addition of a telephone line and of a wireless LAN into the COES-DH. This last improvement would allow doctors to use a tablet PC, one per doctor, to define chemotherapy right at the side of the patient bed.

3.3 Reorganization of the patient care pathway
The oral administration of Navelbina implies some changes in the "As-Is" pathway. The analysis of the oral administration has been conducted according to the restructured "To-Be" care pathway illustrated in Fig. 2.
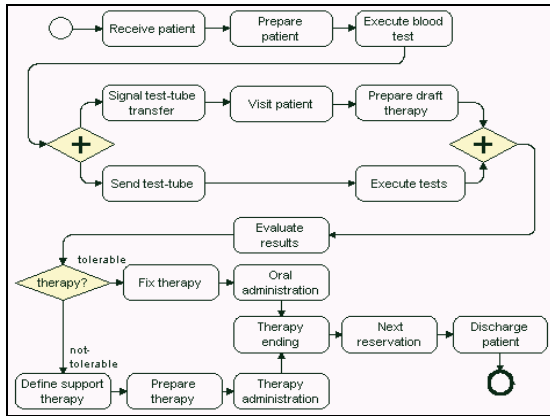
Fig.3.2 – "To-Be Analysis" – the "oral administration" process

The new activities, *Fix therapy* and *Oral administration*, are illustrated in Table 2.

Activity name	Resources	Time (min)
Fix therapy (the therapy is fixed and then the oral drug is delivered to the patient)	1 nurse	UnifDist(2;3)
Oral administration	1 doctor	UnifDist(1;2)

Table 2

Using as a starting point the results of the simulation of the As-Is care process, we have defined two reorganization scenarios:

Scenario A:

The working hypothesis is the variation in therapy administration, from intravenous to oral way. This solution implies a significant variation of the interactions between COES-DH and Pharmacy. Oral administration of the therapy allows reducing the interactions between the two business units. In this hypothesis, the COES-DH has to manage a local warehouse with the oral chemotherapy (Navelbina) supplied by the Pharmacy.

This way all the steps of drug request, drug preparation and waiting time, and all the backup procedures necessary in case of faulty delivery are removed. The doctor once received the exam results and having assessed the patient therapy, delivers the oral chemotherapy to the patient.

Scenario B:

Resource improvements described above, such as the use of a robot for test-tube labelling, a pneumatic mail test-tube system and a certified computerized system to advice doctors that exam results are ready.

The analysis of this new process, using the scenario A, shows the following results:

Resource utilization:
- doctor 47%
- nurse   48%

Cycle time: 141 minutes.

We observe an overall reduction in the patient cycle time of about 31%.

Let's now study scenario B. The process includes the previously described suggestions, which allow reducing crucial waiting time steps, such as exam result notification. The process is the same as in scenario A, we just have changed the temporal characteristics of the activities that are involved in the adoption of new technologies. The analysis of the new
process shows the following results:

Resource utilization:
- doctor 43%
- nurse   42%

Cycle time: 107 minutes.

Thus the overall reduction in the patient cycle time is about 48%

CONCLUSIONS

In this paper we present a methodology that responds to some of the problems organisations are faced with in their process analysis projects. The main objective of the paper is to investigate some potential benefits and outcomes of introducing new processes that could be assessed in advance by using simulation modelling.

In this research, the patient care process in the COES-DH was modelled using a process mapping and simulation tool, iGrafx Process. It has been proved to be a very useful tool for business process analysis and design which offers a way to understand the behaviour of existing and restructured processes without be involved in costly deployment procedures.

This analysis is still under study and the results obtained are influenced by low cardinality of statistical units analyzed. Nevertheless the results obtained seem to be a good estimation of the reality. The benefits of the restructured process has been analysed and two different scenarios were compared.

The first solution just changes the way to administrate the therapy, with no changes in the COES-DH organization.

The second solution is based on the first one, but introduces all the technology innovations previously described. We observe that both solutions provide relevant improvements with respect to the original process. Specifically, referring to the patient cycle time (the overall time a patient spends in the COES-DH), the two solutions allow reducing the cycle time of about 31% and 48% respectively.

It must be pointed out that the development trend of pharmaceutical companies will be based on investments on new molecules with oral administration that can be delivered at patient' home. In the near future we intend to investigate how this trend could impact on the organization of the oncological division and to analyse the benefits of a solution that take into account this new kind of administration.

Bibliography

Abu-Own A, et al, 1999. "Vascular surgical society of Great Britain and Ireland: integrated pathways for vascular surgery" *British Journal of Surgery* 86(5):703.

Bell CM, et al, 2000. "Methodological issues in the use of guidelines and audit to improve clinical effectiveness in breast cancer in one United Kingdom health region" *European Journal of Surgical Oncology* Mar; 26(2):130–6.

Berio G., Di Leva A., Giolito P., and Vernadat F, 1995. "The M*-OBJECT Methodology for Information System Design in CIM Environments" *IEEE Trans. on System, Man, and Cybernetics* 25: 68-85 .

Berio G., Di Leva A, 2000. "The M*-COMPLEX Approach to Enterprise Modeling, Engineering, and Integration", Conceptual Modeling for E-business and the Web, *LNCS Springer Verlag* n.1921, pp. 7-18.

BPMN.org, Business Process Modeling Notation, 2006. *BPMN 1.0: OMG Final Adopted Specification*.

Campbell H, Hotchkiss R, Bradshaw N, Porteous M, 1998. "Integrated care pathways" BMJ;316(7125):133-7, Jan 10.

Detlev H Smaltz et al, 2005. " Making sausage--effective management of enterprise-wide clinical IT projects "*Spring*;19(2):48-55.

IEEE Std 1320.1-1998. IEEE Standard for Functional Modeling Language—Syntax and Semantics for IDEF0. New York: IEEE.

iGrafx.com: Unlocking the Potential of Business Process Management.
http://portal.igrafx.com/downloads/documents/bpm_whitepaper.pdf

Harding MJ, Paul J, Gillis CR, Kaye SB, 1993. "Management of malignant teratoma: does referral to a specialist unit matter?" *Lancet*;341:999–1002.

Lin L., Jomon P, 2005. "Simulation and Lean Six Sigma Integration for Improvement of Hospital Operations" *Proc. 19th Annual Society for Health Systems Conf.,* New Orleans , 2007.; 19:48-55

Business Process Execution Language for Web Services V1.1:
http://dev2dev.bea.com/webservices/BPEL4WS.html

A REACTIVE SCHEDULING FOR INTESIVE CARE UNITS

(a) **Erhan Kozan**

(a) School of Mathematical Sciences, Queensland University of Technology, Australia

(a) e.kozan@qut.edu.au

ABSTRACT

This paper leads to significant improvement in intensive care units (ICU) operating efficiency and productivity by optimising scarce resources. Scheduling of patients in the ICU is complicated by the two general types; elective surgery and emergency arrivals. The job shop approach to reactive scheduling system for ICU promises considerable benefits over existing approaches, and allows problems of large size and complexity to be solved with great accuracy. A general parallel machine job shop model is proposed that fits uncertain arrivals with priorities and the allowance of rejection events at Intensive Care Units. An integer programming model is proposed to handle uncertain arrivals with priorities and while allowing rejections.

Keywords: Scheduling, Health Services, Integer Programming

1. INTRODUCTION

There are a number of studies that have focussed on the ICU. Operations research issues investigated include scheduling of patients and resources, allocation of limited resources and physical design of the facility. These will be discussed as well as general scheduling and resource allocation methods that have been applied to hospitals.

Kim and Horowitz (1999, 2000 and 2002) describe models which are the closest to our proposed model. However they do not encompass the whole of the operating theatre (OT) and ICU. In these studies the ICU is modelled completely, and patients who require the ICU after a surgical intervention are included. Their focus is also on balancing deterministic and stochastic arrivals. To accomplish this they have used a quota system to specify the number of beds available to deterministic arrivals each day. They experimented with one or two weeks scheduling windows. Simulation was used to compare the different scenarios. For the particular hospital studied it was found that a scheme allocating two beds on Monday and Friday, and one bed on Tuesday, Wednesday and Thursday with bookings taken for beds over a two week period was optimal. The proposed model will be looking at scheduling patients on a day by day basis, whereas this model determines the best resource allocation for the ICU.

Similar to research undertaken by Kim and Horowitz (2002) and Ridge et al. (1998) develops a simulation model of the ICU focussing on minimising the number of deterministic arrivals that are rescheduled. Notable inclusions in the model are that rescheduled surgeries re-enter the model, and a queuing theory model is used to verify the output. Some sensitivity analysis was performed to ascertain the effects that important variables had on the system. These variables were the number of beds in the ICU, length of reschedule times and the number of beds reserved for emergency admissions. The results were intended to be used as part of a decision analysis tool to decide allocation of beds. While the main focus was on the number of emergency patient transfers it was concluded a more effective patient admission scheduling system could benefit the hospital being analysed. Due to the stochastic nature of arrivals to the critical care facility it is a requirement to reduce their effect on the objective of the schedule.

Sahinidis (2004) reviews the theory and methods developed to cope with the complexity of optimisation problems under uncertainty. The main approaches to handle uncertainty are stochastic programming, robust stochastic programming, fuzzy programming and stochastic dynamic programming.

Stochastic programming uses a two-stage solution system, where the decision variables are partitioned into two sets. The first stage variables are those that can be decided before any uncertainty is realised. Once the random events impact the system improvements can be made by selecting second stage or recourse variables at some cost. The objective is to minimize the sum of the first stage costs and the expected second stage costs. For problems with continuous parameter distributions convexity properties of the recourse function have been used by Infanger (1994); and Shapiro and Homem-de-Mello (1998) to develop sampling based decompositions and approximation schemes.

Sand and Engell (2004) use a two-stage stochastic integer programming model on a moving horizon to schedule a flexible chemical batch process. They found that most previous work used very defensive strategies to generate robust off-line scheduling which require little emphasis on adjustment in real-time. They look at recourse actions as further opportunity for optimisation.

The system is affected by four types of uncertain processing times; deviation in product quality; machine breakdowns; product demand changes. A tree of possible future schedule horizons was computed to find the best first horizon schedule.

Engels and Karger (2003) look at the objective of minimizing the sum of the weighted completions times of jobs scheduled plus the sum of jobs rejected. Both these papers use a dummy machine to place all those jobs that are not scheduled. This machine is not constrained in the same way as the normal machines to allow any job to be placed on it.

There are numerous factors and systems that influence patient flow (eg: number and rate of patient arrivals, number of beds, length of stay, staffing arrangements, etc.). Patient arrivals fall under two groups, elective and emergency. In the case of hospital the number of patients from the elective and emergency groups is approximately the same. Elective patients are known few weeks in advance. On top of this we have batches of these patients arriving at regular times on weekdays. Emergency patients on the other hand give little or no warning of their need to use the ICU. This uncertainty adds complexity to the scheduling of patients.

The aims of this scheduling model are to: normalise utilisation of the ICU at an increased level; reduce rejection and reschedule rates of patients; and allow real-time adjustment for stochastic arrivals.

Another aspect of the model that adds complexity and sets it apart from many conventional scheduling problems is that some patients may be denied service by the ICU because of insufficient resources.

2. THE MODEL

The model describes the deterministic scheduling of one time window within the system. In the case of the ICU there will be an infinite number of these time windows as the unit never stops processing patients. While the schedule created may be satisfactory for some period of time there will always be unexpected events to manage. To manage these unexpected events we need to create a reactive system made up of many deterministic scheduling steps to keep the ICU running optimally.

Patients are divided into fixed and flexible patients. Fixed patients are those patients currently being treated in the ICU or patients that are in the previous schedule and soon to arrive or fixed for other reasons. Admission time of fixed patients can not be changed. Flexible patients are all other patients, either currently in the previous schedule or newly arriving and their admission time may change.

An integer programming for generating a schedule is developed. The objective is to maximise the utilisation of the ICU and minimising the number of patients rejected. The size of the model is determined by the number of patients, beds and time intervals.

2.1 Notations

i : patients, $i \in \{1...I\}$

j : beds, $j \in \{1...J\}$

t : time, $t \in \{0,1,2,...,T\}$

f : number of flexible patients in the schedule

A_j : first available time of bed j

a_i^o : original admission time, $i > f$

a_i : real admission time of patient i

c_i : priority index of patient i

d_i : discharge time of patient i ($d_i = a_i + p_i \ \forall i$)

$D_j = \begin{cases} 1 & \text{if bed } j \text{ is available on weekends,} \\ 0 & \forall j \end{cases}$

$e_i = \begin{cases} 1 & \text{if patient } i \text{ is an elective patient} \\ 0 & \text{otherwise} \end{cases}$

E : maximum number of elective surgery patients schedule at any particular time

K : total idle time of the system

p_i : length of stay of patient i

r_i : arrival time of patient i

s_j : setup time of bed j

$x_{ijt} = \begin{cases} 1 & \text{if patient } i \text{ is admitted to bed } j \text{ at time} \\ & t, \ \forall i,j,t \ ; \ i \le f \\ 0 & \forall i,j,t \ ; \ t \ne a_i^o; \ i > f \end{cases}$

w_i : waiting time of patient i

w_i^M : maximum waiting time of patient i

2.2 Objective Functions

The objective is to minimise the number of rejected patients by fulfilling the objective to maximise the utilisation of the unit. By including the waiting time in the objective function (1) and scaling by their priority category we aim to minimise rejection of the higher priority patients. Objective function balances the waiting time against the length of stay of the patient, in effect rejecting the patient if the scaled waiting time becomes too large when compared to the length of stay. The following objective function is derived from the waiting time, priority and idle time functions.

$$Minimise \ \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=0}^{T} x_{ijt} c_i (p_i + w_i) \qquad (1)$$

2.3 Constraints

Maximum one patient is scheduled in bed j at time t. The inclusion of s_j allows for bed dependent setup times. When programming these constraints for a given t only include jobs that satisfy $t - p_i \geq 0$.

$$\sum_{i=1}^{I} \sum_{s_j = \max(t-p_i,0)}^{t-1} x_{ijs_j} \leq 1 \quad \forall j, t \tag{2}$$

Equation 3 and 4 ensure that each patient is scheduled once. If a patient is not scheduled Equation 3 is equal to 0, which means the patient is rejected.

$$\sum_{j=1}^{J} \sum_{t=0}^{T} x_{ijt} \leq 0 \qquad for \quad i = 1, ..., f \tag{3}$$

Equation 4 ensures that each patient is scheduled once. For patients that are fixed in the schedule constraint 4 is used

$$\sum_{j=1}^{J} x_{ija_i^o} = 1 \qquad i = f+1, f+2, ..., I \tag{4}$$

This constraint combined with setting all the other x_{ijt} values for these patients to 0 restricts any flexibility to just bed choice at their admission time. The admission time is calculated as follows. This will be 0 when the patient is not scheduled.

$$a_i = \sum_{j=1}^{J} \sum_{t=0}^{T} x_{ijt} t \quad for \ \forall i \tag{5}$$

Waiting time is the difference between the arrival time of patient i and their admission time. This value needs to be 0 when the patient is not scheduled. It is also required to limit the maximum waiting time of a patient to a fraction of their length of stay.

$$0 \leq w_i = \sum_{j=1}^{J} \sum_{t=0}^{T} x_{ijt} (t - r_i) \leq w_i^M \tag{6}$$

Equation 7 ensures that no patient occupies bed j before it becomes available. This constraint is irrelevant for the first time window that is scheduled, but is necessary for subsequent time windows when patients overlap.

$$\sum_{t=0}^{A_j-1} \sum_{i=1}^{I} x_{ijt} = 0 \quad \forall j \tag{7}$$

Equation 8 ensures that there are no more than the specified amount of elective surgery patients scheduled at one time.

$$\sum_{i=1}^{I} \sum_{j=0}^{J} x_{ijt} e_i \leq E \quad \forall t \tag{8}$$

Reducing the number of beds in use over the weekends is handled by setting the value of A_j for the beds that should not be used to either midnight of Sunday if that occurs in the scheduling window or to $T+1$ if it does not. This makes sure that the scheduling algorithm will not schedule any patients on these specific beds.

3. IMPLEMENTATION

The model above describes the deterministic scheduling of one time window within the system. In the case of the ICU there will be an infinite number of these time windows as the unit never stops processing patients. While the schedule created may be satisfactory for some period of time there will always be unexpected events to manage. To manage these unexpected events we need to create a reactive system made up of many deterministic scheduling steps to keep the ICU running optimally.

This model was solved with the aid of the CPLEX libraries. Using the input file the software generates a list of patients that will arrive in the time window specified. Patient arrival times and length of stay values are generated and initially don't coincide with the time slots.

The software rounds the arrival times back and the length of stay values out to the closest time. When a list of patients is generated, it is used for differing number of beds to provide a sensitivity analysis to determine how much impact this factor has on the schedule. This can be measured in number of patients rejected, utilisation of the unit.

At time 0 there are no patients in the system so a warm-up period is used to bring the system to a steady state and from that point descriptive statistics can be taken from the schedule that is developed. A schedule is created with elective patients over the scheduling horizon. Subsequent schedules are generated when emergency patients arrive; the warm up period is over; and no emergency patients have arrived for the number of intervals specified by schedule length.

Patients are split into groups at the point of the expected arrival of the new patient. Basically we have those patients that have their admission time and/or bed fixed and those that are flexible. The patients that are currently being treated in the ICU have their bed and

admission time fixed. Patients that have not arrived yet may have their admission time fixed by the scheduler if their admission time has been changed a certain number of times, or if they are set to arrive within a certain amount of time. Flexible patients may have their admission time and bed changed or they are rejected from the system.

An initial schedule of fixed and flexible Patients is shown in Figure 1. This information is then used to determine bed availability times for subsequent schedules. Once the new schedule is found this information is fed into the next one and so on. The approach for this re-scheduling step is to generate a new schedule that is as close to the previous one as possible. Fixing patients to reduce the impact of future schedule changes on them is one part of this.
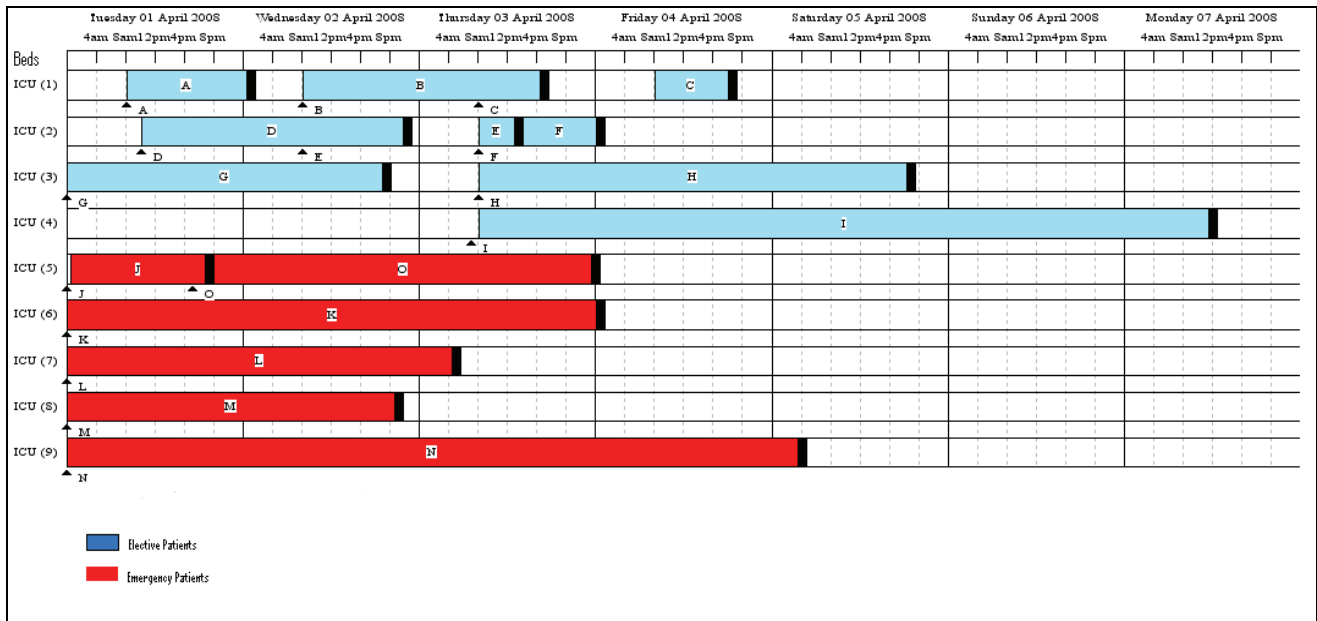
Figure 1: An initial schedule of fixed and flexible Patients

4. CONCLUSIONS

It was our aim to keep the model as general as possible to increase its applicability to other ICU's. The focus of the model is scheduling of humans or jobs that incur a penalty for job start time changes, while allowing for job rejections in a dynamic environment. This model could be modified to fit other units within the hospital, as well as external practices that handle patients in a similar manner.

The next step for this model is to determine the effects of the parameters in the objective function. We can investigate the balance between waiting time and length of stay time has on patient rejections to provide more information to the implementers of the system. The use of the scaling factor on the waiting time value may prove crucial in finding schedules that are good for the system they are applied to.

More investigation should to be done with the schedule window length in relation to patient input parameters and their effect on the objectives of the system. The results above show that there may be a link,

and therefore a different schedule length may need to be applied to different hospitals using the system.

The focus of the model is scheduling of humans or jobs that incur a penalty for job start time changes, while allowing for job rejections in a dynamic environment. This model could be modified to fit other units within the hospital, as well as external practices that handle patients in a similar manner.

REFERENCES

Engels, D. W. and Karger, D. R. (2003). Techniques for scheduling with rejection. *Journal of Algorithms,* 49, 175-191.

Infanger, G. (1994). *Planning under uncertainty: Solving large scale stochastic linear program*s. Boyd and Fraser Publishing Co.

Kim, S. and Horowitz, I. (2002). Scheduling hospital services: the efficacy of elective-surgery quotas. *Omega,* 30, 335-346.

Kim, S. and Horowitz, I. (1999). Analysis of capacity management of the intensive care unit in a hospital.

European Journal of Operational Research, 115, 36-46.

Kim, S. and Horowitz, I.(2000). Flexible bed allocation and performance in the intensive care unit. Journal of Operations Management, 18, 427-443.

McHardy, S. and Kozan, E. (2004). Modelling of Intensive Care Units and Operating Theatres in Public Hospitals. *The Fifth Asia-Pacific Industrial Engineering and Management Systems Conference,* Australia.

Ridge, J. C. and Jones, S. K. (1998). Capacity planning for intensive care units. *European Journal of Operational Research*. 105, 346-355.

Sahinidis, N. V. (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers and Chemical Engineering,* 28, 971-983.

Sand, G. and Engell, S. (2004). Modeling and solving real-time scheduling problems by stochastic integer programming. *Computers and Chemical Engineering*, 28, 1087-1103.

Shapiro, A. and Homem-de-Mello, T. (1998). A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming,* 81, 301-325.

AUTHORS BIOGRAPHY

Erhan Kozan is a Chair Professor of Operations Research, in the School of Mathematical Sciences at Queensland University of Technology, Australia. He has had 33 years industrial, managerial, teaching and research experience in the areas of Operations Research. He worked with the World Bank Group and the United Nations Development Program. He is the National President for the Australian Society for Operations Research. Professor Kozan has acted as principal investigator for over 22 long-term industrial projects, and 16 competitive national and international research grants in the area of health, finance, production, railways and seaports transportation. He is the author of a book, nine softwares and over 150 articles. He is the editor and associate editor of seven journals. He has supervised over 30 postgraduate research students. He is currently supervising five PhD students in the health and transportation area.

MODELLING AND DESIGN OF HOSPITAL DEPARTMENTS
BY TIMED CONTINUOUS PETRI NETS

Mariagrazia Dotoli[(a1)], Maria Pia Fanti[(a2)], Agostino Marcello Mangini[(a3)], Walter Ukovich[(b4)]

[(a)]Dipartimento di Elettrotecnica ed Elettronica, Polytechnic of Bari (Italy)
[(b)] Dipartimento di Elettrotecnica, Elettronica e Informatica, University of Trieste (Italy)

[(1)]dotoli@deemail.poliba.it, [(2)]fanti@deemail.poliba.it, [(3)]mangini@deemail.poliba.it, [(4)]walter.ukovich@deei.units.it

ABSTRACT
This paper proposes a model to describe in a concise and detailed way the flow of patients in a hospital starting from their arrival to the emergency medical service to the assignment of beds in the suitable department and finally the discharge. The model is based on a continuous Petri nets framework, whose fluid approximation allows us to define suitable optimization problems in order to plan the system capacity, e.g., determining the medical and nursing staff dimension and the number of beds. A case study and a simulation and optimization analysis show the efficiency of the model.

Keywords: hospital department, modeling, continuous Petri nets, performance evaluation.

1. INTRODUCTION
Providing high quality healthcare calls for improved organization and management in hospital departments. In such systems the main problems to face may be classified as follows (Xiong, Zhou, Manikopoulos 1994): i) dimensioning the system, i.e., determining the type and number of resources to provide (staff, rooms, beds, etc.); ii) understanding the workflow and detecting anomalies such as bottlenecks, waiting times, etc.; iii) improving efficiency, i.e., using resources in a better way, by decreasing patients length of stay, reacting to problems such as staff absence, etc.; iv) studying the system reactivity with respect to an increased workload.

Simulation and performance evaluation provides a useful tool for capacity planning and efficiency improvement. The hospital system may be effectively described as a Discrete Event System (DES) in order to perform discrete event simulation (Gunal and Pidd 2007, Kumar and Shim 2007). Moreover, Petri Nets (PNs) may be employed to model emergency medical services and hospitals (S.S. Choi, M.K. Choi, Song and Son 2005, Xiong, Zhou, Manikopoulos 1994). Indeed, PNs are analytical and graphical tools that are suitable for modeling asynchronous, concurrent processes in communication, computer and manufacturing systems. However, PN models suffer from the so called state explosion problem. One way to deal with such a problem is to use some kind of relaxation technique, in particular applicable to some discrete event models. Since hospitals can be considered DESs whose number of reachable states is very large, PN formalisms using fluid approximations provide an aggregate formulation to deal effectively with such complex systems, reducing the dimension of the state space (Silva and Recalde 2004).

This paper proposes a model to describe in a concise framework the flow of patients in a hospital starting from their arrival to the emergency medical service to the assignment of a bed in the suitable department and finally the discharge. The model is based on timed continuous PNs (Silva and Recalde 2004). In particular, places with finite capacities model the medical and nursing staff as well as the available beds and surgery theatres, while transitions describe the flow of patients and the operation/examination/treatment actions. The model provides an effective framework to analyze and simulate the workflow in a generic hospital department. Moreover, the fluid approximation allows us to define suitable optimization problems in order to optimize the chosen performance indices. For instance, this formulation provides a tool to determine the optimal number of beds, doctors and nurses to guarantee efficiency and good flow of discharged patients in the considered hospital department.

2. BASICS ON PETRI NETS

2.1. Discrete Petri Nets
A discrete PN is a bipartite graph described by the four-tuple $PN=(P, T, \textbf{Pre}, \textbf{Post})$, where P is a set of places with cardinality m, T is a set of transitions with cardinality n, $\textbf{Pre}$: $P \times T \rightarrow \mathbb{N}^{m \times n}$ and $\textbf{Post}$: $P \times T \rightarrow \mathbb{N}^{m \times n}$ are the *pre-* and *post-incidence matrices*, respectively, which specify the arcs connecting places and transitions. More precisely, for each $p \in P$ and $t \in T$ element $\textbf{Pre}(p,t)$ ($\textbf{Post}(p,t)$) is equal to a natural number indicating the arc multiplicity if an arc going from p to t (from t to p) exists, and it equals 0 otherwise. Note that $\mathbb{N}$ is the set of non-negative integers.

The *m×n incidence matrix* of the net is defined as follows:

$$\mathbf{C} = \mathbf{Post} - \mathbf{Pre}. \tag{1}$$

Given a PN, for each place $p \in P$ the following sets of transitions may be defined: $^\bullet p = \{t \in T: \mathbf{Post}(p,t) > 0\}$, named pre-set of p; and $p^\bullet = \{t \in T: \mathbf{Pre}(p,t) > 0\}$, named post-set of p. Analogously, for each transition $t \in T$ the following sets of places may be defined: $^\bullet t = \{p \in P: \mathbf{Pre}(p,t) > 0\}$, named pre-set of t; and $t^\bullet = \{p \in P: \mathbf{Post}(p,t) > 0\}$, named post-set of t.

The state of a PN is given by its current marking, which is a mapping $\mathbf{m}: P \rightarrow \mathbb{N}^m$, assigning to each place of the net a nonnegative number of tokens. A PN system $\langle PN, \mathbf{m}_0 \rangle$ is a net *PN* with an initial marking $\mathbf{m}_0$.

A transition $t \in T$ is enabled at a marking $\mathbf{m}$ if and only if (iff) for each $p \in {}^\bullet t$, it holds:

$$\mathbf{m}(p) \geq \mathbf{Pre}(p,t) \tag{2}$$

and we write $\mathbf{m}[t\rangle$ to denote that $t \in T$ is enabled at marking $\mathbf{m}$. When fired, t produces a new marking $\mathbf{m}'$, denoted by $\mathbf{m}[t\rangle\mathbf{m}'$ that is computed by the PN state equation:

$$\mathbf{m}' = \mathbf{m} + \mathbf{C}\,\vec{t}\,, \tag{3}$$

where $\vec{t}$ is the firing vector.

Let σ be a sequence of transitions (or firing sequence). The notation $\mathbf{m}[\sigma\rangle\mathbf{m}'$ indicates that the sequence of enabled transitions σ may fire at $\mathbf{m}$ yielding $\mathbf{m}'$. We also denote $\sigma: T \rightarrow \mathbb{N}^n$ the firing vector associated with a sequence σ, i.e., $\sigma(t) = v$ if transition t is contained v times in σ.

A marking $\mathbf{m}$ is said reachable from $\langle PN, \mathbf{m}_0 \rangle$ iff there exists a firing sequence σ such that $\mathbf{m}_0[\sigma\rangle\mathbf{m}$. The set of all markings reachable from $\mathbf{m}_0$ defines the reachability set of $\langle PN, \mathbf{m}_0 \rangle$ and is denoted by $R(PN, \mathbf{m}_0) = \{\mathbf{m} | \sigma: \mathbf{m}_0[\sigma\rangle\mathbf{m}\}$.

2.2. Continuous Petri Nets
This section recalls some basic definitions on the Continuous PN (ContPN) formalism used in this paper. For additional details the interested reader is referred to (Silva and Recalde 2004).

ContPNs are a straightforward relaxation of discrete PNs. Unlike discrete PNs, the amount in which a transition can be fired in ContPNs is not restricted to a natural number. The structure of a ContPN is identical to that of a discrete PN. However, the initial marking $\mathbf{m}_0$ is a vector of non negative *real* numbers. A transition t is enabled at $\mathbf{m}$ iff $\forall p \in {}^\bullet t$, $\mathbf{m}(p) > 0$. The enabling degree of t is:

$$\mathrm{enab}(t,\mathbf{m}) = \min_{p \in {}^\bullet t}\left\{\frac{\mathbf{m}(p)}{\mathbf{Pre}(p,t)}\right\} =$$
$$= \max\left\{k \in \mathbb{R}_0^+ \mid k \cdot \mathbf{Pre}(\cdot,t) \leq \mathbf{m}\right\}, \tag{4}$$

with $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$ and t can fire in a certain amount $\alpha \in \mathbb{R}$, with $0 \leq \alpha \leq \mathrm{enab}(t,\mathbf{m})$ leading to a new marking $\mathbf{m}' = \mathbf{m} + \alpha \cdot \mathbf{C}(\cdot,t)$, where the incidence matrix $\mathbf{C}$ is also called the *token flow matrix*. If $\mathbf{m}$ is reachable from $\mathbf{m}_0$ by the firing of a sequence σ, the fundamental equation $\mathbf{m} = \mathbf{m}_0 + \mathbf{C} \cdot \sigma$ can be written, where $\sigma \in \mathbb{R}_0^{+n}$ is the firing count vector associated with σ.

2.3. Timed Continuous Petri Nets
In this subsection, timing constraints are added to ContPNs. Time can be associated with places, transitions or arcs. In this paper we assume that time is associated with transitions and the following definitions specify and characterize timed ContPNs.

Definition 1: A timed ContPN $\langle PN, \boldsymbol{\lambda} \rangle$ is the untimed ContPN *PN* together with a vector $\boldsymbol{\lambda} \in \mathbb{R}^{+n}$, where $\boldsymbol{\lambda}[t_i] = \lambda_i$ is the firing rate of transition t_i.

Definition 2: A timed ContPN system is a tuple $\Sigma = \langle PN, \boldsymbol{\lambda}, \mathbf{m}_0 \rangle$, where $\langle PN, \boldsymbol{\lambda} \rangle$ is a timed ContPN and $\mathbf{m}_0$ is the initial marking of the net.

The fundamental equation describing the timed ContPN system evolution explicitly depends on time τ and is as follows: $\mathbf{m}(\tau) = \mathbf{m}_0 + \mathbf{C} \cdot \boldsymbol{\sigma}(\tau)$. Taking the derivative of this equation with respect to time, we obtain $\dot{\mathbf{m}}(\tau) = \mathbf{C} \cdot \dot{\boldsymbol{\sigma}}(\tau)$. Using the notation $\boldsymbol{f}(\tau) = \dot{\boldsymbol{\sigma}}(\tau)$ to represent the flow of transitions with respect to time, the state equation becomes:

$$\dot{\mathbf{m}}(\tau) = \mathbf{C} \cdot \boldsymbol{f}(\tau). \tag{5}$$

Depending on the flow definition, different semantics have been defined in the literature for continuous timed transitions, the two most important ones being the so-called *infinite server* (or *variable speed*) and *finite server* (or *constant speed*) semantics (Recalde and Silva 2001). The two semantics correspond to two different approximations of the discrete net system that the ContPN relaxes. For a broad class of nets it is formally proven by Mahulea, Recalde and Silva (2006) that the infinite server semantics always provides a better approximation than the finite server one. Hence, in this paper we consider the infinite server semantics.

Under the infinite server semantics, the flow f_i through a timed transition t_i is the product of its speed $\lambda[t_i]$ and its instantaneous enabling degree, as follows:

$$f_i = \boldsymbol{f}\left[t_i\right] = \lambda_i \cdot \mathrm{enab}(t_i, \mathbf{m}) = \lambda_i \cdot \min_{p_j \in {}^\bullet t_i} \left\{ \frac{m_j}{\mathbf{Pre}(p_j, t_i)} \right\} \quad (6)$$

where $m_j = \mathbf{m}(p_j)$ is the marking of place p_j.

2.4. Optimization of Timed Continuous Petri Nets

The use of timed ContPNs to model a DES allows us to consider off-line problems in which, given the system configuration, the objective is to optimally parameterize it. Among the problems belonging to this class are those devoted to the minimization of a cost function that may be formulated in linear terms with respect to the initial marking elements, i.e., as a weighting of the initial marking $\mathbf{b} \cdot \mathbf{m}_0$, where $\mathbf{b}$ represents a gain vector (e.g., if $\mathbf{m}_0(p_1)$ is to be minimized, $\mathbf{b}(p_1)=1$, while the rest of the weights of the gain vector should be zero). This kind of optimization problems, under some conditions depending by the structure of the ContPN described in (Silva and Recalde 2004), admits a particularly elegant and efficient solution by solving a linear programming problem. Indeed, it is either possible to determine the exact value of the optimal initial marking or in the worse case to obtain an upper bound of $\mathbf{m}_0$.

3. THE CONTINUOUS PETRI NET MODEL OF THE HOSPITAL DEPARTMENT

3.1. The System Description

Figure 1 shows the scheme of the basic hospital workflow model. Patients arrive at the emergency department at random time instants. In general, the emergency department serves various patient categories, each characterized by a different degree of urgency. Typically, patients are classified according to four degrees of urgency: life threatening, urgent, serious, non urgent.

Incoming patients are immediately registered and subsequently redirected to a suitable department on the basis of a performed diagnosis. However, life threatening cases are permitted to by-pass the registration and are generally treated in the surgical department. These patients are operated by the doctors and, if the operation is successful, they are subsequently considered as urgent cases and have to wait for the assignment of a bed place.

Usually, other patients wait in a waiting area until the staff is available for registration. After registration, registered patients have to wait in the patient waiting area until the treatment area and associated staff are available. Several non urgent patients do not need to be hospitalized, while other patients are assigned to a suitable department where they wait for a bed to be assigned. After the bed assignment, patients wait in another queue to receive the examination and prescription by the doctors.
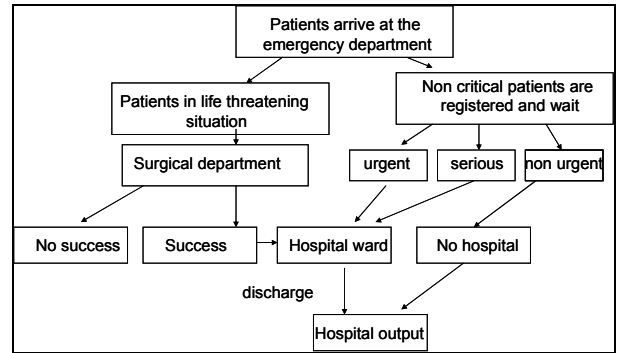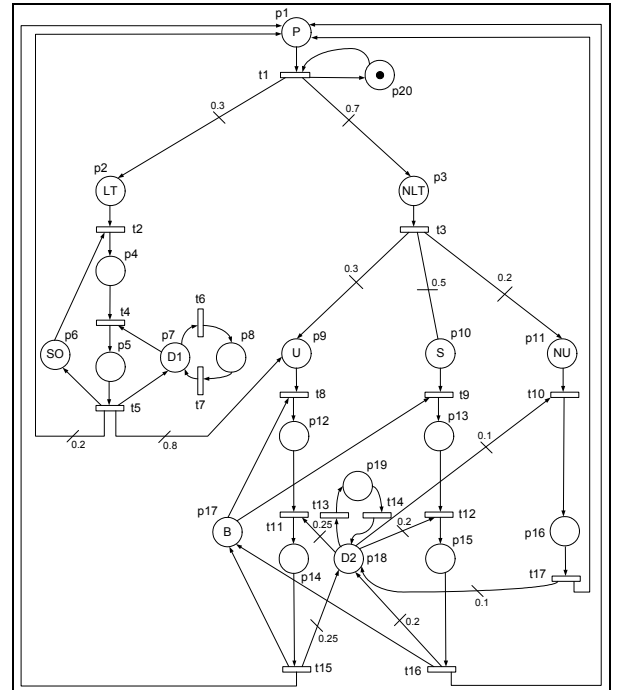

Figure 1. The basic hospital workflow.

Figure 2. The ContPN model of the patient flow in a typical hospital department.

3.2. The System Model by Timed ContPN

We model the hospital department system in the timed ContPN framework (Silva and Recalde 2004), where places with finite capacities model the staffs of doctors and nurses as well as the available beds. In addition, transitions describe the flow of patients and the actions of doctors and nurses.

The ContPN represented in Figure 2 models the process steps of the patient flow. Marking m_1 represents the number of patients that enter the hospital department. Place p_{20} is added to make the flow through the timed transition t_1 constant and equal to λ_1 that represents the speed with which patients enter the department. Some patients (in our case a percentage of the total that is assigned equal to 30%) are in a life threatening situation (marking m_2), others (marking m_3) have to wait for the registration that is represented by the timed transition t_3. The life threatening patients wait for the operating room to be ready (transition t_2) and successively for the availability of doctors (transition t_4). Markings m_7 and m_8 represent respectively the

number of ready to operate and absent surgeons, whereas m_6 represents the number of available beds in the operating room. When the operation is finished (transition t_5) some patients can be dead (in our case a percentage of 20% of the total) while other patients, no longer in a life threatening situation, are now considered as urgent.

As regards the patients that are not in a life threatening situation, after registration they are divided into three degrees of urgency: urgent (marking m_9), serious (marking m_{10}), non urgent (marking m_{11}). The average number of urgent, serious and non urgent patients is given by the weights of the arcs exiting from transition t_3 (in our case corresponding to percentages of 30%, 50% and 20%). The urgent and serious patients have to be hospitalized, hence they have to wait for a bed place (transition t_8 or t_9) and have to be assigned to a doctor to be examined (transition t_{11} or t_{12}). A doctor can be assigned either to urgent patients (in our case to 4 patients at most, since the weight of the arc exiting from place p_{18} is assumed 1/4=0.25) or to serious patients (5 patients at most). When patients recover, they exit from the department (transition t_{15} or t_{16}).

In addition, non urgent patients do not need to be hospitalized, but they only have to wait for the assignment of a doctor (transition t_{10}). In the considered department, a doctor can be assigned to 10 non urgent patients at most. Transition t_{17} represents the exit of non urgent patients from the department.

Finally, markings m_{17} and m_{18} represent the number of available beds and doctors in the department, respectively, whereas marking m_{19} represents the number of absent doctors.

3.3. The Simulation and Optimization of the System

This section presents a system capacity design problem that aims to optimize suitable performance indices on the basis of the defined ContPN model of the hospital. A simulation analysis applies and verifies the obtained policy.

The objective of the considered design problem is establishing the suitable number of doctors, surgeons, beds and surgery theatres in order to obtain good values of some selected performance indices. In our model this means to determine the initial markings of the capacity places p_{18}, p_7, p_{17} and p_6, respectively.

The chosen performance indices are the following: the number of operations per time unit (t.u.), the number of discharged urgent patients per t.u., the number of discharged serious patients per t.u. and the number of discharged non urgent patients per t.u. Hence, we impose a lower bound of the cycle time of the corresponding transitions, i.e., t_5, t_{15}, t_{16} and t_{17}, respectively.

We show that the considered design problem can be solved by defining the following programming problem (Silva and Recalde 2002):

$$\min \mathbf{b} \cdot \boldsymbol{\mu}_0 \qquad (7)$$

$$\text{s.t.} \begin{cases} \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \mathbf{C} \cdot \boldsymbol{\sigma} \\ \boldsymbol{\phi}[t] \le \lambda[t] \cdot \dfrac{\boldsymbol{\mu}[p]}{\mathbf{Pre}(p,t)} \quad \forall t \in T_S, \ \forall p \in {}^\bullet t \\ \boldsymbol{\phi}[t] \le \lambda[t] \cdot \dfrac{\boldsymbol{\mu}[p]}{\mathbf{Pre}(p,t)} \quad \forall t \in T_U, \ p = {}^\bullet t \qquad (8) \\ \mathbf{C} \cdot \boldsymbol{\phi} = 0 \\ \boldsymbol{\sigma}, \boldsymbol{\mu}_0, \boldsymbol{\phi} \ge 0 \\ \boldsymbol{\phi}[t_i] \ge 1/\Gamma_i \end{cases}$$

where $\mathbf{b}$ is the cost vector of the objective function, while $\boldsymbol{\mu}$, $\boldsymbol{\mu}_0$ and $\boldsymbol{\phi}[t]$ are respectively the approximations of marking $\mathbf{m}$, of the initial marking $\mathbf{m}_0$ and of flow $\mathbf{f}[t]$. T_U is the set of transitions with one input place, while T_S is the set of remaining transitions, in which synchronizations are present. Moreover, Γ_i represents the cycle time of transition t_i.

In particular, the constraints (8) follow from the definition of the enabling conditions and of the state equations (Silva and Recalde 2002).

Moreover, we add the following constraints on the initial markings: i) $\boldsymbol{\mu}_0[p_1]$ is equal to P, i.e., the population of patients using the medical service (we assume P=5000); ii) $\boldsymbol{\mu}_0[p_{20}]=1$ to impose at each time instant that flow $\mathbf{f}[t_1]$ is constant and equal to λ_1; iii) the remaining initial markings are set equal to zero. Formally, the following constraints are added to (8):

$$\begin{cases} \boldsymbol{\mu}_0[p_i] = 0 \quad \text{for i=2,...,5,8,...,16,19} \\ \boldsymbol{\mu}_0[p_1] = P \qquad (9) \\ \boldsymbol{\mu}_0[p_{20}] = 1 \end{cases}$$

Furthermore, the last constraints of (8) impose lower bounds of the transitions that represent the selected performance indices of the model. Hence, the flows of transitions t_5, t_{15}, t_{16} and t_{17} are forced in (8) as follows:

$$\begin{cases} \boldsymbol{\phi}[t_5] \ge 0.90 \\ \boldsymbol{\phi}[t_{15}] \ge 1.35 \\ \boldsymbol{\phi}[t_{16}] \ge 1.05 \qquad (10) \\ \boldsymbol{\phi}[t_{17}] \ge 0.40. \end{cases}$$

As previously specified, the aim of the design problem is choosing the minimum number of available doctors, surgeons, beds and surgery theatres in order to satisfy the chosen constraints, in particular equations (10). Consequently, the objective function (7) has to minimize the initial markings $\boldsymbol{\mu}_0[p_6]$, $\boldsymbol{\mu}_0[p_7]$, $\boldsymbol{\mu}_0[p_{17}]$ and $\boldsymbol{\mu}_0[p_{18}]$ that are weighted by a suitable cost vector $\mathbf{b}$. In particular, we assume that the cost of an operating theatre is higher than that of a surgeon, which is in turn higher than that of a doctor, or of a bed

setting $b_6=20$, $b_7=10$, $b_{17}=1$, $b_{18}=5$. Hence, the objective function (7) is specified as follows:

$$\min\left(20\mu_0[p_6]+10\mu_0[p_7]+\mu_0[p_{17}]+5\mu_0[p_{18}]\right). \quad (11)$$

Table 1 reports the firing rates of the transitions. The solution of the programming problem (11)-(8)-(9)-(10) provides the following initial markings:

$$\begin{cases} \mu_0[p_6]=2.4750 \\ \mu_0[p_7]=2.2500 \\ \mu_0[p_{17}]=43.1667 \\ \mu_0[p_{18}]=9.7230 \end{cases} \quad (12)$$

Obviously, since the initial markings m_6, m_7, m_{17} and m_{18} have to be defined by integer quantities, we set:

$$\begin{cases} \mathbf{m}_0[p_6]=3 \\ \mathbf{m}_0[p_7]=3 \\ \mathbf{m}_0[p_{17}]=44 \\ \mathbf{m}_0[p_{18}]=10 \end{cases} \quad (13)$$

Table 1: Transition firing rates.

Transition	Firing rate λ [time units/day]
t_1	3
t_2	4
t_3	3
t_4	2
t_5	0.5
t_6	1
t_7	10
t_8	3
t_9	2
t_{10}	0.8
t_{11}	1
t_{12}	0.9
t_{13}	1
t_{14}	10
t_{15}	0.05
t_{16}	0.08
t_{17}	4
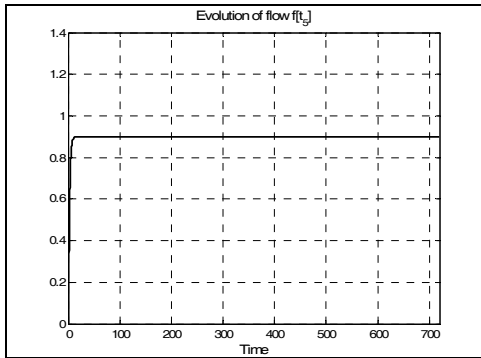
Figure 3: Evolution of flow $f[t_5]$ of finished operations (transition t_5).
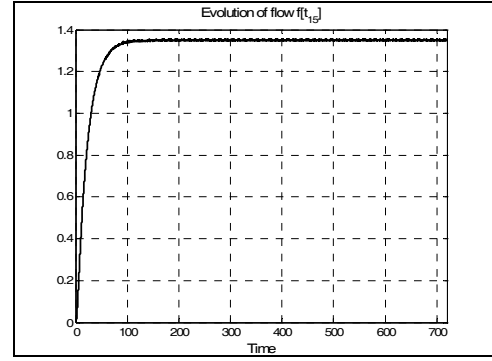
Figure 4: Evolution of flow $f[t_{15}]$ of urgent discharged patients (transition t_{15}).
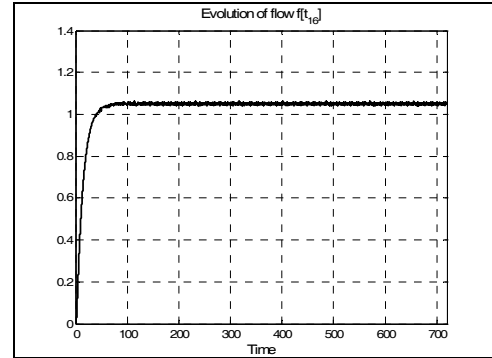
Figure 5: Evolution of flow $f[t_{16}]$ of serious discharged patients (transition t_{16}).
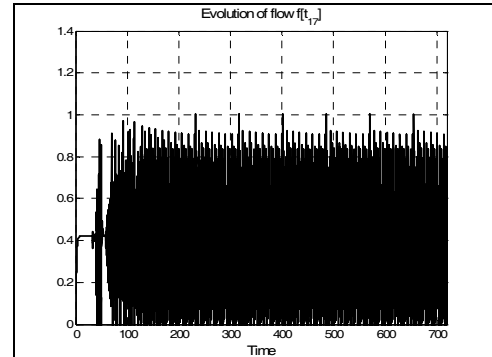
Figure 6: Evolution of flow $f[t_{17}]$ of non urgent discharged patients (transition t_{17}).
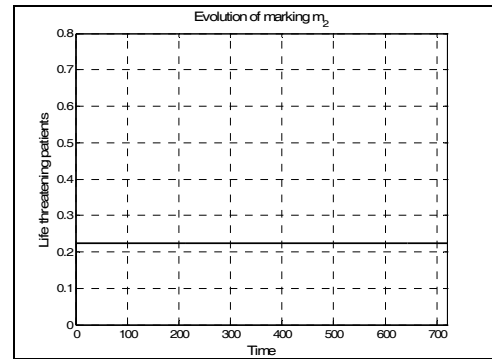
Figure 7: Marking $\mathbf{m}[p_2]$, number of patients in life-threatening situation waiting for treatment.
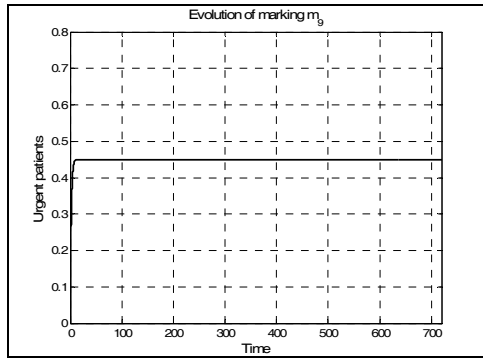
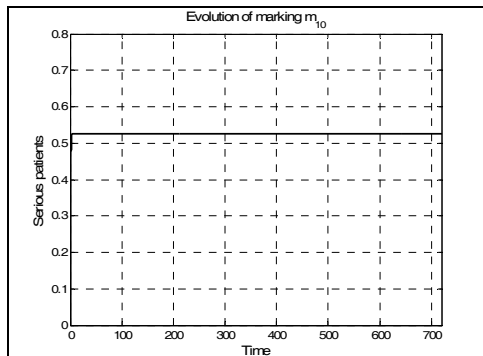Figure 8: Marking $\mathbf{m}[p_9]$, number of urgent patients waiting for treatment.

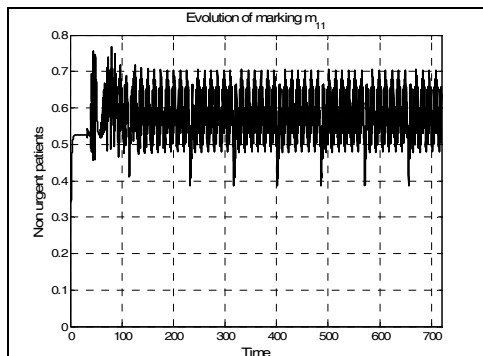Figure 9: Marking $\mathbf{m}[p_{10}]$, number of serious patients waiting for treatment.

Figure 10: Marking $\mathbf{m}[p_{11}]$, number of non urgent patients waiting for treatment.

Successively, the system dynamics is analyzed via numerical simulation in the MATLAB environment (The Mathworks 2006), a well-known and efficient software that allows us to model systems with a large number of places and transitions. Moreover, such a matrix-based software appears particularly appropriate for simulating the dynamics of ContPNs based on the matrix formulation of the marking update. Furthermore, the MATLAB software is able to integrate modeling and simulation of dynamical systems with the execution of control and optimization algorithms.

A one month simulation (with run time 720 t.u. if we associate one hour to one t.u.) of the ContPN model leads to determine the flows $f[t_i]$ with i=5,15,16,17. We show that each flow reaches a value close to the corresponding imposed bound (10) (see Figures 3 to 6). Note that $f[t_{17}]$ oscillates at steady state, but its average value equals 0.4, in accordance with (10). Hence, the simulation shows that the initial markings provided by the solution of the programming problem are appropriate to reach the objective values of the performance indices. Figures 7 to 10 show the average number of patients (respectively in life threatening situation, urgent, serious and non-urgent) still to be treated. Note that in each category there is on average always less than one waiting patient, showing the success of the system capacity design procedure.

4. CONCLUSION

We propose a continuous Petri net model for analyzing and simulating a generic hospital department workflow, starting from the arrival of patients to their discharge. The fluid approximation allows us to define suitable optimization problems in order to determine the optimal value of key hospital parameters. In particular, we consider the planning of the optimal number of operating theatres, beds, doctors and nurses to guarantee efficiency and minimize waiting times.

5. REFERENCES

Choi, S. S., Choi, M.K., Song, W.J., Son, S.H., 2005. Ubiquitous RFID Healthcare Systems Analysis on PhysioNet Grid Portal Services Using Petri Nets. *Proceedings of IEEE ICICS*, 1254-1258.

Gunal, M.M., Pidd, M., 2007. Interconnected DES Models of Emergency, Outpatient and Inpatient Departments of a Hospital. *Proceedings of the 2007 Winter Simulation Conference*, 1461- 1465.

Kumar, A., Shim S.J., 2007, Eliminating Emergency Department Wait by BPR Implementation. *Proceedings of the 2007 IEEE IEEM,* 1679-1683.

Mahulea, C., Recalde, L., Silva, M. 2006. On Performance Monotonicity and Basic Servers Semantics of Continuous Petri Nets. *Proceedings of WODES'06: 8th International Workshop on Discrete Event Systems*, Ann Arbor, USA, July, 345–351.

Recalde, L., Silva, M. 2001. Petri Nets Fluidification Revisited: Semantics and Steady State. *European Journal of Automation* APII-JESA 35, 435–449.

Silva, M., Recalde, L., 2004. On the Fluidification of Petri Nets: from Discrete to Hybrid and Continuous Models. *Annual Reviews in Control*, 28 (2), 253-266.

Silva, M., Recalde, L., 2002. Petri Nets and Integrality Relaxations: a View of Continuous Petri Nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 32 (4), 314–327.

The Mathworks, 2006. *MATLAB Release Notes for Release 14*. Natick, MA: The Mathworks.

Xiong, H.H., Zhou, M.C., Manikopoulos, C.N., 1994. Modeling and Performance Analysis of Medical Services Systems Using Petri Nets. *Proc. IEEE International Conference on Systems, Man and Cybernetics,* 2339-2342.

NETWORKS OF QUEUES WITH MULTIPLE CUSTOMER TYPES: APPLICATION IN EMERGENCY DEPARTMENT

Jihène. JLASSI [(a)], Abederrahman. EL MHAMEDI [(b)] and Habib. CHABCHOUB[(c)]

[(a), (c)] ISGI de Sfax Route Mharza Km 1.5-B.P954- 3018 Sfax., Tunisia
[(b)] MGSI (Modélisation et Génie des Systèmes Industriels) IUT de Montreuil, 140 Rue de la Nouvelle France- 93100 MONTREUIL.

[(a)] jihene2000tn@yahoo.fr   [(b)] a.elmhamedi@iut.univ-paris8.fr

ABSTRACT

Emergency department becomes a useful way to the access to hospital and it is a subject of study for many researchers. The research developed in this paper aims to improve the performance of the emergency department(ED) of Sfax Hospital by analytical method. So, a network of queues with multiple customer types is proposed. Different indicators of performance are used.

Keywords: Performance, Analytical method, networks of queues, multi class, priority patient class

1. INTRODUCTION

In general, evaluation techniques, that are methods by which performance evaluation indices are obtained, can be subdivided into two main categories: measurement (or empirical) techniques and modelling techniques. Empirical techniques require that the system or network to be evaluated exists and direct measurements of the evaluation target have to be taken. On the other hand, modelling techniques only require a model of the system. Modelling techniques are of two types: simulation and analytic.

The ED of hospital is a complex unit where the fight between life and death is always a hair's breath away, requiring a high degree of coordination and interrelations between human and material elements. (Jinn-Yi and Wen-Shan, 2007).

In the precedent paper (Jlassi et al, 2007), we proposed a simulation model which enables us to definite indicators to evaluate the performance of the ED of Sfax hospital. The study consisted of drawing a passage from a graphic model IDEF3x to a WITNESS model. Which reflect more the reality in a clearer way. Several numbers of indicators was defined. Different variation of the WITNESS model was proposed, showing the impact of variation in several parameters on the process performance.

This paper surveys the contributions and applications of queuing theory in the Emergency Department of Sfax hospital. It is organized as follows. We start with describing queuing theory and queuing network. In section3, we present literature review, in section 4, we describe the Emergency Department In section 5 we illustrate the proposed approach. Finally the last section presents concluding remarks and perspectives for future research.

2. QUEUING THEORY

Queuing theory has been a prominent analytical technique in operations research for more than half a century. A queuing model can save valuable time by providing analytical short cuts thereby improving the timeliness of management interventions and be helpful in the process of external scrutiny (e.g. via the Health Care Commission). In practice it may only be possible to do some of these things because of data and other limitations. Queuing theory provides an analytical approach to showing the variability in processes.

The theory permits the derivation and calculation of several performance measures including the average waiting time in the queue or the system, the expected number waiting or receiving service and the probability of encountering the system in certain states, such as empty, full, having an available server or having to wait a certain time to be served.

In many applications it is desirable to give certain classes of customers' preferential treatment; the queue is ordered and the higher priority customers are served first.

In a priority queueing system customers are divided into K ¸ 2 classes numbered 1; 2,…, K:

The lower the priority, the higher the class number. In other words, priority i customers are given preference over priority j customers if i < j.

The model with multiple stations is called a queueing network model. A variety of queueing network frameworks have been developed to represent various system mechanisms.

3. LITERATURE REVIEW

While there are many published healthcare research studies that apply queuing models, those that incorporate an Emergency Department concept are rare.

The study of Linda V. et al (2005) illustrates how data analysis and queuing models can be used to identify staffing changes that can decrease the delays in being seen by a provider and, thus, the fraction of patients who leave without being seen, without necessarily increasing capacity. It also highlights the need to establish patient delay standards, preferably by triage class, and to establish information systems to collect and track data on provider service times and patient delays in seeing a provider.

William et al (2005) explore conceptual development and practical application of the spreadsheet model. Particular features include: constructing lookup tables by hour of day containing estimates of minimum and maximum rates, using the rand between function to randomly select model inputs from a uniform distribution, developing frequency distributions to assist in output interpretation, illustrating conditional formatting, output graphing, etc. One can observe multiple samples of hourly patient fluctuations based on unit open beds and midnight census. Number of patients waiting can be shown at varying levels of system utilization. As utilization approaches approximately eighty percent, patient waiting time increases disproportionately. The spreadsheet model is a dynamic, visual illustration of how variation in individual process times can affect total process capability. Its use is primarily intended as a teaching tool for those new to simulation modeling.

In Siddharthan et al article (1996), Investigates the increased waiting time costs imposed on society due to inappropriate use of the emergency department by patients seeking non-emergency or primary care. Proposes a simple economic model to illustrate the effect of this misuse at a public or not-for-profit hospital. Provides evidence that non-emergency patients contribute to lengthy delays in the ER for all classes of patients. Proposes a priority queuing model to reduce average waiting times.

Marianov et al (1995), presented, in their paper, a model which seeks to maximize the population covered by emergency vehicles with availability &. A probabilistic formulation is structured in which availability is computed utilizing queueing theory.

Au-Yeung et al (2000) report's developed a multiclass Markovian queuing network model of patient flow in the Accident and Emergency Department of a major London hospital.

4. CONTEXT OF THE STUDY

Habib Bourguiba hospital is presented as follows:

- Academic Hospital centre since 1985 ;
- Erected in public establishment of health since 1993.
- The missions of the hospital are:
- To lavish current pathology cares and essentially cares of reference;
- To assure the convenient formation of basis and retrain the medical and decorate-medical personnel
- To develop the activity of research in the medical domain and cares male nurses.

The hospital includes 18 Departments; the most important one is the Emergency Department that represents the entrance door of the hospital.

The geographical situation of Sfax City in the centre of the country, point of link between the south and the north of the country makes that the Emergency Department receives an important number of casualties of the public way. As second economic and industrial pole sheltering a lot of companies, the city knows a rather elevated number of work casualties.

The department includes:

- 2 rooms of care;
- 1 room of plaster;
- 2 post offices;
- 1 room of general surgery ;
- 1 room of orthopaedic;
- 1 room for medical visits;
- the overage number of patients per day is $\pm$ 300,
- the number of personnel is 13;

The process by wish a patient passes is: For the arrival of the patient, we distinguish 2 possibilities: either he will come by himself or, in dangerous cases, he will be transported by the ambulance. At the arrival, the patient will pass by a triage process that will define the Emergency degree and then the process that will be taken by the patient. In fact, there are two processes that can be taken by the patient, given the result of the triage process.

In both cases, the remaining process will be almost the same. In fact, the main difference concerns the administrative process. Thus, for a dangerous case, this process will be reported at the end of the patient process while for a non dangerous case, this process will be at the beginning of the whole process.

The remaining tasks of the process are the same as those of a normal consultation process. In fact, the patient will see a doctor who will determine the patient's state and if necessary ask for supplementary analysis. Given the analysis results,

the doctor will either care the patient by him self or ask for a specialist. In both cases, the patient could be hospitalized or go home.

5. THE NETWORK MODEL

The queueing models considered here consist of J single-server stations and K customer types (Two patients types) , with each customer type having a fixed route through the network of stations. Customers of types k = 1, 2, . , K arrives to the network according to independent renewal processes. Each type k customer visits the sequence of stations

$$r_k(1), r_k(2), ..., r_k(s(k))$$

After which he leaves the system. This is an open network in the sense that all customers originate from external sources and all eventually leave the system. We require a feed forward routing pattern, meaning that the stations can be numbered j = 1, 2,. , J in such a way that for each k the route of a type k customer forms an increasing sequence. Intuitively this means that work flows in a directed fashion through the network. In particular, no customer may visit any work center more than once. The s (k) service times along the route of a given type k customer are allowed to have an arbitrary joint distribution, but are independent of the arrival processes and of the service times of any other customer. For each station j we define

$$C_j = \{ K : r_k(S) = j \quad 1 \le S \le S(k) \}$$

We call the C, the "constituency" of station j; it is the set of customer types that visit j at some stage of their routes. The queue discipline at station j is defined by a partition of C, into subsets Gj and Hj, which represent equivalence classes with respect to priority level. Customers of type's $K \in H_j$ (high priority) have pre-emptive resume priority over customers of types $K \in G_j$, at station j. Within a priority class, customers are served in first-come-first-served (FCFS) order regardless of type. At a station j where there is only one priority level (i.e., the discipline is simply FCFS), we set $G_j = C_j$, and $H_j = \phi$. There is no bound on the number of customers that can be queued at any station.

We take as primitive a probability space (Ω, F, P) on which the following sequences of nonnegative random variables and vectors are defined. For k = 1, .. . , K, let $\left\{ u_k^l \quad l \ge 1 \right\}$ be mutually independent i.i.d.

sequences with $E\left[u_k^l \right] = 1$. Independent = of these, we have K mutually independent i.i.d. sequences $\left\{ \widetilde{v}_k^l, \quad l \ge 1 \right\}$ of random vectors

$$\widetilde{v}_k^l = (v_{rk(1),k}^l, ..., v_{rk(s(k)),k}^l)$$

With $E\left[\widetilde{v}_k^l \right] = (1, ..., 1)$ we define

$$
\begin{aligned}
&\alpha_k^2 = \text{var}\left(u_k^l \right) \ge 0, \quad k=1, ...K \\
&\beta_{jk}^2 = \text{var}\left(v_{jk}^l \right) \ge 0, \quad j=1, ...J \quad k \in C_j \\
&\beta_{ijk} = \text{cov}\left(v_{ik}^l, v_{jk}^l \right) \quad j=1, ...J \quad k \in C_i \cap C_j
\end{aligned}
\tag{1}
$$

The $\left\{ u_k l \right\}$ represent normalized (i.e., unit mean) interarrival times, and the $\left\{ v_{jk}^l \right\}$ represent normalized service times. The actual interarrival and service times for the queueing model are constructed from these sequences by specifying constants $\lambda_k \ge 0$ for k = 1 , .. . , K, and $m_{jk} \ge 0$ for j = 1,. . . , J and $k \in C_j$. We then let $(\lambda_k)^{-1} u_k^l$; be the interarrival time for the lth type k customer, and let $m_{jk} v_{jk}^l$ be this customer's service time at station j on his route. In this scheme, the λ_k , are the average arrival rates, and the m_{jk} are the mean service times. Observe that α_k^2 is the squared coefficient of variation (the variance divided by the square of the mean) of the interarrival time distribution for type k customers, and β_{jk}^2 is the squared coefficient of variation of the type k service times at station j. The parameters β_{ijk}, which for given (i, j, k) represent the covariance of the type k service times at station i and station j divided by the product of their means, are the analogous normalizations of the covariance in the service time vectors.

Let the renewal counting processes $N_k = \{ N_k(t), \quad t \ge 0 \}$ be defined for k = 1 , .. . ,K by

$$N_t(t) = \begin{cases} \max\{l \geq 0; u_k^l + ... + u_k^l \leq t\}, u_k^l \leq t \\ 0 \qquad\qquad\qquad u_k^l \geq t \end{cases}$$

(2)

The arrival counting process A, for type k customers is now given by $A_k(t) = N_k(\lambda_k t),\quad t \geq 0$ (we assume that the network is empty at time t = 0). Next, for j = 1 , ... ,J and $k \in C_j$, define the compound processes L_{jk} by

$$L_{jk}(t) = \sum_{l=1}^{N_k(t)} v_{jk}^l, \quad t \geq 0$$

(3)

Then $m_{jk}\, l_{jk}(\lambda_k t)$ represents the amount of type k work for server j that arrives to the network during [0, t]. We define the "netput" process Xj at station j by

$$X_j(t) = (\sum_{k \in C_j} m_{jk}\, l_{jk}(\lambda_k t)) - t \quad t \geq 0$$

(4)

$$\hat{X}^n(t) = \varepsilon_n X^n(t/\varepsilon_n^2), \quad \hat{W}^n(t) = \varepsilon_n W^n(t/\varepsilon_n^2),$$

$$\hat{I}^n(t) = \varepsilon_n I^n(t/\varepsilon_n^2), \quad \hat{Z}^n(t) = \varepsilon_n Z^n(t/\varepsilon_n^2),$$

$$\hat{U}^n(t) = \varepsilon_n U^n(t/\varepsilon_n^2), \quad \hat{Q}^n(t) = \varepsilon_n Q^n(t/\varepsilon_n^2), \qquad 0 \leq t \leq 1,$$

$$\hat{S}^n(t) = \varepsilon_n S^n(t/\varepsilon_n^2),$$

Here $X_j(t)$ represents the total workload input for server j over [0, t] minus the work which would be completed if the server were never idle. The vector netput process X = (XI,...,X,) is a fundamental building block for our theorem. For j = 1,. ..,J let $I_j(t)$ denote the cumulative time during [0, t] that server j is idle. Then the process defined by

$$W_j(t) = X_j(t) + I_j(t), \quad t \geq 0 \quad (5)$$

Represents the total amount of work for server j present anywhere in the network at time t. However, not all of the customers whose service times are accounted for in Wj.(t) will reach station j by time t; some will be in queue or in service at stations preceding j on their routes. Let $Z_j(t)$ denote the unfinished workload which is actually present at station j at time t. By our last comment, we have

$$Z_j(t) \leq W_j(t), \qquad t \geq 0.$$

The central result of this paper is a heavy traffic limit theorem for the vector process Z(t) = (Z,(t), .. .,Z,(t)). It turns out that the J-dimensional limit process for this vector of station workloads contains all the information necessary to describe the heavy traffic behavior of the system at the customer level.

More precisely, let $Q_{jk}(t)$ denote the number of type k customers in queue or in service at station j at time t, and let $U_{jk}(t)$ denote unfinished work represented by these customers (of course, $\sum_k U_{jk}(t)$ is simply $Z_j(t)$). Also, let SK(t) denote the sojourn time in the network for the first type k customer arriving at time greater than or equal to t. Then heavy traffic limit results for the processes

$$Q(t) = (Q_{11}(t),...,Q_{1K}(t),...,Q_{J1}(t),...,Q_{JK}(t)),$$

$$U(t) = (U_{11}(t),...,U_{1K}(t),...,U_{J1}(t),...,U_{JK}(t)), \qquad t \geq 0,$$

$$S(t) = (S_1(t),...,S_K(t)),$$

can be obtained jointly with the result for Z, with the limit processes for Q, U and S expressed as linear transformations of the limit process for Z. The limit results are obtained under conditions of heavy traffic, which we now describe. Define

The parameters p_{jk} and p_j by:

$$p_{jk} = \lambda_k m_{jk} \quad j = 1,...,J \quad k \in C_j$$

$$p_j = \sum_{k \in C_j} p_{jk} \quad j = 1,...,J$$

(6)

The results gotten while applying the formulas of William Peterson (1991), present that the specialist physician and the generalist physician in the second

passage of the patients are the most occupied step by which pass a patient (see Appendix A). And the patients in these two steps of process wait a lot to be served. What confirms the results gotten by the simulation model by Witness.

It is why we propose to add a physician (specialist or generalist who has experience) to be able to minimize the waiting time and therefore to increase the number of patients treaties

6. CONCLUSION

This paper studied the evaluation of the emergency department of Sfax hospital by using the multi class queuing networks technique.

First, we presented the queuing network. Then, we provided the primordial role of Emergency Department (ED) and we identified numerous studies which have been made in order to evaluate the out put of the ED. Next we presented the ED of Sfax Hospital. Finally, we presented the main theorem.

The results showed the necessity to add in case of need of a physician. To choose between what physicians to add, we proposed to apply a method of multiple criteria decision aid as a PROMETHEE II method.

APPENDIX

In the Emergency Department there are two patients Types
- type 1 : no urgent
- Classe 2 : urgent case (high priority)

In the networks there are 6 steps:
T1: administrative procedure;
T2: generalist physician (first)
T3 : Analyze
T4 : Radiology ;
T5 : generalist physician (second)
T6 : specialist physician.

$$p_{jk} = \lambda_k m_{jk} \quad j=1,...,J \quad k \in C_j$$

$$p_j = \sum_{k \in C_j} p_{jk} \quad j=1,...,J$$

$$p_{11} = \lambda_1 m_{11} = 55,59$$

$$p_{12} = \lambda_2 m_{12} = 0$$

$$p_{21} = \lambda_1 m_{21} = 107.008$$

$$p_{22} = 71.25$$

$$p_{31} = 146.3$$

$$p_{32} = 95$$

$$p_{41} = 96$$

$$p_{42} = 85.5$$

$$p_{51} = 133$$

$$p_{52} = 95$$

$$p_{61} = 188$$

$$p_{62} = 118$$

$$R = I - \Gamma + \Gamma^2 + ... + (-1)^{J-1} \Gamma^{J-1}$$

Γ is strictly upper triangular guarantees that $(I + \Gamma) \equiv R$

$$\gamma_{ij} = \sum_{k \in G_i \cap C_j} p_{jk} / \sum_{k \in G_i} p_{ik} \quad i \le j$$

$$\Gamma = \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1.9 & 1 & 0 & 0 & 0 & 0 \\ 1231 & 6.32 & 1 & 0 & 0 & 0 \\ 17.8 & .91 & 0.14 & 1 & 0 & 0 \\ 11 & 5.66 & 0.89 & 6.18 & 1 & 0 \\ 22.8 & 11.71 & 1.85 & 12.79 & 2.069 & 1 \end{matrix}$$

$$Z_j^*(t) = \begin{bmatrix} 94 \\ 199.3 \\ 975.7 \\ 920.15 \\ 1023.25 \\ 209215 \end{bmatrix}$$

No urgent case

- First step:

$$Q_{jK}^*(t) = \frac{\lambda_k}{P_{jG}} Z_j^*(t) = \frac{1}{m_{jk}} U_{jk}^*(t)$$

$$Q_{11}^*(t) = 7.068$$

- Others :

$$Q^*_{ik}(t) = \frac{\lambda_k}{P_{iG}} Z^*_i(t) \delta_{ik}$$

$$\delta_{ik} = \begin{cases} 1 & \text{si } k \in G_i \\ 0 & \text{others} \end{cases}$$

$$Q^*_{21}(t) = 7.78$$

$$Q^*_{3.1}(t) = 27.93$$

$$Q^*_{41}(t) = 40.06$$

$$Q^*_{51}(t) = 32.15$$

$$Q^*_{61}(t) = 46.51$$

<u>High priority patient :</u>

$$Q^n_{jk}(t) = A^n_{jk}(t) - A^n_{jk}(T^n_j(t))$$

$$A_k(t) = N_k(\lambda_k t), \quad t \geq 0$$

$$N_t(t) = \begin{cases} \max\{l \geq 0; u^l_k + ... + u^l_k \leq t\}, u^l_k \leq t \\ 0 \qquad\qquad\qquad u^l_k \geq t \end{cases}$$

$$T^n_j = \frac{1}{p_{jG}} Z^*_j$$

$$E(X) = \frac{1}{N} \sum X_i$$

$$V(X) = (X_i - \overline{X})^2$$

Si $X_i \to N(\mu, \sigma)$

$$\frac{X - E(X)}{\sqrt{V(X)}} \to N(0,1)$$

<u>Waiting time :</u>

$$S^*_k(t) = \sum_{l=1}^{S(k)} S^*_{lk}(t) = \sum_{\{j: k \in G_j\}} \frac{Z^*_j}{P_{ij}}$$

$$S^*_k(t) = 1.69 + 1.86 + 9.58 + 7.69 + 11.12 = 31.94$$

REFERENCES

Carely. J. Jurishica. (2005), "Emergency department simulations: medicine for building effective models", Proceedings of the 2005 Winter Simulation Conference.

Green, L. (2006) Queueing analysis in healthcare, in Patient Flow: Reducing Delay in Healthcare Delivery, Hall, R.W., ed., Springer, New York, 281-308.

Jinn-Yi Yeh, Wen-Shan Lin. (2007), "Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department", Expert Systems with Application 32 (2007) 1073-1083

Linda V. Green, Joa~o Soares, James F. Giglio, Robert Alan Green, Society for Academic Emergency Medicine November 2005

William A. Stout, Jr. Barbara Tawney Proceedings of the 2005 Systems and Information Engineering Design Symposium Ellen J. Bass, ed.

William, Peterson; A heavy traffic limit theorem for networks of queues with multiple customer types. Mathematics of operations research, vol, 16, No1. Feb, 1991, pp 90-118.

CRITICAL NEWBORN TRANSPORT IN VENETO REGION: MODELS AND SIMULATION

Francesca Bortolato[(a)], Anna Ferrante[(b)], Giorgio Romanin-Jacur[(c)], Laura Salmaso[(d)]

[(a) (c)]Department of Management and Engineering, University of Padova, Italy
[(b) (d)] Epidemiology and Community Medicine Unit, Department of Paediatrics, University of Padova, Italy

[(a)]romjac@dei.unipd.it, [(b)]epi@pediatria.unipd.it, [(c)]romjac@dei.unipd.it, [(d)]epi@pediatria.unipd.it.

ABSTRACT

The paper studies the problems connected with the transport of critical newborns from their birth centre towards specialized assistance centres, where they can undergo suitable intensive cares. The scope is to analyze whether the employed resources are sufficient for a correct service course; in fact on one side the transport requires apt means and specifically trained personnel, on the other side the structures where critical newborns can be admitted and may receive intensive neonatal cares are few, shall be equipped with special instrumentation and require specialized staffs. Here a simulation model describing critical newborns transport in Veneto Region has been built up and has elaborated data collected in all birth centres. Simulation results revealed that resources of centres with neonatal intensive care units are often insufficient to satisfy all admission requests. The model, implemented in language MicroSaint, is easily extendable to other Italian or foreign regions presenting analogous situations.

Keywords: critical newborn, transportation, neonatal intensive care unit

1. INTRODUCTION

The paper is a part of an extended study, operated by Veneto Region in Italy, about the characters of pregnancy and birth course. The critical newborn transport is an integral component of this study.

Newborn transport is necessary whenever a critical newborn, which has been born in a non specialized birth centre, is in a severe condition and shall be urgently transferred in a centre which is able to supply high level cares.

An accurate bibliographic research pointed out that many papers have been written about the matter, see Cornette (2004), Das and Leuthner (2004), Fenton, Lesile and Skeoch (2004), Fowlie, Booth and Skeoch (2004), Lupton and Pendray (2004), Nicholson, Burr and Powell (2005), Rashid, Bhuta and Berry (1999), Wright (2000), but none of them have developed a model describing the critical newborn transport phoenomenon.

In this paper such a phoenomenon has been studied with simulation techniques. By the implemented model every request for a newborn transport, coming from a non specialized birth centre, is assigned to the nearest third level centre (able to supply neonatal intensive cares) which has resources at disposition. Obviously if the birth took place in a third level centre, then the admission request is assigned to the centre itself, if resources are at disposition. Model data have been obtained from regional data bases, in particular from "birth assistance papers" (CEDAP) and "hospital discharge papers" (SDO), related to newborn transfer requests and to births in third level centres. From model runs we were able to determine whether resources employed for critical newborn transport and admissions in third level centres are sufficient to satisfy all requests; we observed that third level centres have not sufficient resources at disposition to satisfy all admission requests.

The model was utilized related to Veneto Region situation in a given time interval, but it can be easily employed for other both Italian and foreign regions presenting similar problems.

The paper is organized as follows. First pregnancy-birth course is analyzed to evidence the causes of critical newborn birth. Then the newborn transport problem is treated in detail. Model data sources are considered and discussed. The model and its implementation is presented and accurately described. Eventually model results, related both to current situation and to possible changed situations to obtain a better service are examined.

2. PREGNANCY AND BIRTH COURSE

Pregnancy is the state of a woman who carries the product of conception; during pregnancy woman's organism undergoes many variations which are useful to assure the best condition for embryo installation, nutrition and growing. We may distinguish two types of pregnancy: physiological, which follows a natural course without complications, and pathological, with complications both for the mother and for the embryo (e.g., diabetes, malformations, infections, …). For what concerns birth, we may distinguish birth without complications, birth with expected complications (e.g.,

foetal-pelvic disproportion, placental deformity, etc.) and finally birth with non expected complications (e.g., foetal distress, haemorrage, premature birth, etc.).

On the basis of complications encountered during pregnancy and at the birth, newborns may be classified in three levels: first level concerns normal newborns, which need to get neonatal generic assistance; second level concerns newborns presenting light pathologies or mild prematureness, which need to get neonatal pathology assistance (incubator); third level concerns newborns requiring neonatal intensive care (ventilated incubator) by a specialized neonatal intensive care unit. Every birth centre is characterized by the assistance levels it can give: in most centres we have only the first level, in some specialized centres we have first and second level, in few highly specialized centres (called third level centres) we have all three levels.

Every pregnant woman is generally admitted to a birth centre where the assistance level is appropriate to the expected birth complications. Anyway sometimes newborn conditions may appear more severe than expected, so that the resources at disposition in the birth centre are not sufficient and an urgent transfer to a higher level centre is required.

3. CRITICAL NEWBORN TRANSPORT GENERAL CHARACTERISTICS

Initial cares supplied to a newborn represent one of the key factors which determine its survival; critical newborn transport is an integral part of such cares. The procedure starts when the birth hospital identifies the newborn emergency situation; the reasons which bring to decide the transport may be many, as follows (Fowlie, Booth and Skeoch 2004): unsuitable structures, insufficient medical staff, lack of places (full neonatal intensive care unit or specific care unit), unexpected birth far from home, transfer towards local structures. Obviously the structure where birth took place shall be able to supply at least base cares and a first stabilization while waiting for the transport team. Alerted transport team moves from its base towards the birth centre, carrying all transport equipment; once arrived at the birth centre it devotes a time interval, which may be up to one or two hours, to newborn stabilization, so to make it able to survive to the trip; then the transport takes place towards the nearest third level centre with admission resources at disposition; during the trip assistance is continuous; once arrived at the admission centre, the newborn is delivered to the centre team. If all places of arrival centre are busy at the arrival instant, the critical newborn is admitted in another department of the same hospital.

We may have two types of transport supply: "dedicated" and "on demand"; the dedicated service is made up of a specialized team using specialized means, whose only job is neonatal transport, with the advantages of timely response and continuous improvement of supplied cares; the on demand service is carried out by people usually engaged in other jobs and available for emergencies.

The transport team includes specially prepared personnel, trained to supply neonatal intensive cares under stress in difficult conditions, generally formed of medical doctors and nurses (of specialized nurses only in Canada and U.S.A.); the transport means may be different according to the distance and to the conditions of the place: the ambulance is generally apt for a maximum trip duration of 1.5 hours, the aircraft for long distances, over 350 km, and the helicopter for medium distances (with the advantage of overcoming obstacles and the disadvantage of reduced working space); the main transport equipment is an incubator with forced ventilation, provided of independent electrical supply, of all measurement sets and of all necessary gases, liquids and drugs. The transport shall generally be preceded by a suitable stabilization, operated by the transport team.

In Veneto Region births are about 45,000 per year, and take place in 41 birth centres, of which 8 are of third level: Padova with 10 places, Verona with 8, Treviso with 6, Vicenza, Camposanpiero and Rovigo with 4 places each, Mestre and Belluno with 2 places each. Two critical newborn transport services are working, in Padova and in Verona, each one serving a well defined part of the region. Every transport centre works on demand, and the team resources include: a special ambulance, with two serving persons, a complete incubator, a specialized neonatal doctor and a specialized nurse.

Every birth centre is equipped to supply assistance to the mother (gynaecologic-obstetric unit) and to the newborn (neonatology-paediatric unit), at one of three distinct assistance levels, corresponding to newborn possible pathologies, as seen above; every pregnant woman is generally admitted to a birth centre where the assistance level is appropriate to the expected birth complications, therefore the majority of critical newborns are born where they can be suitably assisted; anyway the 2-2.5% of newborns require a transport towards another centre because of: unforeseen risks, scarcity of places, necessity of special exams or of special treatments; a special transport is the return to the original birth centre once the acute phase is over (back transport).

4. DATA SOURCES

The Epidemiology and Community Medicine Unit manages a large amount of national and regional data concerning population health and health assistance.
The most interesting data for our work are contained in two papers, the CEDAP (birth assistance paper) and the SDO (hospital discharge paper). The CEDAP reports information about the parents, the pregnancy, the birth and the newborn. The SDO reports information about the patient, the hospital, the admission, the admission department (or departments), the disease and the cares supplied; it is emitted also for newborns. By use of suitable filters we have obtained: the number and the length of stay of third level newborns who have been born in the third level birth centres and have been

QUEUE

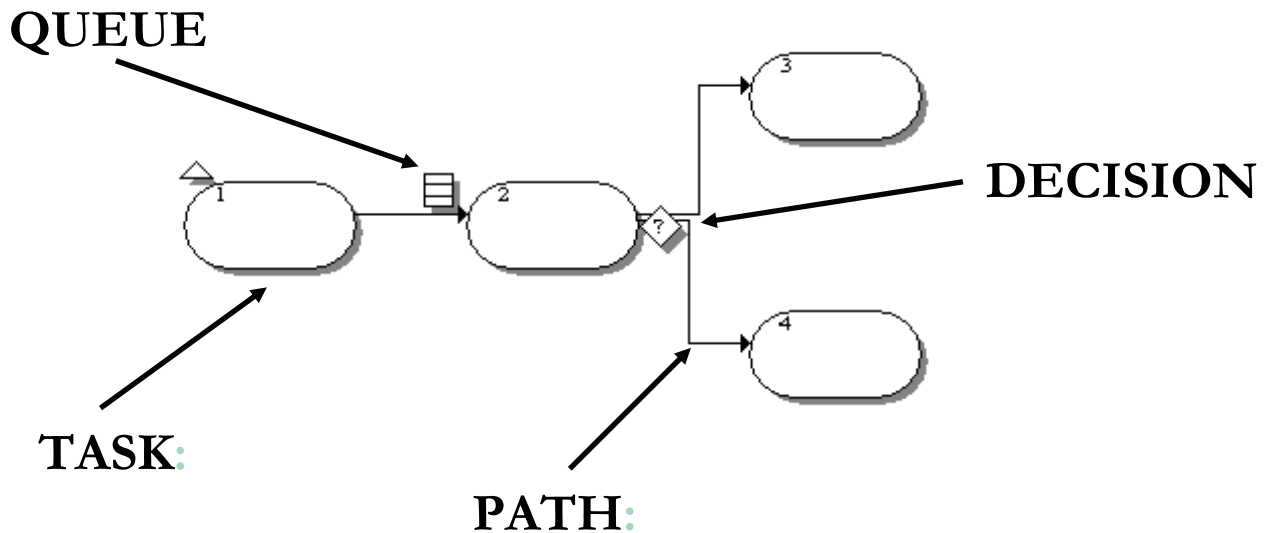**DECISION**

TASK:

PATH:

Fig. 1: Essential elements of graphic model representation by MicroSaint

admitted in the neonatal intensive care unit of the same centre; the number and the length of stay of newborns which have been transferred from other birth centres to third level centres (for which the day of discharge from the first centre is equal to the day of admission in the second centre). All above data are used to compute patient flows for transport and admissions.

From other orderly sources we could accurately compute motor car run times: between the two transport service bases and all birth centres; between all birth centres and all third level assistance centres; between all third level assistance centres and the transport service bases. Such run times are used to compute transportation times in the model.

5. THE MODEL

The model is devoted to analyse whether resources at disposition are sufficient for what concerns newborn transport and newborn admissions in the region. Therefore the model includes two main interacting parts: the first one concerns newborn transfers and the second one concerns newborn admissions. Newborn transfer considers the following actions: critical newborn generation in one of the birth centres of first or second level, wait for the transport team set up and trip, newborn stabilization in the birth centre, trip towards the closest third level centre with places at disposition, newborn admission in the third level centre and team return to the operation place. Newborn admissions in the third level centres considers the following actions: entry, both for local births and for arrivals after transport, stay and discharge: all these arrival-discharge pairs determine disposition of places for further admissions.

For model implementation we used the software package Micro Saint 3.2, a powerful and flexible instrument, which can be used both by simulation's experts and programmers and by less qualified users. Model implementation uses a graphic model

representation where essential elements, shown in Fig. 1, are: ellipses ("tasks") representing activities (time running and/or variable computing), directed arcs ("paths") representing activity sequences, rhombs ("decisions") representing (multiple, probabilistic or deterministic) decisions and striped rectangles ("queues") representing queues.

The whole model graphic representation is reported in Fig. 2. There task 1 gives the start to the whole simulation and activates task 2 (birth activator), task 19 (transport requests to Padova service activator) and task 42 (transport request to Verona service activator). On its own task 19 activates twenty transport request generators (tasks 20 to 39), each one corresponding to a (first or second level) birth centre which is served by Padova service; as can be seen in Fig. 2 and 3, every generated request is processed by tasks 89 (call to Padova service), 40 (activation of Padova team), 41 (trip of Padova team from Padova to the birth centre), 58 (newborn stabilization), 67 (newborn transport) 60 to 66 plus 69 (newborn admission in one of the third level centre) and eventually 92 to 99 (come back to Padova); if all places in the arrival third level centre are busy, the newborn is admitted in another department of the same hospital (task 87). Correspondingly task 42 activates thirteen transport request generators (tasks 43 to 55), each one corresponding to a (first or second level) birth centre which is served by Verona service; as can be seen in Fig. 2 and 3, every generated request is processed by tasks 91 (call to Verona service), 56 (activation of Verona team), 57 (trip of Verona team from Verona to the birth centre), 59 (newborn stabilization), 68 (newborn transport) 60 to 66 plus 69 (newborn admission in one of the third level centre) and eventually 92 to 99 (come back to Verona); if all places in the arrival third level centre are busy, the newborn is admitted in another department of the same hospital (task 88). As can be seen in Fig. 2 and Fig. 4, on its own
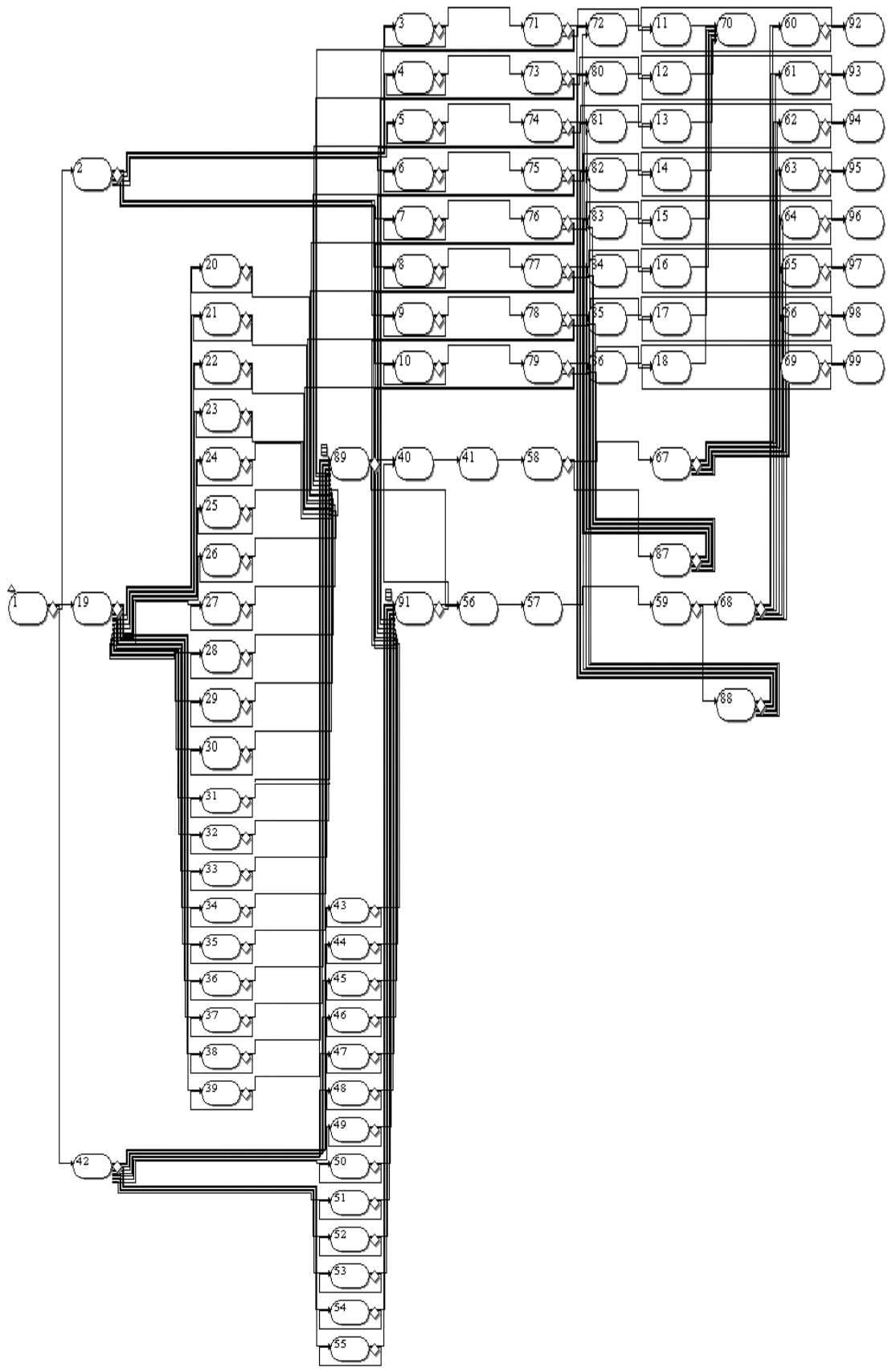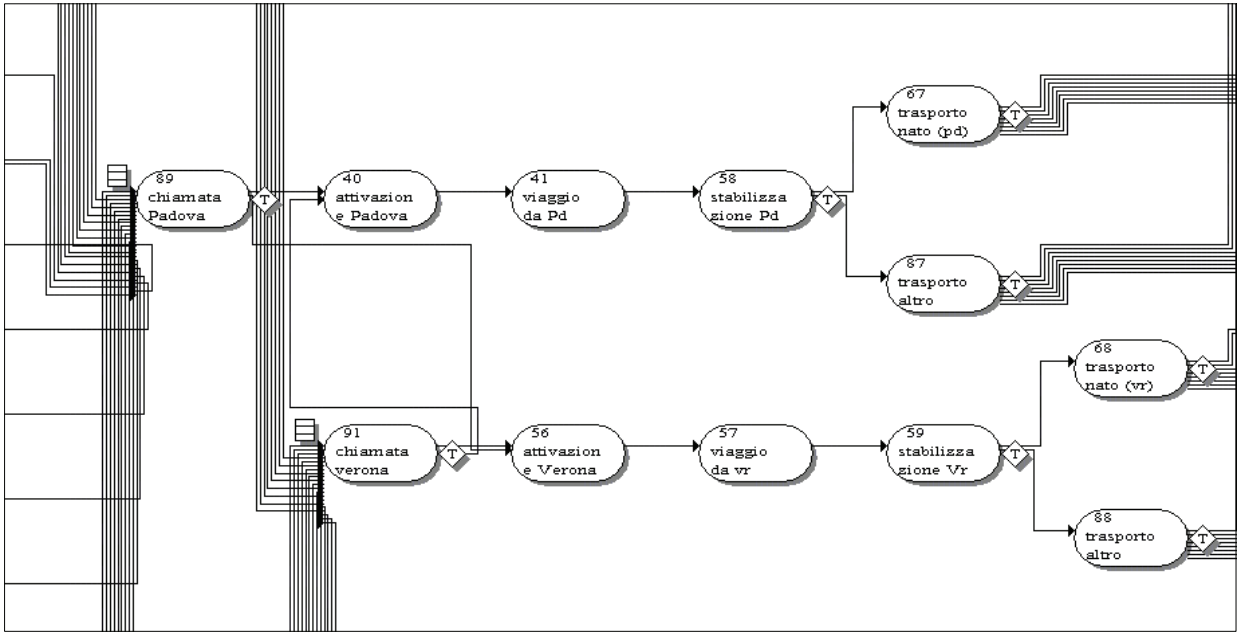
Fig.2 Whole model graphic representation

190

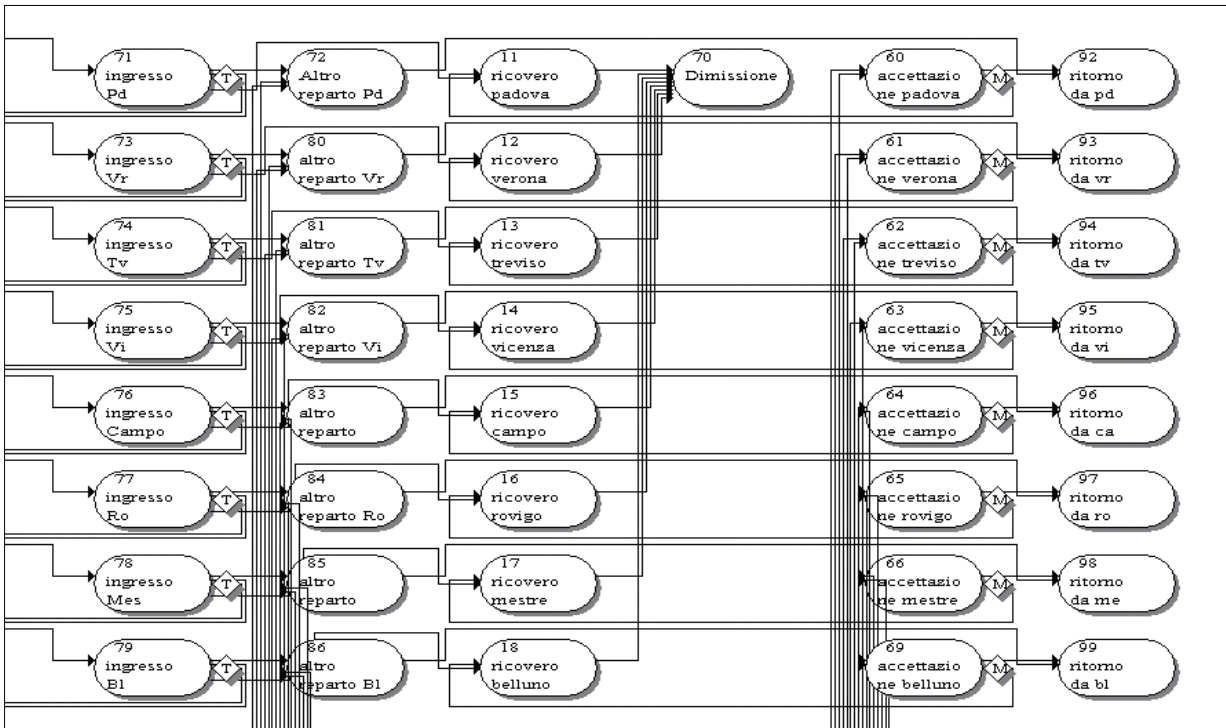Fig.2: Graphic representation of critical newborn transport functions.

Fig. 3: Graphic representation of admission and discharge functions at the eight third level centres

task 2 activates all births which take place in the third level centres (tasks 3 to 10); such births are processed by tasks 71 plus 73 to 79 (entry in the corresponding third level centre) and 11 to 18 (admission in the centre); if all places are busy, they are admitted in another department of the same hospital (tasks 72 plus 80 to 86).

All times associated with tasks have been obtained from data sources considered in Section 4.
.

6. ANALYSIS OF RESULTS

Main simulation results concern: mean transport teams utilization times, actual newborn transport time between call and admission in the third level arrival centre, queue analysis. Transport team utilization times vary between 4 and 5 hours for the Padova service and is about 6 hours for the Verona service, depending on the third level centre where the newborn is transferred; average actual newborn transport time is 4 and 5 hours respectively, where the difference is due to the fact that the city of Verona is comparatively less central in Veneto Region. For what concerns waiting times, a wait due to busy service is rare both for Padova and Verona, but it may rise up to 6 and 8 hours respectively.

An in-depth analysis about the two services results was carried out, and it concluded that they may be considered as satisfactory; what is not satisfactory is the amount of resources for critical newborn admissions. Two different modifications to be effected in the structures regarding critical newborns were suggested, and the respective consequences on the transport service examined by the model. The first suggestion was an increase in the number of neonatal intensive care places in three most required hospitals: the number of new places was 5, 2 in Padova, 1 in Verona and 2 in Vicenza, on a total of 40 current places in the whole region; the suggested place increase were obtained from the number of critical newborn admitted in hospital departments different from neonatal intensive care units; the modification result is a dramatic reduction of "irregular" admissions for all region third level centres; moreover we notice a soft reduction of actual transport time, and the cancellation of waiting times. The second suggestion was the institution of a second ambulance in Padova: for this case some light improvements were obtained in the reduction of waiting times.

7. CONCLUSIONS

A model of critical newborn transport in Veneto Region has been built up and implemented. The model was usefully employed to validate current service and to test the consequences of some possible structural modification; it may be employed to test new situation as for instance a sensible increase of the amount of critical newborns in next years; finally it may be employed to test different situations in different regions

REFERENCES

Cornette, L, 2004. Transporting the sick neonate. *Current Pediatrics* 14, pp. 20-25.

Das, U. G., Leuthner, S. R., 2004. Preparing the neonate for transport. *Pediatric Clinic North Am.* 51 (3), pp. 581-598.

Fenton, A. C., Lesile, A., Skeoch, C. H., 2004. Optimising neonatal transfert. *Arch. Dis. Child Fetal Neonatal* 89, pp. 215-219.

Fowlie, P. W., Booth, P., Skeoch, C. H., 2004. Moving thr preterm infant. *BMJ* 329, pp. 904-906.

Lupton, B. A., Pendray, M. R., 2004. Regionalized neonatal emergency transport. *Seminars in neonatology 9*, pp. 125-133.

Nicholson, P, Burr, J, Powell, J, 2005. Becoming an advanced practitioner in neonatal nursing: a psyco-social study of the relationship between educational preparation and role development. *Journal of Clinical Nursing* 14, pp. 727-738.

Rashid, A., Bhuta, T., Berry, A., 1999. A regionalised transport service, the way ahead? *Arch. Dis. Child.* 80, pp. 488-492.

Wright, J. D., 2000. Before the transport team arrives: neonatal stabilization; *The Journal of Perinatal & neonatal Nursing* 13 (4), pp. 87-107.

AUTHORS BIOGRAPHY

Francesca Bortolato was gradued in Management Engineering at the University of Padova with a thesis developed at the Epidemiology and Community Medicine Unit.

Anna Ferrante obtained a medical degree at the University of Padova. After a period at the Padova Local Health Unit, where she worked on territorial health assistance, she works now at the Epidemiology and Community Medicine Unit.

Giorgio Romanin-Jacur was gradued in Electrical Engineering at the University of Padova where he is now Full Professor of Operations Research.

Laura Salmaso was gradued in Statistics at the University of Padova and obtained a PhD in Epidemiology. She is now working at the Epidemiology and Community Medicine Unit.

.

A NON-NEWTONIAN FLOW MODEL FOR SIMULATING THE FABRIC COATING PROCESS

Zhao X[a], Stylios G K[b]

[a]Research Institute for Flexible Materials, Heriot-Watt University (SBC), UK
[a]x.zhao@hw.ac.uk, [b]g.stylios@hw.ac.uk

ABSTRACT

The peak stress value (PSV) at the interface between a coating polymer solution and a fabric during coating has been calculated using generalised Navier-Stokes and Carreau models for non-Newtonian flow. The model suggested that knife gap is the most important machine parameter affecting the coating process. A 20% decrease in knife gap for example leads to 100% increase in PSV in the coating paste, which subsequently can lead to poor adhesion between the paste and the fabric during coating. Coating speed was found to have less effect on the PSV, a 50% increase in speed leading to only a 30% increase in PSV, implying that this factor is less significant. The validity of the rheological models was confirmed by experimental investigations.

Kew words: Modelling, simulation, non-Newtonian flow, fabrics, coating

1. INTRODUCTION

Manufacturers of advanced technical textiles are facing increasing pressure to improve product performance and quality while simultaneously reducing costs. In order to reduce production costs, mathematical modelling and numerical simulation are more and more extensively used in product development, process optimization and quality control, which can contribute significantly to meeting these challenges (1).

Previous studies (2) on the rheological behaviour of the paste formulations in coating indicated that the coating paste corresponds to a non-Newtonian, shear thinning fluid as the viscosity decreases with increasing shear rate. The Carreau model was fitted with measured data describing viscosity dependence of the paste formulations on shear rate, with coefficients varying at different solvent volume fraction.

Coating quality can be assessed by the penetration of paste into the fabrics, adhesion between the paste and fabrics, and the uniform spreading of the paste on the fabrics (1). Knowing the parameters of paste formulations, a computer simulation of the coating process can be carried out to relate the coating quality with coating material properties and coating process parameters by modelling the polymer paste flow during the coating process.

2. MODEL OF POLYMER SOLUTION DURING COATING

Most theoretical studies on the spreading of polymer solutions onto a solid surface have been focussed on shear-thinning behaviour (3–7), as attempts to account for other non-Newtonian properties concluded that normal stress effects are not important (7, 8). Based on this assumption, the property of coating materials can be characterized in a phenomenological way, yielding equations of motion that have the form of generalized Navier-Stokes equations 1, 2 (9, 10).

$$\rho\frac{\partial u}{\partial t} - \nabla \cdot \eta(\nabla u + (\nabla u)^T) + \rho(u \cdot \nabla)u + \nabla p = F \quad (1)$$

$$\nabla \cdot u = 0 \quad (2)$$

where η is the dynamic viscosity, ρ is the density, u the velocity field, p the pressure, and F a volume force field such as gravity.

To characterise the dynamic viscosity during coating, the Carreau model (Equations 3 and 4) was employed based on the rheometric analyses of coating materials (2).

$$\eta = \eta_\infty + (\eta_0 - \eta_\infty)\left[1 + \left(\lambda \dot{\gamma}\right)^2\right]^{\frac{(n-1)}{2}} \quad (3)$$

Where

$$\dot{\gamma} = \sqrt{\frac{1}{2}\left((2u_x)^2 + (u_y + v_x)^2 + (2v_y)^2\right)} \qquad (4)$$

where η_0 represents the zero shear rate viscosity, η_∞ is the infinite shear rate viscosity, λ is a parameter with units of time, and n represents the shear rate sensitivity $(0 < n < 1)$.

The rheological behaviour of the polymer paste during coating can be defined by equations 1, 2, 3, and 4, together with a set of appropriate boundary conditions. The formulated mathematical model was then solved via a computer based numerical technique.

3. SYSTEM GEOMETRY

For the knife-over-roll coating, a solvent borne, aromatic, one-component polyurethane fabric coating, (Clear Durane® Basecoat 51144, Raffi and Swanson Inc., Wilmington, MA) was used to coat a number of woven fabric substrates. The fabric was placed between the coating knife and the roller (Figure 1), and the coating material was applied uniformly along the coating knife. The gap distance was adjusted by moving the blade up or down, and the fabric pull-through speed was also adjusted.
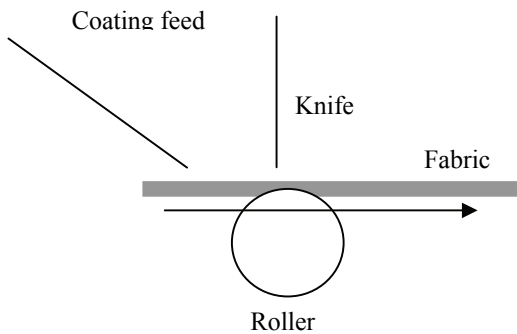
Figure 1: A simplified knife over roller coating mechanism

Coating machines which use knife coating are in general simple to operate and can be used for a wide variety of coating thicknesses. Experimental wok was carried out using a Coatema base coater, which is a pilot scale knife coating machine. A simple sketch of the knife position is shown in Figure 1.

The model equations in this study were used in the finite element method (FEM) to simulate the behaviour of the polymer solution (coating materials) during the coating process. FEM is used because of its ability to efficiently resolve the non-linear phenomena

with a non-uniform grid mesh, the ease with which it handles variable boundary conditions and its firm theoretical foundation. The equations were solved using the MATLAB software package under selected conditions. The geometry and Finite Element mesh of the system can be illustrated in Figure 2.
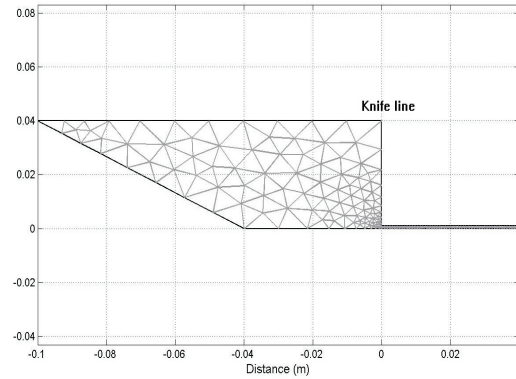
Figure 2: Geometry and Finite Element mesh of the system

4. SIMULATION RESULTS

Based on the model equations 1, 2, 3 and 4, and the system geometry, the results of computer simulation of the properties of the polymer paste within the defined system were presented. These results are mainly presented in graphical forms.

Figure 3 shows the velocity and pressure distribution of the polymer paste before knife (within paste tray) and after knife (on fabric). The colour expresses the pressure distribution. Hotter colours indicate higher pressures. Arrows represent the velocity of polymer paste within the system. The length of the arrow denotes the magnitude of the velocity, and the arrow heads indicate the direction. This figure demonstrates the flow behaviour of the paste during coating. For example, there is a peak value of both velocity and pressure just near the knife.
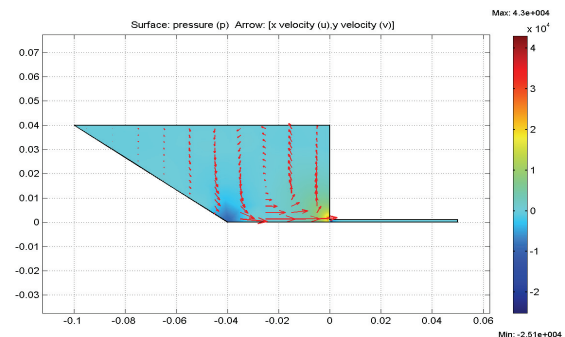
Figure 3: Velocity and pressure distribution of polymer paste within the coating system

To evaluate the quality of coating, the stress values at the interface between a coating polymer solution and a fabric during coating were calculated. They subsequently affect the degree of the penetration of paste into the fabric, the adhesion between the paste and the fabric, and the uniform spreading of the paste on the fabric during coating.
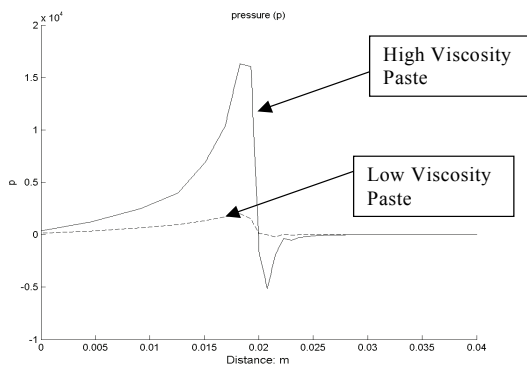
Figure 4: Stress profiles for two polymer pastes

Figure 4 provides stress profiles of two polymer paste samples (dashed curve for sample with solvent concentration of 20/1000, and solid curve for a thicker paste sample of 10/1000) before and after the knife along the interface of the paste on the fabric (x = 0.02 line is the knife position). It can be seen that for both samples there is a steady increase of the stress when the fabric approaches the knife line and a sharp decease when paste and fabric just pass the knife.

This is an expected property due to the chain nature of polymers. Their molecular structure tends to align when subjected to a shear field. After the knife position, the relaxation of the molecular chain causes a spring-back illustrated by the peak. Higher peak stress value (PSV) after the knife position should be avoided during the coating process since it will cause a poorer adhesion between the paste and the fabric. This has been observed from experimental work.

It can also be seen from Figure 4 that a thicker paste leads to a higher PSV. The simulation result indicates that a decrease of viscosity of coating paste by adding small amount of solvent will remarkably decrease the PSV, hence improving adhesion and penetration in the system resulting in a better quality product.
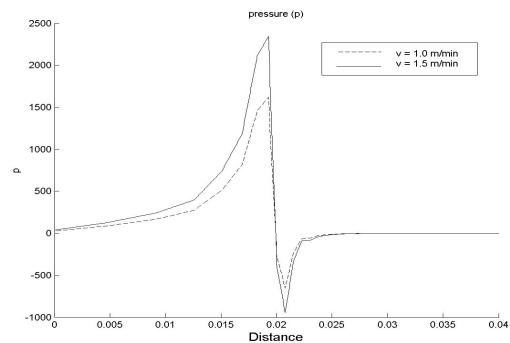
Figure 5: The effect of coating speed on the stress profile of the polymer paste

Figure 5 shows stress profiles of the polymer paste of different coating speeds before and after the knife along the surface of the paste on the fabric (x = 0.02 line is the knife position). It can be seen that a higher coating speed will lead to a higher PSV. A 50% increase (from1.0 to 1.5 m/min) in coating speed caused a 30% increase for PVS. This result implies that keeping other factors constant, the coating quality is affected by changing the coating speed, but not significantly.

Figure 6 illustrates the effect of knife gap on coating. It is remarkable that only a 20% decrease (from 1000 to 800 micron) in knife gap leads to a 100% increase for PSV (from -600 to -1200) indicating that knife gap is the most important processing parameter which need to be carefully controlled during the coating process.
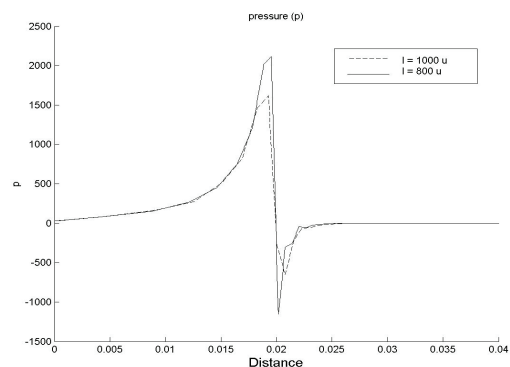
Figure 6: The effect coating knife gap on the stress profile of the polymer paste during coating

Figure 7 presents the viscosity distribution of the polymer paste before the knife (within the paste tray) and after the knife (on the fabric). Hotter colours indicate higher values of viscosity. This result clearly shows the non-Newtonian behaviour of the polymer

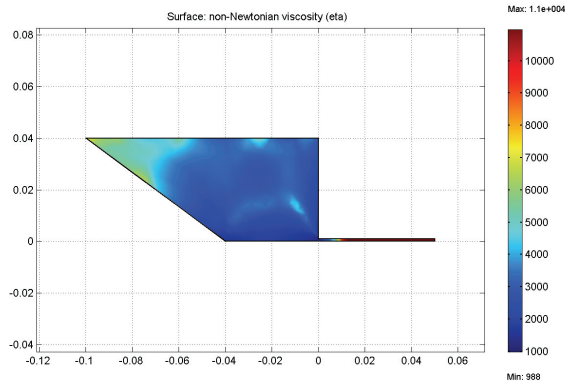paste. Viscosity decreases significantly with the increase of the shear rate.

Figure 7: Viscosity distribution of the polymer paste within the system

5. VALIDATION OF RHEOLOGICAL MODELS

To verify the numerical prediction of coating quality from rheological modelling, seven commercial nylon and polyester fabrics of various woven structures were used as base fabrics for the experimental investigation. Factorial experimental designs were used to produce over 50 samples of coated fabrics by altering 5 variables (knife gap, curing temperature, fabric speed, solution viscosity and fabric structure).

The mechanical and physical properties of the coated and uncoated fabrics where characterised and their tensile, tear, compression and surface properties were measured using the Kawabata Evaluation System for Fabrics (KESF) and the Instron tensile machine. Water vapour permeability and water resistance were measured to evaluate the performance of the samples. Other characteristics were also observed or measured, e.g. through Scanning Electron Microscope (SEM), Differential Scanning Calorimetry (DSC), and chemical tests. Online visual images during coating were also collected and analysed. Using this pool of experimental data, the relationships between coating parameters and textile properties were established and the validity of the rheological models was confirmed.

The strong influence of knife gap, as identified from the rheological models, was confirmed. Increasing this parameter led to improved hydrostatic head resistance, lower water vapour permeability, tear resistance and improve surface uniformity. The viscosity of the solution paste was also found to have an effect on several properties of the coated fabric, including tear resistance and adhesion, as validated with tests and observed by SEM (Figure 8). This is also in agreement with the rheological predictions.
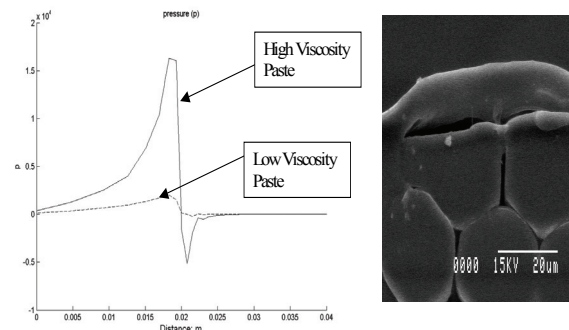

Figure 8: Stress profiles during coating and corresponding SEM images of poor adhesion for low knife gap height (400 nm) and thicker polymer paste (10/1000)

6. CONCLUSIONS

A comprehensive Finite Element model for polymer pastes under given coating conditions opens up possibilities for process optimization and prediction of coating qualities using different polymers and different coating variables such as knife gap, knife angle, fabric speed, etc. Simulation results, which are based on non-Newtonian theories, combined with experimental results provide a clear understanding of the coating process.

The interrelationships between coating materials and coating process conditions have been established for obtaining different coating penetration depths and adhesion, which was verified by user performance measurements. A narrow width continuous base coater has been modified and the findings have been implemented as a test bed for new product development and/or pre-industrial runs for industrial and academic uses.

This model can also be used to improve the design of new and existing coating machinery. For example, the type of knife, the knife blade geometry, the knife introducing angle, the roller diameter, and the dimension of the polymer paste tray, can be simulated and optimized. The selection of suitable materials for the construction of the equipment parts based on the pressure predictions under different coating conditions can also be carried out.

Reference:

[1] Handbook of Technical Textiles, Edited by Horrocks A R and Anand S C, 2000, Published by Woodhead Publishing Limited in association with The Textile Institute

[2] Zhao X, Stylios G K, and Christie R M, 2007, Rheological Behaviour of Polymer Solutions during Fabric Coating, *Journal of Applied Polymer Science,* Volume 107, Issue 4, 2317 - 2321

[3] King, J R, 2001, The spreading of power-law fluids, In IUTAM Symposium on Free Surface Flows (ed. A. C. King & Y. D. Shikhmurzaev), pp. 153–160, Kluwer

[4] Starov, V M, Tyatyushkin, A. N., Velarde, M. G. & Zhdanov, S A, 2003, Spreading of non-Newtonian liquids over solid substrates. *J. Colloid Interface Science*, 257, 284–290

[5] Neogi, P, Ybarra, R M, 2001, The absence of a rheological effect on the spreading of some drops. *J. Chem. Phys.* 115, 7811–7813

[6] Betel S I, Fontelos M A, 2003, Capillarity driven spreading of power-law fluids, *Appl. Math. Letter*, 16, 1315–1320

[7] Rosenblat S, Davis S H, 1985, How do liquid drops spread on solids, In Frontiers in Fluid Mechanics (ed. S. H. Davis & J. L. Lumley), pp. 171–183, Springer

[8] Neogi P and Ybarra R M, 2001, The absence of a rheological effect on the spreading of small drops, *J. Chem. Phys.* 115, 7811-7812

[9] M.J. Crochet, A.R. Davies and K. Walters, 1984, Numerical Simulation of Non-Newtonian
Flow, Elsevier, Amsterdam

[10] Darby, R. 1996, Chemical Engineering Fluid Mechanics, Marcel Dekker, Inc

AUTHORS BIOGRAPHY

Dr **Zhao** has been doing research work at Heriot-Watt University, UK, in the field of protective textiles since 2004. Before joined Heriot-Watt University, he was an Associate Professor (since 1996) at Xinjiang
University, China. His researches involved in the areas including fluid dynamics in textile design and process; mass and heat transfer in textile assemblies and clothing systems; modelling, simulation, prediction, optimization, and control of textile process, and functional materials, etc.

Professor **Stylios** is the director of RIFleX and a senior research professor at Heriot-Watt University, UK. He has always been interested in the interface between textiles and clothing from a number of aspects including providing models for the drapability of cloth, the handle of textiles and the study of complex machinery/material interactions by integrating the mechanics of the material with the variables of the machinery and/or the process.

APPLICATION OF SIMULATION MODELLING TO THE SHIP STEAM BOILER SYSTEM

Enco Tireli[a] , Josko Dvornik[b] , Srđan Dvornik[b]

[a]University of Rijeka
Faculty of Maritime Studies
Studentska ulica 2, 51000 Rijeka
+385 51 338 411

[b]University of Split
Faculty of Maritime Studies
Zrinsko-Frankopanska 38, 21000 Split
+385 21 380 762

[a] tirelli@pfri.hr, [b]josko@pfst.hr, [b]sdvornik@pfst.hr

ABSTRACT

The purpose of this paper is to present the efficiency of the application of the system dynamics simulation modelling (System Dynamics Modelling – Jay Forrester – MIT) in investigating the behaviour dynamics of the ship steam boiler. The ship steam boiler is presented in POWERSIM simulation language through mental – verbal, structural and mathematical – computer models. In addition to the mathematical model for the marine steam boiler, we have as well presented the mathematical model for the feed water heater that is fitted in the system. These models serve as a basis for qualitative and quantitative simulation models for complex systems.

The results presented in the paper have been derived from the scientific research project „SHIPBOARD ENERGY SYSTEMS, ALTERNATIVE FUEL OILS AND REDUCTION OF POLLUTANTS EMISION" supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

Keywords: ship steam boiler, simulation modelling, continuous and discrete simulation

1. SIMULATION MODELLING OF THE MARINE STEAM BOILER

1.1. Mathematical model for the marine steam boiler

The boiler may be considered as a homogenous device, a thermal accumulator, i.e. a homogenous thermal capacity. The equations of thermal balance of such a thermal accumulator (capacity) suggest determining the equation for the level of the water in the boiler, (Nalepin and Demeenko 1975).

System dynamics mathematical model for the marine steam boiler is defined by explicit form of differential equations, (Nalepin and Demeenko 1975):

Equation of the boiler dynamics for the steam pressure

$$\frac{d\varphi_K}{dt} = \frac{1}{T_{a1}} \cdot (\mu_G + a_1 \cdot \mu_V - a_2 \cdot \mu_P - a_3 \cdot \frac{d\mu_P}{dt} - k \cdot \varphi_K) \quad (1)$$

Equation of the boiler dynamics for the water level

$$\frac{d\varphi_Y}{dt} = \frac{1}{T_{a2}} \cdot (\mu_V - b_1 \cdot \varphi_K - b_2 \cdot \frac{d\varphi_K}{dt} - b_3 \cdot \mu_P - b_4 \cdot \frac{d\mu_P}{dt}) \quad (2)$$

Where the following denote:

φ_K - relative state of the steam pressure in the boiler,

φ_Y - relative state of the water level in the boiler,

T_{a1} - time constant of the steam boiler for the steam pressure [s],

T_{a2} - time constant of the steam boiler for the water level [s],

k - coefficient of self-regulation of the steam boiler,

μ_G - relative change of the position of the fuel valve,

μ_V - relative change of the valve of the feed water,

μ_P - relative change of the position of the steam discharge valve,

$\dfrac{d\varphi_K}{dt}$ - speed of the change of the relative increment of the boiler steam pressure,

$\dfrac{d\mu_P}{dt}$ - speed of the relative change of the position of the steam discharge valve,

$a_{1,2,3}$ - coefficients of the steam boiler for the steam pressure,

$b_{1,2,3,4}$ - coefficients of the steam boiler for the water level.

1.2. Mathematical model for the feed water heater in the marine steam boiler system

Equation of the feed water heater dynamics has been derived on the basis of thermal balance and has the following form:

$$T_{1X} \frac{DT_{X2}}{dt} + T_{X2} = T_{2X} \frac{DT_{X1}}{dt} + K_{1X}T_{X1} + K_{2X}T_{CX} \quad (3)$$

Where the following denote:

T_{1X} - time constant for the water medium,

$\dfrac{DT_{X2}}{dt}$ - speed of the relative change of the outlet water temperature,

T_{X2} - relative change of the outlet water temperature,

$\dfrac{DT_{X1}}{dt}$ - speed of the relative change of the inlet water temperature,

K_{1X} - transient coefficient of the water medium,

T_{X1} - relative change of the inlet water temperature,

K_{2X} - transient coefficient of the water medium,

T_{CX} - temperatures of the cooling surface wall of the water medium.

1.3. System dynamics structural model for the marine steam boiler

On the basis of the stated mental- verbal models (Dvornik and Tireli 2007), it is possible to produce structural diagrams of the marine steam boiler, as shown in Figures 1, 2 and 3.
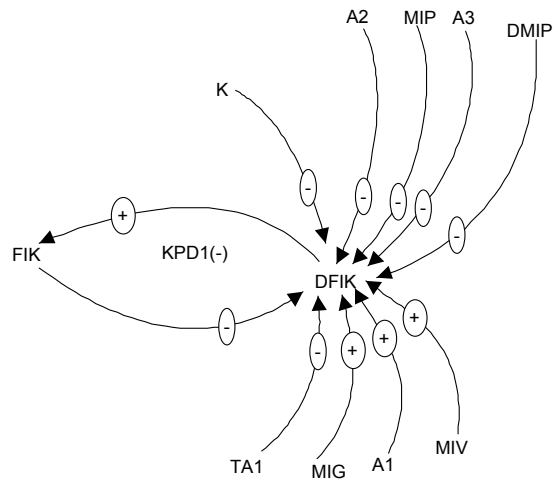

Figure 1: Structural Model for the Marine Steam Boiler – for Steam Pressure

In the observed system there is a feedback loop (KPD1).

KPD1(-):FIK=>(-)DFIK=>(+)DFIK=>(+)FIK; which has self-regulating dynamic character (-), because the sum of negative signs is an odd number.
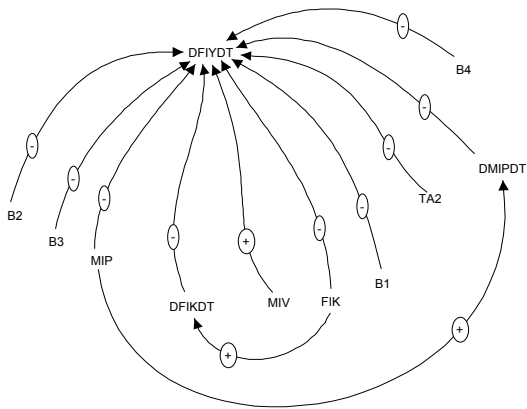
Figure 2: Structural Model for the Marine Steam Boiler – for the Water Level

In the observed system there is not a feedback loop KPD, because the dynamic process of the performance of the steam boiler with natural circulation for the water level does not have the self-regulation property.
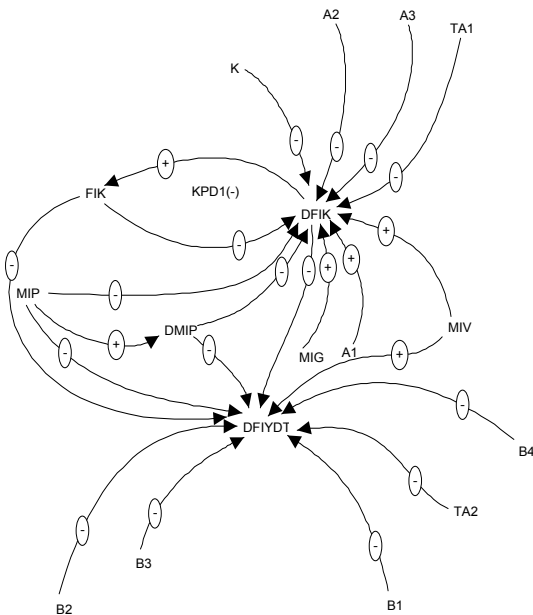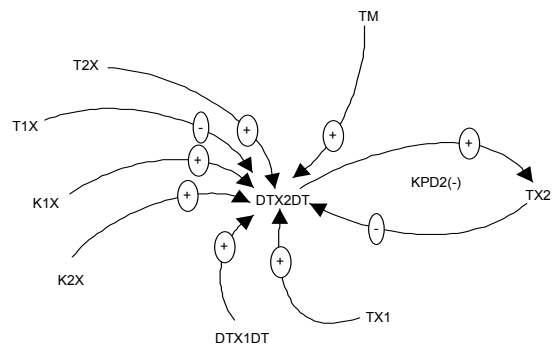

Figure 3: Global Structural Model for the Marine Steam Boiler

1.4. System dynamics structural model for the feed water heater in the marine steam boiler system

On the basis of the dynamics, i.e. stated mental - verbal model (Dvornik and Tireli 2007), it is possible to produce structural model for the feed water heater, as shown in Figure 4.

Figure 4: Structural Model for the Feed Water Heater

In the observed system there is a feedback loop (KPD2).

KPD2(-):TX2=>(-)DTX2DT=>(+)DTX2DT=>

(+)TX2; which has self-regulating dynamic character (-), because the sum of negative signs is an odd number.

1.5. System dynamics flowchart of the marine steam boiler

On the basis of the produced mental - verbal and structural models, the flowchart of the marine steam boiler in POWERSIM simulation language is produced (Byrknes 1993).
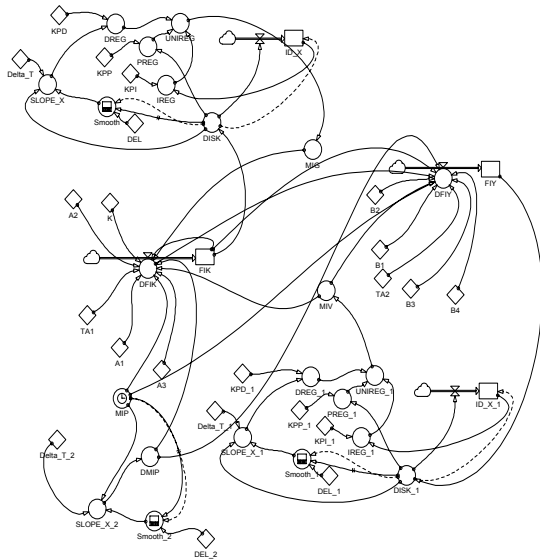
Figure 5: Global Flowchart of the Marine Steam Boiler with the built-in PID Governors

2. INVESTIGATING PERFORMANCE DYNAMICS OF THE MARINE STEAM BOILER IN LOAD CONDITIONS

After system dynamics qualitative and quantitative simulation models have been produced, in one of the simulation packages, most frequently DYNAMO (Richardson and Aleksander 1981) or POWERSIM (Byrknes 1993), all possible operating modes of the system will be simulated in a laboratory.

After an engineer, a designer or a student has conducted a sufficient number of experiments, or scenarios, and an insight has been obtained into the performance dynamics of the system using the method of heuristic optimisation, the optimisation of any parameters in the system may be performed, provided that the model is valid.

In the presented scenario the simulation model for the marine steam boiler for the steam pressure and the water level with two built-in PID governors will be presented.

1. Consumption of the steam is determined by the impulse function of 50-sec duration, which means from 200 – 250 sec, and MIP = 0, FIK = 0, FIY = 0.9999 at the initial TIME = 0.
2. Fuel supply MIG is determined as an outlet of PID-governor, at which inlet there is the discrepancy of the steam pressure (1-FIK) and correspondingly, the water supply MIV is outlet

of the other PID-governor, to which the inlet is discrepancy (1- FIY).
3. Other parameters of the marine steam boiler equal nominal values.
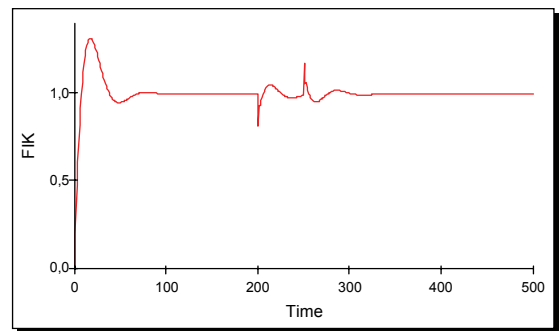
Graphic results of the simulation:

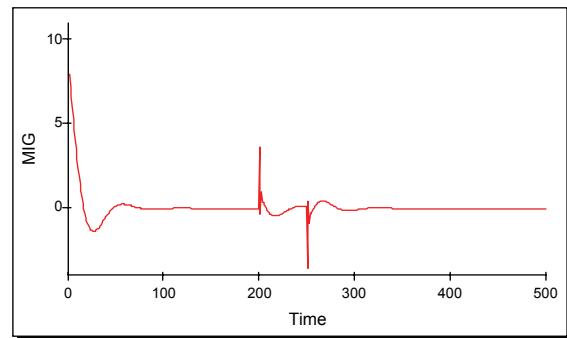Figure 6: Relative State of the Steam Pressure in the Steam Boiler

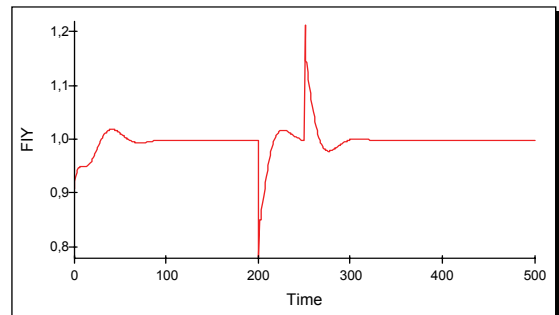Figure 7: Relative Change of the Position of the Fuel Valve

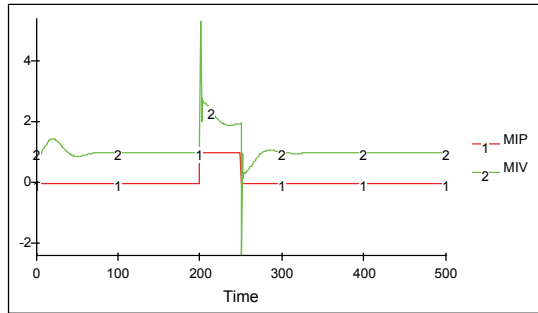Figure 8: Relative State of the Water Level in the Steam Boiler

Figure 9. Relative Change of the Position of the Feed Water Valve and Relative Change of the Position of the Steam Discharge Valve

From the results of the simulation it may be observed that the model shows real performance dynamics and that by applying PID governors and adequate values of coefficients a better levelling and attenuation of the transition occurrence of the variables FIK and FIY will be achieved.

3. CONCLUSION

System dynamics is a scientific method that allows simulation of the most complex systems. The method used in the presented example demonstrates a high quality simulation of complex dynamic systems, and provides an opportunity to all interested students or engineers to apply the same method to modelling, optimising and simulating any scenario of the existing elements.

Furthermore, the users of this method of simulating continuous models in digital computers have an opportunity to acquire new information on dynamic systems performance. The method is also important because it does not only refer to computer modelling, but also clearly determines mental, structural and mathematical modelling of the elements of the system.

This brief presentation gives to an expert all the necessary data and the opportunity to collect information about the system using a fast and scientific method of investigation of a complex system.

Which means: Do not simulate the performance dynamics of complex systems using the method of the "black box", because the education and designing practice of complex systems has confirmed that it is much better to simulate using the research approach of the "white box," i.e. System dynamics methodology.

REFERENCES

Byrknes, A. H.,1993. *Run-Time User's Guide and Reference Manual*. Powersim 2.5, Powersim Corporation, Powersim AS, 12007 Sunrise Valley Drive, Reston: Virginia USA.

Dvornik, J., Dvornik, S., 2007. Simulation modelling and heuristic optimization of the ship steam boiler. *The 26th IASTED International Conference on Modelling, Identification and Control, MIC 2007,* pp. 105-108. February 12-14, Innsbruck (Austria).

Dvornik, J., Dvornik, S., Tireli, E., 2007. System dynamics simulation model of the ship steam boiler.*11th World Multiconference on Systemics, Cybernetics and Informatics WMSCI,* pp. 25-29. July 8-11, Orlando (USA).

Dvornik, S., Dvornik, J., 2007. Contribution to the development of simulation model of the ship steam boiler. *The 2007 European Simulation and Modelling Conference, ESM 2007,* pp. 65-67. October 22-24, St. Julians (Malta).

Forrester, Jay W., 1973/1971. *Principles of Systems*. Cambridge: Massachusetts USA.

Hind, A., 1968. *Automation in merchant ship*. London.

Isakov, L.I. and Kutljin, L.I., 1984. *Kompleksnaja avtomatizacija sudovljlh dizeljnih i gazoturbinmljih ustanovok*. Leningrad: Sudostreonnie.

Munitić, A., 1989. *Computer Simulation with Help of System Dynamics*.Croatia : BIS Split.

Nalepin, R.A. and Demeenko, O.P., 1975. *Avtomatizacija sudovljih energetskih ustanovok*. Leningrad: Sudostroennie.

Richardson, G. P. and Aleksander, L., 1981. *Introduction to System Dynamics Modelling with Dynamo*. MIT Press, Cambridge : Massachusetts USA.

Suprun, G. F., 1972. *Sintezsistem elektroenergetiki sudov*. Leningrad : Sudostroenie.

Šneller, S., 1996. *Pogon broad I – generatori pare*. Sveučilište u Zagrebu , Fakultet strojarstva i brodogradnje:Zagreb.

Šretner, J., 1975. *Brodski parni kotlovi*. Sveučilište u Zagrebu : Zagreb.

AUTHORS' BIOGRAPHIES
Enco Tireli
He was born in Rijeka in 1947. He completed undergraduate studies at Faculty of Engineering in Rijeka in 1970 and was awarded the BSc degree – marine engineer. He was awarded the master's degree at Faculty of Engineering in Ljubljana, Slovenia, in 1975 – MSc in technical sciences, the field of energetics. In 1978 he won the doctor's degree at Faculty of Engineering in Ljubljana, Slovenia – PhD in technical sciences. In 1970 he worked as a planner at Rade Končar factory in Rijeka. From 1970 to 1973 he was the Development and production manager at the factory of thermal installations TTU in Labin. From 1974 to 1991 he was the Plant manager of the thermo-electric power plant TE Plomin 1, as well as the Construction manager of the TE Plomin 2. Since 1992 he has been Vice-dean for business relations at Faculty of Maritime Studies in Rijeka. Since 2004 he has been Full professor in charge of courses in Failure diagnosis, Marine thermal turbines, Exploitation of marine steam boilers, Exploitation of marine thermal turbines, Fuels, lubricants and water, Optimisation of the ship's propulsion, and Thermal science.

Joško Dvornik
He was born in Split in 1978. He completed undergraduate studies at Faculty of Maritime Studies in Split in 2001 and was awarded the BSc degree – graduated engineer in Maritime transport, marine engineer. In 2001 he employed at the Maritime Faculty in Split as Junior Researcher at the scientific project No 01717007, 2004 graduated from post graduate studies at the Faculty of Electrical Engineering, Mechanical Engineering and Shipbuilding as a Master of Science in Technical Sciences, filed of Mechanical Engineering. In 2001 he won the doctor's degree at Faculty of Maritime Studies in Rijeka – doctor of Tehnical Science, field of Traffic and Transport Technology, branch of Marine and River transport. He published over 50 scientific papers about the field of System Dynamics Computer Simulation Modelling, of which over 20 paper relate to the ship steam and gas turbines, ship engines and complex ship propulsion systems.

Srđan Dvornik
He was born in Split in 1966. He completed undergraduate studies at Faculty of Maritime Studies in Split in 1990 and was awarded the BSc degree – graduated engineer in Maritime transport, marine engineer. From 1990 to 2005 he sailed as an engineer officer on domestic and foreign merchant ships. He passed the exam for Chief engineer officer on a ship powered by the main propulsion machinery of 3,000 kW or more in 1995 in Split. Since 2005 he has been an assistant at the Department of marine engineering of Faculty of Maritime Studies in Split in the area of Technical sciences, the field of Traffic and transport technology. He is the author of more than 10 scientific papers in the area of System dynamics computer simulation modelling, dealing with various systems and processes.

CONTROLLER DESIGN USING COMBINATION OF SYMBOLIC AND NUMERIC CALCULATIONS IN MAPLE

Pavol Bisták[(a)], Peter Ťapák [(b)]

[(a)] Slovak University of Technology in Bratislava, Fac. of Electr. Eng. & Info. Technology
[(b)] Slovak University of Technology in Bratislava, Fac. of Electr. Eng. & Info. Technology

[(a)] Pavol.Bistak@stuba.sk, [(b)] Peter.Tapak@stuba.sk

ABSTRACT

The paper shows how it is possible to combine symbolic computations with numerical methods in order to achieve the controller action under specified conditions. The corresponding calculations are based on the nonlinear state space model resulted from dynamics decomposition. The controlled system is represented by the triple integrator with input saturation. The design combines the well known time optimal control with the linear pole assignment control, i.e. the control consist of n phases similar to the time optimal control, however the transients between these phases are "smooth" and the dynamics of the transients is given by the closed loop poles. For modeling and simulations the computer algebra system Maple has been used.

Keywords: nonlinear model, symbolic calculations, pole assignment control, time optimal control

1. INTRODUCTION

Recent decade in the control design is characteristic by a revival of theory of constrained systems. The minimum time control, which was dominating the control design from late 40-ties to the beginning of 70-ties in the 20th century (see e.g. Pavlov 1966; Athans and Falb 1966) is, however, replaced by several new approaches as the predictive control (see e.g. Bemporad et al. 2002; El-Farra et al. 2003), different anti-windup solutions, positive invariance sets, etc. Motivation comes from different fields – from the traffic control, robot control, control of unstable systems, etc. A common feature of the new approaches is that they are rather complex - even in the case of simple control problems. So, they are not easy to understand and to apply. Traditionally, the engineering community preferred simpler solutions, as e.g. that one proposed by Kiendl and Schneider (1972), which was later used in robot control (Kunze 1984; Patzelt 1981).

Parallel to this, family of new not yet widely known solutions (Huba 1994; Huba and Bisták 1995; Huba and Bisták 1999; Huba 2006) was developed also thanks to new possibilities given by the modelling and simulation software tools. They are relatively simple for understanding, easy to implement and so appropriate also for extremely fast application and easy to tune by a

procedure that can be considered as a generalization of the well-known method by Ziegler and Nichols. For the controller design purposes we usually use the computer algebra system Maple (mainly for symbolic calculations) beside the well-known Matlab (suitable for numeric methods and simulations). The whole presented design is carried out in the Maple environment.

In this paper, the family of the already known approaches is extended by the control design, which is based on the decomposition of the closed loop dynamics into particular modes defined by chosen real closed loop poles. This decomposition results in nonlinear state space model that is built in the Maple environment. Then the control law is calculated after localizing the current state in the nonlinear model and is related to the distance from the corresponding target object.

2. LINEAR POLE ASSIGNMENT CONTROL BASED ON MODES DECOMPOSITION

Let us consider the 3rd order integrator

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{b}u \quad \begin{matrix} \dot{x} = y \\ \dot{y} = z \\ \dot{z} = u \end{matrix} \quad \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u \quad (1)$$

whereby $\mathbf{x}$ is the representative point, $\mathbf{b}$ is the input vector and $\mathbf{A}$ is the system matrix.

The linear pole assignment controller fulfills three rules:

1. It decreases the distance of the representative point from the plane.
2. It decreases the oriented distance of the representative point lying in that plane from the line.
3. Along that line the controller decreases the oriented distance from the origin.

Let us consider closed loop system with the following poles $\alpha_3 < \alpha_2 < \alpha_1 < 0$. Since the eigenvectors

$$\mathbf{v}_i = [\alpha_i \mathbf{I} - \mathbf{A}]^{-1} \mathbf{b} = \begin{bmatrix} \dfrac{1}{\alpha_i^3} & \dfrac{1}{\alpha_i^2} & \dfrac{1}{\alpha_i} \end{bmatrix}^T \tag{2}$$

are not collinear, they form a basis, which can be used for expressing any state as a sum of three modes

$$\mathbf{x} = q_1 \mathbf{v}_1 + q_2 \mathbf{v}_2 + q_3 \mathbf{v}_3 , q_1, q_2, q_3 \in R \tag{3}$$

Then one can write

$$\begin{aligned}
\dot{\mathbf{x}} &= \dot{\mathbf{x}}_1 + \dot{\mathbf{x}}_2 + \dot{\mathbf{x}}_3 = \\
&= \mathbf{A}\mathbf{x} + \mathbf{b}u = \\
&= \mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) + \mathbf{b}(u_1 + u_2 + u_3)
\end{aligned} \tag{4}$$

After denoting $\mathbf{x}_i = q_i \mathbf{v}_i$ each subsystem can be expressed as

$$\begin{aligned}
\dot{\mathbf{x}}_1 &= \mathbf{A}\mathbf{x}_1 + \mathbf{b}u_1 = \alpha_1 \mathbf{x}_1 \\
\dot{\mathbf{x}}_2 &= \mathbf{A}\mathbf{x}_2 + \mathbf{b}u_2 = \alpha_2 \mathbf{x}_2 \\
\dot{\mathbf{x}}_3 &= \mathbf{A}\mathbf{x}_3 + \mathbf{b}u_3 = \alpha_3 \mathbf{x}_3
\end{aligned} \tag{5}$$

So the 3rd order dynamics can be decomposed into three 1st order ones. The appropriate interpretation and the appropriate choice of the oriented distance measurement from the representative point $\mathbf{x}$ to the plane (or the line) lead to three control phases well known from the time optimal control. All of the coordinates ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$) are equivalent in this linear case. Let us assign following:

1. The 1st equation describes the transient into the origin along the line given by $\mathbf{v}_1$. The dynamics is given by α_1.
2. The 2nd equation describes the transient in the plane given by $\mathbf{v}_1, \mathbf{v}_2$ to the line given by $\mathbf{v}_1$ measuring oriented distance in the direction of $\mathbf{v}_2$. The dynamics of the transient is given by the second pole α_2.
3. Similarly to 1. and 2. the 3rd equation describes the transient to the plane using the measurement direction of $\mathbf{v}_3$. The dynamics of this transient is given by α_3.

The total control signal is then the following sum

$$u = \sum_{i=1}^{3} u_i = \sum_{i=1}^{3} q_i \tag{6}$$

The control algorithm described above guarantees that all three control phases are running in parallel and this is equal to the result of the well-known Ackerman's formula. The Fig. 1 shows the new base described above. This new base creates the model for linear pole

assignment control that is simply represented by three eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$. The control low can be computed by calculation of the distance between the representative point $\mathbf{x}$ and lines given by these three vectors. In the nonlinear case when the input constraints are present the model for pole assignment control will be more complicated because it consists of line segments (corresponding to linear control), curves a surfaces as it is shown below.
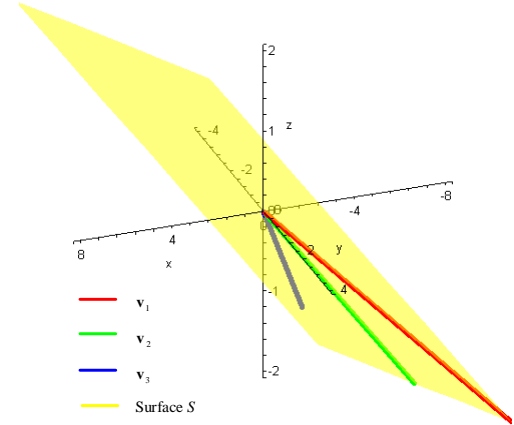

Figure 1: New basis of phase space

3. NON-LINEAR (CONSTRAINED) POLE ASSIGNEMNT CONTROL BASED ON MODES DECOMPOSITION

Let us consider a constrained control signal

$$u \in \langle U_1 \quad U_2 \rangle \tag{7}$$

The transient of the representative point with an initial state on the line given by the eigenvector $\mathbf{v}_1$ outside the proportional band does not follow this line. The particular subsystems do not change themselves only in their own coordinates $\mathbf{x}_i$, but also in the other subsystems coordinates. There is necessary to assign the coordinate to each of the control phases. The following assignment has been used:

1. The transient of the representative point in the phase space to the reference surface (RS) is given by $\alpha_3, \mathbf{v}_3$.
2. The transient along the RS to the reference curve (RC) is given by $\alpha_2, \mathbf{v}_2$.
3. The transient along the RC into the origin is given by $\alpha_1, \mathbf{v}_1$.

The control phases in this control algorithm combine the time optimal control with the linear pole assignment control. Each control phase is given by particular constraint of the control signal, but the transition between the control phases is described by the chosen corresponding closed loop pole. So each control phase is given by two parameters:

α_i - describes the dynamics of the transient of the representative point $\mathbf{x}_i$ inside the linear subsystem (proportional band)

t_i - the time, which is needed to reach the "linear" subspace under influence of limit control values..

One part of the final control phase (the linear one), corresponding to the transient into the origin, when the other coordinates $\mathbf{x}_3$ and $\mathbf{x}_2$ are zeroes, is given by the closed loop pole α_1 only along the line given by the eigenvector $\mathbf{v}_1$, where the coordinate

$$\mathbf{x}_1 \in \left\langle U_j \mathbf{v}_1 \quad U_{3-j} \mathbf{v}_1 \right\rangle \tag{8}$$

Let us assign the constraint of the control signal in the final control phase as U_j where

$$j = 1,2 \tag{9}$$

The linear control interval is restricted by

$$q_1 \in \left\langle U_j \quad U_{3-j} \right\rangle \tag{10}$$

The second part of the final control phase (non-linear) is described by the time t_1, that represents the time of the transient of the representative point using $u = U_j$ to the border points of the linear subsystem $^j X_1 = U_j \mathbf{v}_1$. The 2nd part is created by points $\mathbf{x}_1$ as the result of backward integration of (1) on the interval $t \in \left\langle 0 \quad t_1 \right\rangle$ using $u = U_j$ and starting from the point $^j X_1$. One gets

$$\mathbf{x}_1(t_1) = \begin{bmatrix} 1 & -t_1 & \dfrac{t_1^2}{2} \\ 0 & 1 & -t_1 \\ 0 & 0 & 1 \end{bmatrix} q_1 \mathbf{v}_1 + \begin{bmatrix} \dfrac{-t_1^3}{6} \\ \dfrac{t_1^2}{2} \\ -t_1 \end{bmatrix} U_j \tag{11}$$

So the $\mathbf{x}_1$ represents all the points of the 1st subsystem, and these points are given by the parameter q_1 for the points $t_1 = 0$, $\mathbf{x}_1 \in \left\langle U_j \quad U_{3-j} \right\rangle \mathbf{v}_1$ and also by $t_1 \neq 0$, $q_1 = U_j$ for the points outside the linear part (8). So let us introduce the following generalized denotation $\mathbf{x}_1(q_1, t_1)$, i.e. the points lying in the proportional band (the control signal is not saturated) are represented as $\mathbf{x}_1(q_1, 0)$ and the points lying outside (8) with saturated control signal are represented as $\mathbf{x}_1(U_j, t_1)$. The control signal for points in the proportional band $\mathbf{x}_1 = \mathbf{x}_1(q_1, 0) \in \left\langle U_1 \mathbf{v}_1 \quad 0 \right\rangle$ is

$u_1 = z_1 \alpha_1$. The control signal for the points outside the proportional band $\mathbf{x}_1(U_j, t_1)$ is $u = U_j$. Such a representation of these points describes the transient along the RC with dynamics given by the closed loop pole α_1 with respecting the control constraints. The RC is invariant set of the system $\dot{\mathbf{x}}_1 = \mathbf{A}\mathbf{x}_1 + \mathbf{b}u_1$ with constrained control signal $u_1 \in \left\langle U_j \quad U_{3-j} \right\rangle$, i.e. after approaching the RC, the system remains on the RC.

In the second control phase is $\mathbf{x}_3 = 0$. There is the transient in the surface given by $\mathbf{x}_1$, $\mathbf{x}_2$ towards the RC (characterized by $\mathbf{x}_1$) described in this phase. The goal is to approach the RC, where $\mathbf{x}_2 = 0$. The dynamics of this transient is given by α_2 only in the proportional band of the second subsystem

$\mathbf{x}_2 \in \left\langle 0 \quad (U_{3-j} - q_1) \right\rangle \mathbf{v}_2$, i.e.

$$q_2 \in \left\langle 0 \quad (U_{3-j} - U_j) \right\rangle \tag{12}$$

that gives the 1st part of the surface. The second part of the surface is given as the result of backward integration of (1) on the interval $t \in \left\langle 0 \quad t_2 \right\rangle$ starting from the points where the q_2 is saturated using $u = U_{3-j}$. Using generalized denotation of the RC one gets

$$\mathbf{x} = \mathbf{x}_2((U_{3-j} - q_1), t_2) + \mathbf{x}_1(q_1, t_1) =$$

$$= \begin{bmatrix} 1 & -t_2 & \dfrac{t_2^2}{2} \\ 0 & 1 & -t_2 \\ 0 & 0 & 1 \end{bmatrix} (\mathbf{x}_1(q_1, t_1) + q_2 \mathbf{v}_2) + \begin{bmatrix} \dfrac{-t_2^3}{6} \\ \dfrac{t_2^2}{2} \\ -t_2 \end{bmatrix} U_{3-j} \tag{13}$$

In the starting control phase the 3rd subsystem $\mathbf{x}_3$ is in the proportional band for

$\mathbf{x}_3 \in \left\langle (U_j - q_1 - q_2)\mathbf{v}_3 \quad (U_{3-j} - q_1 - q_2)\mathbf{v}_3 \right\rangle$, i.e.

$$q_3 \in \left\langle U_j - q_1 - q_2 \quad U_{3-j} - q_1 - q_2 \right\rangle \tag{14}$$

Similarly to the previous subsystems the result of backward integration on the interval $t \in \left\langle 0 \quad t_3 \right\rangle$ starting from the points where q_3 is saturated using $u = U_j$, gives the 3rd subsystem. Using generalized denotation $\mathbf{x}_2(q_2, t_2)$, the general point of the surface can be represented as $\mathbf{x}_1(q_1, t_1) + \mathbf{x}_2(q_2, t_2)$ and any point of the state space can be expressed using modes decomposition as

$$\mathbf{x} = \mathbf{x}_3(q_3, t_3) + \mathbf{x}_2(q_2, t_2) + \mathbf{x}_1(q_1, t_1) =$$

$$= \begin{bmatrix} 1 & -t_3 & \dfrac{t_3^{\,2}}{2} \\ 0 & 1 & -t_3 \\ 0 & 0 & 1 \end{bmatrix} (\mathbf{x}_1(q_1, t_1) + \mathbf{x}_2(q_2, t_2) + \qquad (15)$$

$$+ q_3 \mathbf{v}_3) + \begin{bmatrix} \dfrac{-t_3^{\,3}}{6} \\ \dfrac{t_3^{\,3}}{2} \\ -t_3 \end{bmatrix} U_j$$

Note that the whole state space can be described by $\mathbf{x}_1 = \mathbf{x}_1(q_1, t_1)$, $\mathbf{x}_2 = \mathbf{x}_2(q_2, t_2)$, $\mathbf{x}_3 = \mathbf{x}_3(q_3, t_3)$, using parameters $q_i, i = 1,2,3$ that describe the length of the vectors in the proportional bands of the subsystems and $t_i, i = 1,2,3$ describing these vectors outside theirs proportional bands (the control signal of the subsystem is saturated). The coordinates $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ create the model for pole assignment control of saturated system. The model is sophisticated because it consists of several segments of surfaces that are depicted below. The sequential choice of the coordinates of the subsystems guarantees, that the control signal is saturated only if the third subsystem is saturated also.

The control algorithm is the same as the linear one, but it depends on achieving the parameters $q_i, t_i, i = 1,2,3$ which are more difficult to obtain. The only way how to get them is to use symbolic calculations to derive the equation of a corresponding segment of surface and then to numerically evaluate the distance from the segment. The control is

$$u = \sum_{i=1}^{3} u_i \qquad (16)$$

where u_i are control signals of particular subsystems. The RS is given by the points where $u_3 = 0$, but we divide it according to the parameters t_1, t_2, j. Let us denote following:

$RS^{\,j}{}_0$ - the part of RS, where $t_1 = 0, t_2 = 0$

$RS^{\,j}{}_1$ - the part of RS, where $t_1 > 0, t_2 = 0$

$RS^{\,j}{}_2$ - the part of RS, where $t_1 = 0, t_2 > 0$

$RS^{\,j}{}_{12}$ - the part of RS, where $t_1 > 0, t_2 > 0$

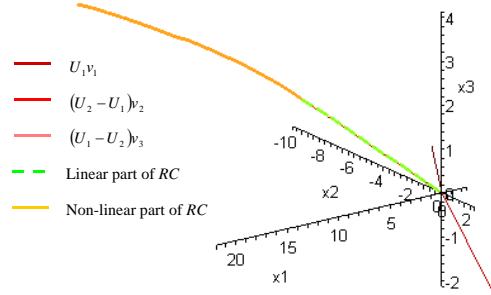Following figures shows the RC and the RS for $j = 1$, $\alpha_1 = -0.5$, $\alpha_2 = -1$, $\alpha_3 = -2$, $u \in \langle -1 \quad 1 \rangle$.

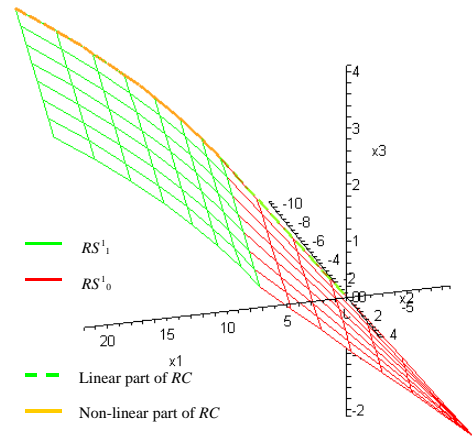Figure 2: Eigenvectors $\mathbf{v}_i, i = 1,2,3$ and RC^1
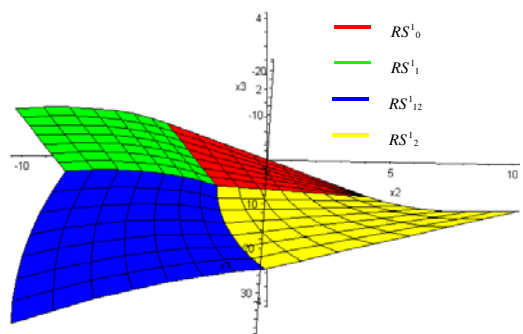

Figure 3: $RS^1{}_0$, $RS^1{}_1$

Figure 4: RS^1

The proportional band (PB) of the system (1) is given by points where $q_3 \in \langle U_1 - q_1 - q_2 \quad U_2 - q_1 - q_2 \rangle$. Let us denote $q_{3\min} = U_1 - q_1 - q_2, q_{3\max} = U_2 - q_1 - q_2$. Fig. 5 and Fig. 6 show the parts of PB corresponding to particular segments of RS.
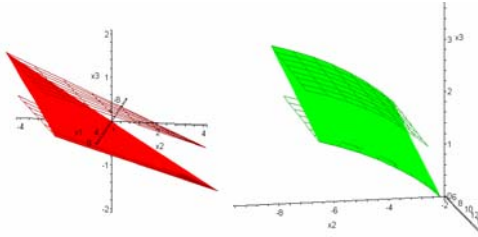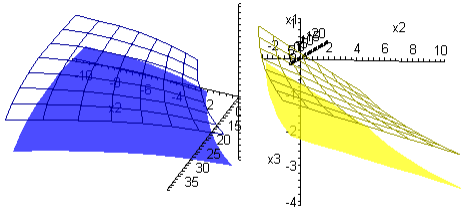
Figure 5: *PB* corresponding to the segments $RS^{j}{}_{0}$, $RS^{j}{}_{1}$

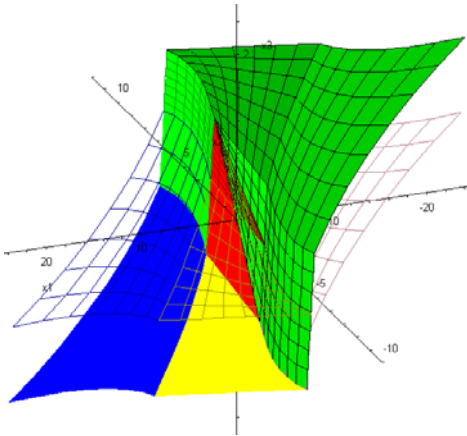Figure 6: *PB* corresponding to the segments $RS^{j}{}_{12}$, $RS^{j}{}_{2}$

Figure 7: *RS* and *PB* in the phase space

4. NON-LINEAR CONTROL ALGORITHM

The control algorithm is based on achieving parameters $q_i, t_i, i = 1,2,3$. However, parameter t_3 is not needed, because there is enough to know whether $\mathbf{x}_3$ is in proportional band or saturation, so let us assign $t_3 = 0$. To find these parameters there is necessary to solve (15), however the results obtained by symbolic solutions can be used now. The formula for evaluation of $\mathbf{x}_3$ differs for each segment and it depends on $\mathbf{x}_1, \mathbf{x}_2$, so the control algorithm can be divided into following steps:

1. START
2. $RS^{j}{}_{0}$
   a. q_1, q_2, q_3 are unknown
   b. $t_1 = 0, t_2 = 0$
   c. solve (15)
   d. (10), (12) are not fulfilled GOTO 3

e. $sat(q_3), u = \sum_{i=1}^{3} q_i$ GOTO 6

3. $RS^{j}{}_{1}$
   a. q_2, q_3, t_1 are unknown
   b. $q_1 = U_j, t_2 = 0$
   c. solve (15)
   d. $t_1 > 0$, (12) are not fulfilled GOTO 4

   e. $sat(q_3), u = \sum_{i=1}^{3} q_i$ GOTO 6

4. $RS^{j}{}_{2}$
   a. q_1, q_3, t_2 are unknown
   b. $q_2 = U_{3-j} - q_1, t_1 = 0$
   c. solve (15)
   d. $t_2 > 0$, (10) are not fulfilled GOTO 5

   e. $sat(q_3), u = \sum_{i=1}^{3} q_i$ GOTO 6

5. $RS^{j}{}_{12}$
   a. q_3, t_1, t_2 are unknown
   b. $q_1 = U_j, q_2 = U_{3-j} - U_j$
   c. solve (15)
   d. $t_1 > 0, t_2 > 0$ are not fulfilled GOTO ERROR

   e. $sat(q_3), u = \sum_{i=1}^{3} q_i$ GOTO 6

6. If the distance of the representative point from the desired state is greater than ε GOTO START
7. END

5. VERIFYING THE CONTROL ALGORITHM BY SIMULATION

The first simulation shows the three control phases presented in this paper.

Parameters of the simulation with initial state outside the *PB* are:

Chosen closed loop poles $\alpha_1 = -1.5, \alpha_2 = -3, \alpha_3 = -6$

Initial state $\mathbf{x} = [14.145, -6.097, 0.833]^{\mathrm{T}}$
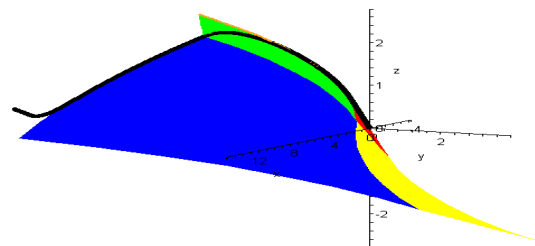

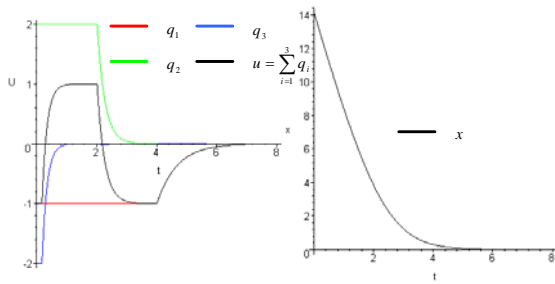Figure 8: Trajectory in the phase space

Figure 9: Control signal and output

Fig. 9 shows that the control signal approaches the constraint three times.

In the next simulation there is shown the trajectory starting at the x-axis. The symbolic solution of the algorithm for RS^j_2 is used in this case. Before, we had the symbolic solution only for the particular closed loop poles and for the particular constraints. The chosen closed loop poles in the simulation are $\alpha_1 = -1.5, \alpha_2 = -3, \alpha_3 = -6$ and the initial state is $\mathbf{x} = [30, 0, 0]^T$.
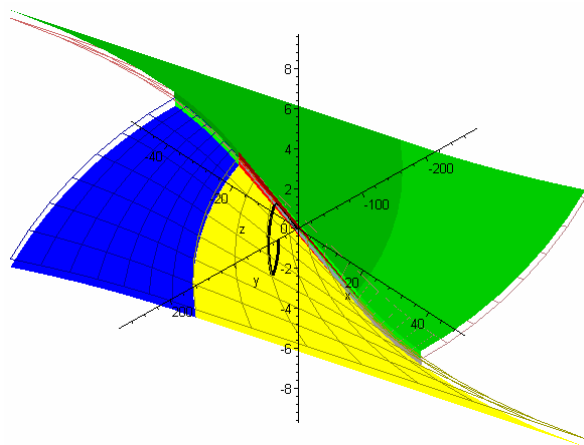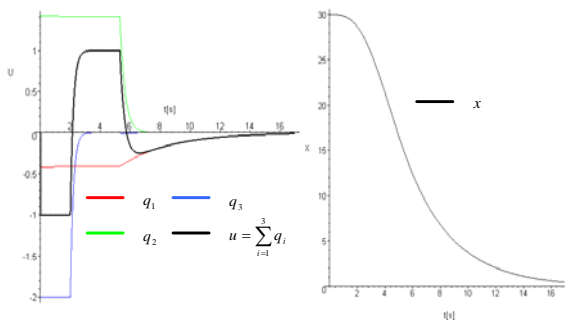

Figure 10: Trajectory in the phase space

Figure 11: Control signal and output

6. CONCLUSION

The simulations have shown, that designed control consists of three phases well known from time optimal control, however the transients between these phases are "smooth" and they have dynamics given by the closed loop poles that gives the advantage to this algorithm in the field of nonlinear control. It can be used to control

systems with parasitic transport delays etc. The interesting approach to the design of controller has shown that under the specified conditions (in this case given by input saturation limits) the symbolic computation are able to produce sophisticated solutions that can be finally evaluated by numerical calculations. Due to the symbolic solutions made in the MAPLE computer algebra system, the controller design process can be easily applied to different 3rd order systems.

REFERENCES

Athans, M., Falb, P.L., 1966. *Optimal Control, An Introduction to the Theory and its Applications*. MacGraw-Hill, New York.

Bemporad, A., Morari, M., Dua, V. and Pistikopoulos, E.N., 2002. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38, 3-20.

Bisták, P., Huba, M., Skachová, Z., 1996. Controller design for nonlinear systems with constrained output. *Proceedings of 5th Electronical and Computer Science Conference ERK´96*, Vol. A, Portorož, Slovenia

El-Farra, N.H., Mhaskar, P. and Christofides, P.D. 2003. Uniting bounded control and MPC for stabilisation of constrained linear systems. *Automatica*, 40, 101-110.

Huba, M., 1994. Saturating Pole Assignement Control-Construction and geometrical interpretation on the phase plane. *Proceedings of 1st IFAC-Workshop „NTDCS"*, Smolenice, Slovakia.

Huba, M., Bisták, P., 1995. Saturating control algorithms for some typical motion control task. *Proceedings of IFAC Conference on Motion control*, Munich, Germany.

Huba, M. and Bisták, P., 1999. Dynamic Classes in the PID Control. *Proceedings of the 1999 American Control Conference*. AACC, San Diego.

Huba, M., 2006. Constrained pole assignment control, In: *Current Trends in Nonlinear Systems and Control*, Boston: Birkhäuser, 163-183.

Kiendl, H. and Schneider, G., 1972. Synthese nichtlinearer Regler für die Regelstrecke const/s2 aufgrund ineinandergeschachtelter abgeschlossener Gebiete beschränkter Stellgröße. *at 20*, 289-296.

Kunze, H.B., 1984. Regelungsalgorithmen für rechnergesteuerte Industrieroboter. *Regelungstechnik*, 32, 7, 215-225.

Patzelt, W., 1981. Zur Lageregelung von Industrierobotern bei Entkopplung durch das inverse System. *Regelungstechnik 29*, 411-422.

Pavlov, A.A., 1966. *Synthesis of relay time-optimal systems*, in Russian, Publishing house "Nauka".

BÉZIER FITTING TO ALMOST OVAL GEAR DEVICES

Karol Gajda [(a)], Piotr Krawiec [(b)], Adam Marlewski [(c)]

[(a)] Institute of Mathematics, Poznan University of Technology, Poland
[(b)] Chair of Basics of Machine Design, Poznan University of Technology, Poland
[(c)] Institute of Mathematics, Poznan University of Technology, Poland

[(a)]karol.gajda@put.poznan.pl, [(b)] piotr.krawiec@put.poznan.pl, [(c)] adam.marlewski@put.poznan.pl

ABSTRACT

Bézier curves are widely applied in computer aided design. In the paper first we shortly say what these curves are and how they are constructed, with examples added to. Next we present their application to model almost oval gear devices.

Keywords: Bernstein polynomials, Bézier curves

1. BERNSTEIN POLYNOMIALS AND BÉZIER CURVES

In 1886 Karl Weierstrass proved that any continuous function on a closed and bounded interval can be uniformly approximated on that interval by polynomial to any degree of accuracy. This result, known as Weierstrass first approximation theorem, has been later shown in many ways. In 1912 Sergei Natanovich Bernstein presented the constructive proof where there are introduced polynomials known as Bernstein polynomials. A **Bernstein polynomial**, $B_{f,n}$, of the function $f \in C_{<0,1>}$ and of degree n is defined as follows

$$B_{f,n}(t) := \sum_{k=0}^{s} f\left(\frac{k}{s}\right) \cdot p_{s,k}(t), \qquad (1)$$

where

$$p_{s,k}(t) := \binom{s}{k} \cdot t^{k} \cdot (1-t)^{s-k}. \qquad (2)$$

The polynomial $p_{s,k}$ is called k-th **basic Bernstein polynomial**, and the set $B = [\, p_{s,0}, p_{s,1}, ..., p_{s,s} \,]$ is called s-th (**standard**) **Bernstein base**. The standard book discussing their properties still remains (Lorentz 1953). Bernstein polynomials and its generalizations are still investigated both in theoretical and applied aspects, and find new applications in various areas, see e.g. (Chen 2000, Ding and Zhang 2003, Hong and Mitchell 2007, John 2007, Kowalski 2006, Madi 2004).

Polynomials $p_{s,k}$ appeared no later than in 1713 when it was issued the book *Ars conjectandi*. Its author, Jacob Bernoulli (1654-1705), described there so-called **binomial distribution**, later known also as **Bernoulli distribution**. This is a discrete probability distribution,

which takes values 1 and 0 with the success probability t and value 0 with the failure probability $1-t$, resp. Then the probability to gain k times in s trials is equal $p_{s,k}(t)$. For this reason Bernstein's proof is known also as a probabilistic one.

Let $c = (c_0, c_1, ..., c_s)$ be a sequence of numbers. **Bernstein polynomial generated by** this **sequence** is the polynomial defined by the formula

$$B_c(t) := \sum_{k=0}^{s} c_k \cdot p_{s,k}(t), \ 0 \le t \le 1, \qquad (3)$$

so it is the Cauchy product, $B_c = B \cdot c$, of the Bernstein base B and the column vector $c = [\, c_0, c_1, ..., c_s \,]^{\mathrm{T}}$ corresponding to c. A number s is called a **potential degree** of this polynomial.

The definition of Bernstein polynomial can be extended by substituting a vector c by $((s+1), n)$-matrix. This matrix denoted by P, $P = [\, P_{j,k} \,]_{\, j=0,1,...,s;\, k=1,2,...,n}$, we set (again applying Cauchy multiplication) $B_P = B \cdot P$, so at each point t there holds $B_P(t) = B(t) \cdot P$. Just defined vector B_P is called a **Bernstein (polynomial) set generated by matrix** P.

With $n = 2$ this set is composed of $n = 2$ polynomials. If $P_{j,1} = t_j = j/(s+1)$ and $P_{j,2} = c_j$, then $B_P(t) = [\, t, B_c(t) \,]$, so the set B_P is simply a parametric representation of the Bernstein polynomial which is explicitly described by the explicit formula $y = B_c(t)$. This is called a **standard Bernstein approximation**.

For arbitrary values of two-column matrix P the elements of the vector B_P are polynomials, the first one is determined by the first column of P, and the second one by the 2nd column of P, in consequence B_P is a parametric equation of a flat curve. Analogously, for $n = 3$ the set B_P generated by three-column real matrix P is a parametric representation of a curve in the space $\boldsymbol{R}^3$. Usually the parameter is denoted by t, the spaces $\boldsymbol{R}^2$ and $\boldsymbol{R}^3$ are equipped in orthocartesian systems Oxy and $Oxyz$, resp. Lines of the matrix P are the coordinates of the points. The graph of the vector B_P, i.e. the set $\{\, B_P(t) : t \in \boldsymbol{R} \,\}$, is called **Beziér curve** and we say that this curve is generated by the matrix P, or is determined by so-called **control points**, i.e. the points which coordinates are given by lines of P. Referring to this interpretation we can call any line of the matrix P as its

point. The polygonal line joining control points as they are listed in *P*, is called a **Bézier polygon**, or an **initial contour of Bernstein approximation**.
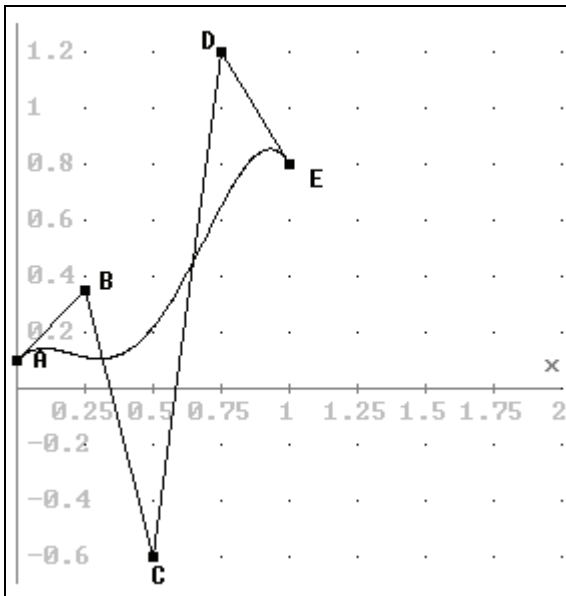

Figure 1: Bézier polygon and curve determined by the sequence *ABCDE*, where   *A* = (0, 0.1), *B* = (0.25, 0.35), *C* = (0.5, –0.6), *D* = (0.75, 1.2) and *E* = (1, 0.8).
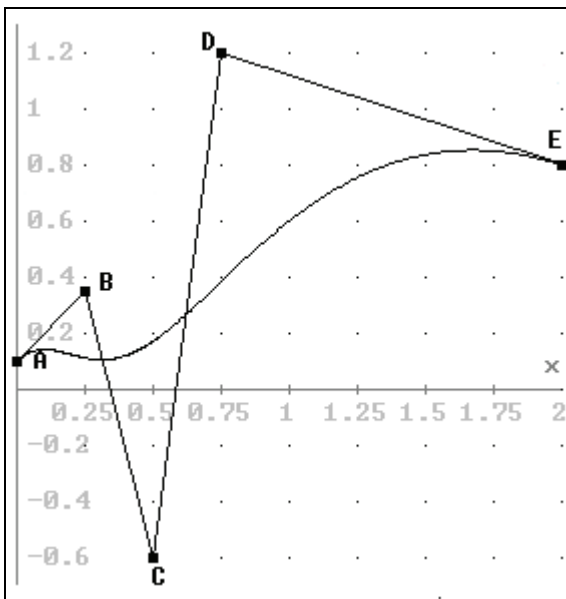

Figure 2: Bézier polygon and curve determined by the sequence *ABCDE*, where *A* = (0, 0.1), *B* = (0.25, 0.35), *C* = (0.5, –0.6), *D* =(0.75, 1.2) and *E* = (2, 0.8).

Example 1. The vector c = [0.1, 0.35, –0.6, 1.2 2]^T, the sequence   *ABCDE* = (*A* = (0, 0.1),   *B* = (0.25, 0.35), *C* = (0.5, –0.6), *D* = (0.75, 1.2), *E* = (1, 0.8)) and the matrix

$$P = \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 0 & 0.1 \\ 0.25 & 0.35 \\ 0.5 & -0.6 \\ 0.75 & 1.2 \\ 1 & 0.8 \end{bmatrix}, \tag{4}$$

determine the Bernstein set $B_P(t) = B(t) \cdot P = [t, -8.9 \cdot t^4 + 15.8 \cdot t^3 - 7.2 \cdot t^2 + t - 0.1]$, see Figure 1.

Example 2. The vector c = [0, 0.25, 0.5, 0.75, 2]^T, as well as the matrix *P* differening form the above one only at the entry $P_{4,2} = E_2$, now E_2 = 2, generate the Bernstein polynomial $y = B_c(t)$, where $B_c(t) = B(t) \cdot P = t^4 + t$, see Figure 2.

Example 3. With the points *A*, *B*, *C*, *D* and *E* as in Example 2 we see that both the sequence *BACDE* and the matrix *P* with lines storing these points in the indicated order generate the Bernstein set $B_P(t) = B(t) \cdot P = [2.25 \cdot t^4 - 4 \cdot t^3 + 4.5 \cdot t^2 - t + 0.25, -7.65 \cdot t^4 + 11.8 \cdot t^3 - 2.7 \cdot t^2 - t + 0.35]$.

The elements of this vector denoted by *x* and *y*, we have the relation $[x, y] = [2.25 \cdot t^4 - 4 \cdot t^3 + 4.5 \cdot t^2 - t + 0.25, -7.65 \cdot t^4 + 11.8 \cdot t^3 - 2.7 \cdot t^2 - t + 0.35]$.

It is the parametric equation of the curve drawn, as the parameter $t \in <0, 1>$, in Figure 3.

Example 4. With the same points *A*, *B*, *C*, *D* and *E* as above, but taken in the different order, namely as the sequence *BACED*, we have Bernstein equation $[x, y] = [-4 \cdot t^4 + t^3 + 4.5 \cdot t^2 - t + 0.25, -5.65 \cdot t^4 + 10.2 \cdot t^3 - 2.7 \cdot t^2 - t + 0.35]$. Its graph, as well as the initial contour of Bernstein approximation, is drawn in Figure 4.

In the years 1959-62 Pierre Bézier and Paul de Casteljau started, in Renault and Citroën car enterprises, to approximate shapes by curves governed by Bernstein sets. This approach gained a wide popularity in the mathematical modeling of various shapes and the designers using it started to call the graphs of Bernstein equations as Bézier curves. Within the area at hand they are used only when the parameter *t* runs from 0 to 1, and therefore a **Bézier curve** is an arc starting at the point produced by $t = 0$ and having its other end for $t = 1$; these points are called a **starting**, or **zero**, point and a **last**, or *s*-th, **point**, if this curve is determined by *s*+1 control points. Directly from the definition of the polynomial B_P it is easy to see that $B_P(0) = P_0$ and $B_P(1) = P_s$, if the lines of the generating matrix *P* are indexed from 0 up to *s* and, at the same time, *j*-th line identifies the *j*-th control point P_j, *j*=0,1,2,...,*s*.
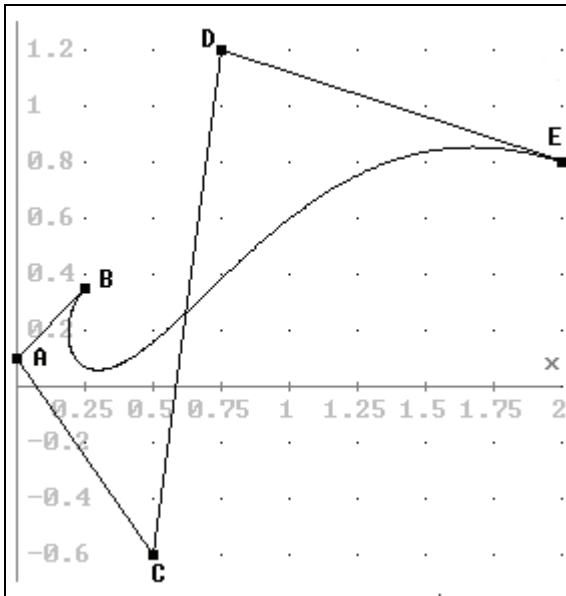
Figure 3: Bézier polygon and curve determined by the sequence *BACDE*, where *A*, *B*, *C*, *D* and *E* are as in Figure 2.
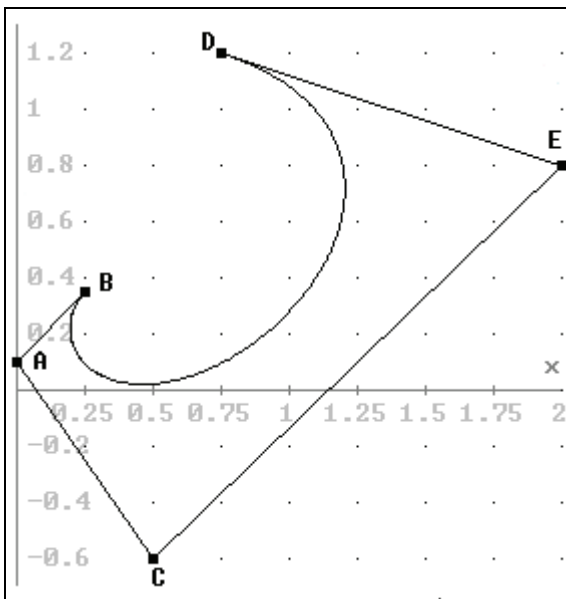

Figure 4: Bézier polygon and curve determined by the sequence *ABCED*, where *A*, *B*, *C*, *D* and *E* are as in Figures 2 and 3.

2. DE CASTELJAU ALGORITHM

Direct computation of any value $B_P(t)$ with $t \in (0, 1)$ needs to know the values of $p_{s,k}(t)$, $k=1,2,...,s-1$, so there have to be known the binomial coefficients $s!/\{k! \cdot (s-k)\}$. These operations are hard in programming: there are quickly produced really large numbers (e.g. $20! = 2\,432\,902\,008\,176\,640\,000$, $10! = 3\,628\,800$, $20!/10!^2 = 184\,756$) and the range for integer type variables (this overcome usually is not signalized by computers), in floating-point arithmetic there arise roundings/truncations which may totally destroy results. Fortunately, these disadvantages take no place when we apply de Casteljau algorithm. This is

the sequence of operations, and each one of them maps the matrix and a number *t* (it is, as always, the parameter of the representation) into a matrix having one less line. Lines of the input matrix denoted by $a_1, a_2, ..., a_r$, this transformation, *T*, works as follows

$$T \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ ... \\ a_{r-1} \\ a_r \end{bmatrix} = \begin{bmatrix} (1-t) \cdot a_1 + t \cdot a_2 \\ (1-t) \cdot a_2 + t \cdot a_3 \\ (1-t) \cdot a_3 + t \cdot a_4 \\ ... \\ (1-t) \cdot a_{r-1} + t \cdot a_r \end{bmatrix}. \tag{5}$$

Thus *T* acts over each column independently. It can be represented via the matrix, it transforms a *r*-dimensional vector W_r into the $(r-1)$-dimensional vector W_{r-1}, simply by Cauchy multiplication $W_{r-1} = TW_r := T(W_r) = T_r \cdot W_r$, where T_r is the matrix of operation *T*, i.e. it is the $(r-1, r)$-matrix

$$T_r = \begin{bmatrix} u & t & & & \\ & u & t & & \\ & & \ddots & \ddots & \\ & & & u & t \\ & & & & u & t \end{bmatrix}, \tag{6}$$

where $u := 1 - t$, and all elements outside both filled lines are 0.

As above, $P_0, P_1, ..., P_s$ are control points, they are consecutive lines of the generating matrix *P*, $P = [\, P_j \,]_{j=0..s}$, and they form the sequence $P_0P_1P_2...P_s$. **de Casteljau algorithm**, realized on the sequence $P_0P_1P_2...P_s$, i.e. on the matrix $W_{s+1} := P$, and with the value *t*, is the computation of consecutive vectors $W_s := T_{s+1}W_{s+1}$, $W_{s-1} := T_sW_s$, ..., $W_1 := T_2W_2$. One sees that the final value

$W_1 = T_2W_2 = T_2T_3W_3 = ... = T_2T_3...T_{s+1}W_{s+1} = TP = B_P(t)$, where *T* is the superposition of all transformations T_j, so $T := T_2T_3...T_s$. (and, in matrix approach, the product of all matrices T_j,). The matrix W_i is called *t*-th **de Casteljau matrix**, and its graph is named **de Casteljau contour**.

Example 5. Let's illustrate the work of de Casteljau algorithm with the initial contour *BACED*, where *A*, *B*, *C*, *D* and *E* are as in Examples 3 and 4, and with $t = 0.8$. We have

$$W_5 = P = \begin{bmatrix} B \\ A \\ C \\ E \\ D \end{bmatrix} = \begin{bmatrix} 0.25 & 0.35 \\ 0 & 0.1 \\ 0.5 & -0.6 \\ 0.75 & 1.2 \\ 2 & 0.8 \end{bmatrix} \tag{7}$$

and it is the first of de Casteljau matrices. Next ones are

$$W_4 = \begin{bmatrix} 0.05 & 0.15 \\ 0.4 & -0.46 \\ 0.7 & 0.84 \\ 1.75 & 0.88 \end{bmatrix}, \tag{8}$$

$$W_3 = \begin{bmatrix} 0.33 & -0.338 \\ 0.64 & 0.58 \\ 1.54 & 0.872 \end{bmatrix}, \tag{9}$$

$$W_2 = \begin{bmatrix} 0.578 & 0.3964 \\ 1.36 & 0.8136 \end{bmatrix} \tag{10}$$

and, finally, the last one is

$$W_1 = [1.2036, 0.73016] = B_P(0.8). \tag{11}$$

The graph of the input matrix P and graphs of all other de Casteljau matrices are shown in Figure 5. The work with $t = 4/5$ produces successive control points sitting on segments of polygonal lines and divided any of them in the proportion 4:1. The last matrix, W_1, has one line only, its graph is the point. In Figure 5 this is the point W. It is the point corresponding to the value $t = 0.8$.
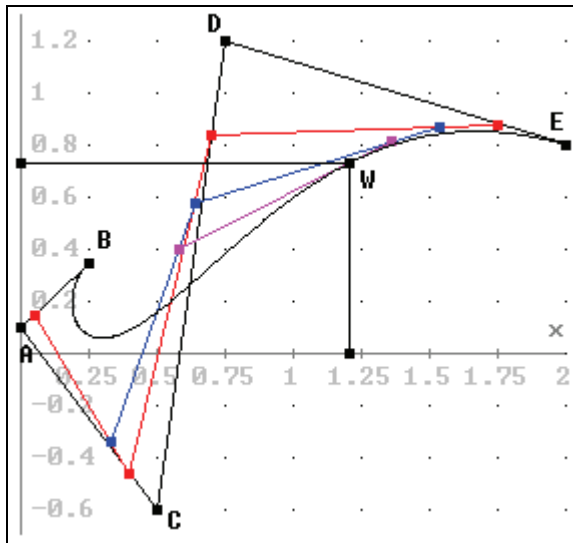

Figure 5: Consecutive de Casteljau contours generated by the points sequenced *BACED* and with $t = 0.8$; the last contour is the point $W = B_P(t = 0.8)$.

3. DEGREE RISING

In modeling a Bézier curve, the basic operations are the change of (coordinates of) control points, the increasing in number of control points, the deletion of a control point. It means the change of elements in the generating matrix P, the increase or decrease of lines of P, respectively. In general, any such change affects the entire curve.

   Potential degree rising of Bézier curve is the process resulting in the replacing of a given matrix P by the matrix $\widetilde{P}$, which row degree is greater by 1 than

that of P and producing the same curve. It is enough to describe this procedure in a standard case, i.e. when the Bézier curve is defined by the vector $c = [\, c_0, c_1, ..., c_s\,]^T$. In view of the identity $c_k = t \cdot c_k + (1-t) \cdot c_k$ it is easy to see that

$$B_c(t) = \sum_{k=0}^{s} p_{s,k}(t) \cdot c_k =$$
$$\sum_{k=0}^{s+1} p_{s+1,k}(t) \cdot w_k = B_w(t), \tag{12}$$

where $w_0 = c_0$, $w_{s+1} = c_s$ and

$$w_k = \frac{k}{s+1} \cdot c_{k-1} + \left(1 - \frac{k}{s+1}\right) \cdot c_k \quad \text{for } k=1,2,...,s. \tag{13}$$
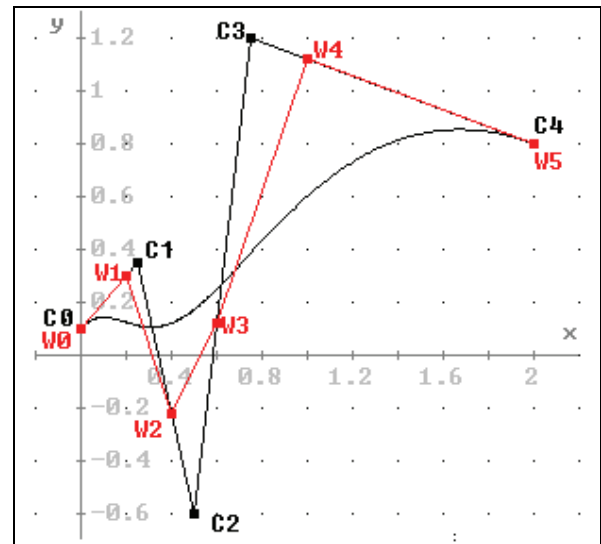
Figure 6: Bézier curve generated by any of vectors:
$c = [\, 0, 0.25, 0.5, 0.75, 2\,]^T$,
$w = [\, 0.1, 0.3, -0.22, 0.12, 1.12, 0.8]^T$

Example 6 (see Figure 6) The vectors $c = [0, 0.25, 0.5, 0.75, 2\,]^T$ and $w = [\, 0.1, 0.3, -0.22, 0.12, 1.12, 0.8]^T$ determine the same Bézier curve, $y = t^4 + t$. Denoting the augmented matrices of c and w by C and W, respectively, we have

$$C = \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} 0 & 0.1 \\ 0.25 & 0.35 \\ 0.5 & -0.6 \\ 0.75 & 1.2 \\ 2 & 0.8 \end{bmatrix} \tag{14}$$

$$W = \begin{bmatrix} W_0 \\ W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \end{bmatrix} = \begin{bmatrix} 0 & 0.1 \\ 0.2 & 0.3 \\ 0.4 & -0.22 \\ 0.6 & 0.12 \\ 1 & 1.12 \\ 2 & 0.8 \end{bmatrix}. \qquad (15)$$

Degree rising is applied when Bézier curve approximates a designed shape in a degree not good enough and probably the change in the positioning of the control points does not results in a better fitting.

In spite of the name of the considered procedure, the polynomials B_c and B_w are of the same algebraic degree, but, in general, even a very small change in w makes the degree of B_w higher. The inverse process to the rising, the deletion of any point results, in general, in decrease of the algebraic degree, so it produces another Bézier curve. We apply it when a Bézier curve determined by $s+1$ points is good enough and we want to check whether it is possible to have good approximation with s control points.

4. SMOOTH JOINT OF BÉZIER CURVES

We say that two polynomials, U and W, **smoothly meet** each other at a point τ, if at this point they have the common value and their derivatives do it, too, i.e. if $U(\tau) = W(\tau)$, $U'(\tau) = W'(\tau)$.

Therefore Bernstein polynomials B_c and B_d, determined by vectors $c = [\, c_i \,]_{i=0..\gamma}$ and $d = [\, d_j \,]_{j=0..\delta}$, resp., are smoothly joined at a point A, if

- A is the last point of the curve B_c and the initial point of B_d,
- A is collinear with the point no. γ–1 of B_c and the point no.1 of B_d.

For $n = 2$, so on the plane R^2, the smooth joint takes place if $\gamma \cdot \{c_\gamma - c_{\gamma-1}\} = \delta \cdot \{d_1 - d_0\}$. Consequently, Bézier curves generated by matrices C and D meet smoothly at the point $A = C_\gamma = D_0$, if $\gamma \cdot \{A - C_{\gamma-1}\} = \delta \cdot \{D_1 - A\}$, where C_i and D_j denote i-th point of C, i.e. i-th line of the matrix C, and j-th point of D.

The inverse process, the **subdivision** of a given Bézier curve generated by a matrix B into two Bézier curves which meet each other at its arbitrary point, Q, means algebraically to produce two matrices, C and D, generating such curves. Now Q is the last point of the matrix C and the initial point of the matrix D.

5. INVERSE BERNSTEIN APPROXIMATION

The **inverse (standard) Bernstein approximation** consists in determination of the vector c such that Bézier curve generated by c, i.e. $y = B_c(t)$ and $0 \le t \le 1$, coincides with a given curve $y = f(t)$ at $s+1$ points, see e.g. (Becker 1979). We find the vector $c = [c_0, c_1, c_2, ..., c_{s-2}, c_{s-1}, c_s]^T$ by taking $s+1$ points $F_j = (t_j, f_j)$ with abscissas $t_j := j \cdot h$, where $h := 1/s$, $j=0,1,2,...,s$, and ordinates $f_j = f(t_j)$. The defining condition takes form $B_r(t_j) = f_j$ for $j = 0, 1, ..., s$, and its

matrix form is $M \cdot c = f$, where, with no reason to confuse by using the same letter f, $f = [f_0, f_1, f_2, ... , f_{s-2}, f_{s-1}, f_s]^T$, $M = [\, m_{j,k} \,]_{j, k = 0, 1, 2, .., s}$, $m_{j,k} = p_{s,k}(j \cdot h)$.

The resolving system $M \cdot c = f$ may be at once reduced by 2, because from the equalities $B_r(0) = r_0$, $B_r(1) = r_s$ it follows $c_0 = f_0$, $c_s = f_s$. If we know values $d_0 = f'(0)$ and $d_s = f'(1)$, then we may reduced it by 2 again, because $s \cdot \{c_1 - c_0\} = d_0$, $s \cdot \{c_s - c_{s-1}\} = d_s$.

Example 7. We find the vector c, which generates the Bernstein polynomial B_c assuming at points $j \cdot h$, $j=0,1,..,s$, $h=1/s$, $s=6$, values $f(j \cdot h)$, where $f(t) = \sin(\pi \cdot t)$. Now

$$f = [\, 0, 1/2, \sqrt{3}/2, 1, \sqrt{3}/2, 1/2, 0\,]^T, \qquad (16)$$

$$M = \frac{1}{46656} \begin{bmatrix} 46656 & 0 & 0 & 0 & 0 & 0 & 0 \\ 15625 & 18750 & 9375 & 2500 & 375 & 30 & 1 \\ 4096 & 12288 & 15360 & 10240 & 3840 & 768 & 64 \\ 729 & 4374 & 10935 & 14580 & 10935 & 4374 & 729 \\ 64 & 768 & 3840 & 10240 & 15360 & 12288 & 4096 \\ 1 & 30 & 375 & 2500 & 9375 & 18750 & 15625 \\ 0 & 0 & 0 & 0 & 0 & 0 & 46656 \end{bmatrix} \quad (17)$$

and the solution of the resolving system $M \cdot c = f$ is

$$c = [0, 6160 - 3375q, 9720q - 16208, 22536 - 12555q,$$
$$9720q - 16208, 6160 - 3375q, 0]^T/600, \qquad (18)$$

where $q := \sqrt{3}$. In Figure 7 we see points P_j with ordinates equal to these numbers, they sit on the polygon of the Bézier curve determined by them. Moreover, we see the graph of the approximation error (to be visible, it is magnified 10000 times). The Bézier curve we produced practically coincides with the given arc $y = \sin(\pi \cdot x)$.
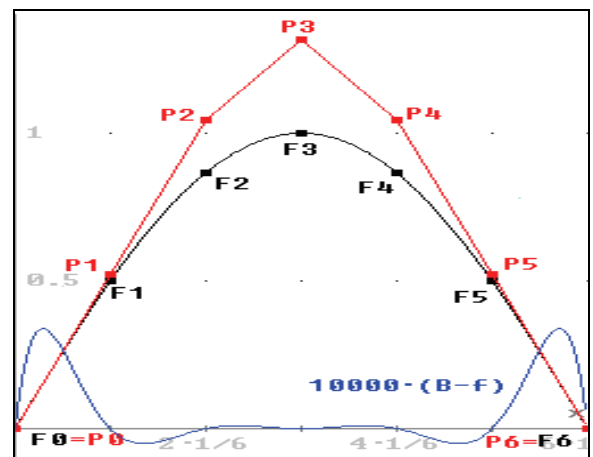

Figure 7: Graphs $y = f(x) = \sin(\pi \cdot x)$, $x \in <0, 1>$, points $F_j = (j/6, f_j = f(j/6))$, $j = 0, 1, ..., 6$, sitting on this graph, points $P_j = (j/6, c_j)$ generating the Bernstein polynomial B_P and the graph $y = 10^4 \cdot \{B_P - f\}(x)$ showing the error of the approximation of the function f by the polynomial B_P.

6. APPROXIMATING NONCIRCULAR WHEEL

Let's consider a variable-speed toothed belt transmission system created by combining the geometric and kinematic characteristics of a non-circular transmission system with a timing belt transmission system.

The required degree of speed variability is obtained by the use of pulleys constructed with wheel rims having shapes of ellipses, ovals or non-circular disks (figure 8).
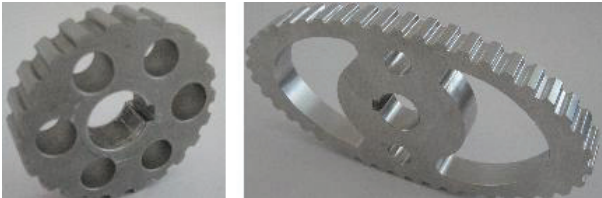

Figure 8: Wheels of the variable-speed transmission system: a) the noncircular wheel, b) the elliptical wheel

The construction must meet some conditions. Namely, the length of the belt must be equal to the length of system envelope. In order to ensure correct operation of a variable-speed transmission system the active and passive sections of the belt must be tight thanks to constant action of applicable force. Circumferences of the wheels are the product of the pitch of the belt and an integer number. Thus, one is able to determine the average transmission ratio of the system as the relation of wheels' circumferences or their number of teeth. Driving cyclicity can be guaranteed only by toothed belts whose plastic strain will increase during operation only slightly. At the same time the belts must be initially pre-tightened in order to avoid belt's slip or skipping on the teeth of the wheels.

A model of a two-wheel transmission system consisting of an elliptical driving wheel and a noncircular driven wheel is presented in Figure 9.
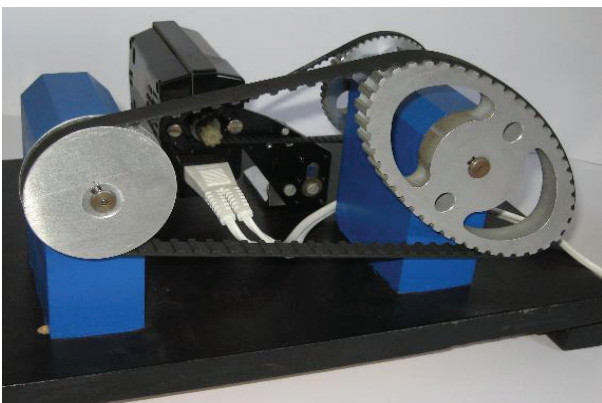

Figure 9: The view of variable-speed toothed belt transmission system with an eccentrically mounted wheel

In the research on effective parts of some devices (see Figure 8 – photo) there appear problems to get their mathematical expressions for the border line in aim to, e.g., to reproduce them on computer-controlled punch

or cutting machineries, to project teeth of gears. In considered case the curve is close to, but it is not an ellipse or any other standard shape. Therefore we look for the Bernstein parameterization. There are offered computer programs to find it, e.g. (Zhao and Shene 1999), but we worked in our own program, called BezierFit, elaborated within Delphi system from Code Gear.
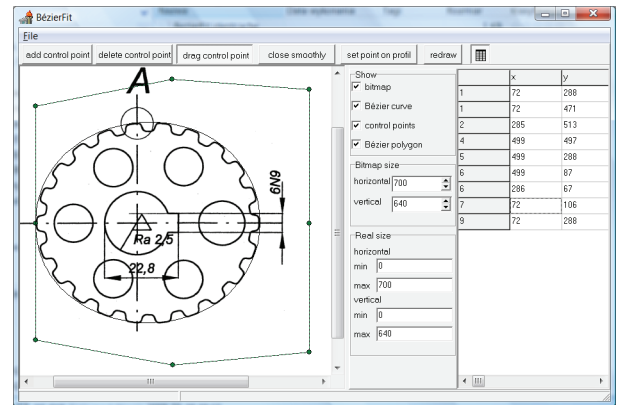

Figure 10: BezierFit program by Karol Gajda

Working in this program we look at the screen, see Figure 10. In the background of its working area there is displayed, from a BMP file, the line to be modeled, and there is the reason we opt for our own program. In this area we may place controlling points (the first placing defines the input matrix P), we can move them, we can delete any point, we can add points in the way to keep up-to-now curve unchanged. Every operation is at instance visualized as the Bézier curve. In the operational area of the screen there are buttons activating offered operations. Here are, e.g., the positioning of control points (by mouse or by setting the coordinates), the memorizing of produced coefficients (they are saved as a text file *.BFP), smooth closing (choosing it we close smoothly the profile, i.e., in the way that the last point coincides with the starting point P_0), the positioning of points on a given profile (and then, by the inverse Bernstein approximation, the program produces the appropriate control points).

REFERENCES

Becker, M., 1979. An elementary proof of the inverse theorem for Bernstein polynomials. *Aequationes mathematicae*, 19(1), 145–150.

Chen, Y., 2000. Some thoughts on the polynomial approximations to a given function. *Chinese J.of Physics*, 38(5), 927–938.

Ding, R., Zhang, Y., 2003. The extension of the dual de Casteljau algorithm, *Proceedings of 4th Int.Conf. on Parallel and Distributed Computing, Applications and Technologies*, 68–92, 2003, Chengdu, IEEE Press.

Hong, X., Mitchell, R.J., 2007. Hammerstein model identification algorithm using Bezier-Bernstein approximation. *IET Control Theory Appl.*, 1(4), 49–59.

John, Ch.T., 2007. All Bézier curves are attractors of iterated function systems. *New York J.Math.*, 13, 107–115.

Kowalski, E, 2006. Bernstein polynomials and Brownian motion, *American Mathematical Monthly*, 12.

Lorentz, G.G., 1953. *Bernstein polynomials*, Toronto: University of Toronto Press.

Madi, M., 2004. Closed-form expressions for the approximation of arclength parametrization for Bézier curves, *Int.J.Appl.Math.Comput.Sci.*, 14(1), 33–41.

Zhao, Y., Shene, C.K., 1999. *DesignMentor program*. Department of Computer Science, Michigan Technological University. Available from: http://lumimath.univ-mrs.fr/~jlm/cours/DMmanuals/curve/ [accessed 26 April 2008].

MODELING AND SIMULATION NEEDS IN FUSION ENERGY RESEARCH

Gábor Veres[a]

[a]KFKI Research Institute for Particle and Nuclear Physics, P.O.Box 49, H-1525 Budapest, Hungary

[a]veres@rmki.kfki.hu

ABSTRACT

Fusion energy research has reached a state, where a transition from laboratory scale physics experiments to industrial scale reactors has become possible. Current devices can routinely reach plasma conditions needed for a fusion reactor, but positive energy balance can only be achieved after a further increase of the machine size by about a factor of two (compared to the currently largest device – JET – in Culham, UK). Although this step seems to be straightforward, it involves some remarkable qualitative changes relative to present day devices: dominance of plasma self-heating by alpha particles, huge neutron load on all components, activation of the structural materials, etc. This paper gives an overview on the status of fusion research, key components of a fusion reactor and simulation possibilities and needs in the hope that modelers in other fields can contribute to the development of fusion energy.

Keywords: fusion energy, confinement, large scale modeling, tokamaks, ITER.

1. INTRODUCTION

Since the discovery of nuclear reactions, controlled nuclear fusion has been considered as a great opportunity for energy production. Unfortunately, fusion reaction occurs only if the nuclei approach each other much closer than their own size. Moreover, due to their identical electric charge, nuclei repel each other, and the requirement for a fusion even to take place is that the particles have enough kinetic energy to overcome the repulsion. The necessary energies for the fusion of protons are in the 10 keV (kilo electronvolt) range.

Such energies can easily be achieved for example in an accelerator, therefore nuclear fusion reactions had already been studied in detail some 50 years ago. From these measurements we know that the reaction cross-section of fusion processes is so low compared to Coulomb scattering, that a fast particle beam spends most of its energy for heating the target and the net energy gain from fusion reactions in accelerator driven systems is in the 1% range.

The situation is completely different when considering a high temperature medium with mean particle thermal energies in the 10 keV range. Under such conditions collisions are necessary to only maintain the thermal equilibrium and every fusion reaction provides net energy production. Ten kilo electronvolt thermal energy corresponds to about 100 million °C, and thus in the devices for fusion energy production matter needs to be heated and sustained at these extreme temperatures.

Among various possible reactions the fusion of Deuterium and Tritium into Helium and a neutron (DT reaction) occurs at the lowest threshold energy. This reaction releases about 17 MeV energy, of which 14 MeV (80%) is carried away by the neutron. This energy per fusion reaction is very high and for a 1GWe power plant only a couple of hundred kg of fuel material would be needed annually. As for the fuel materials, Deuterium is abundant on Earth but Tritium is practically not available. To produce the necessary Tritium one can make use of other nuclear processes, where Tritium is generated from Lithium nuclei using the neutrons released in the fusion reactions. Using a combination of these processes the scheme of a fusion reactor can be established as shown in Fig. 1.
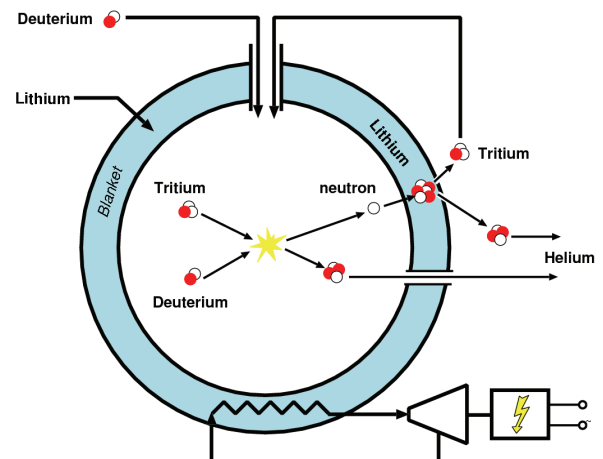
Figure 1: Theoretical scheme of a fusion reactor.

As it can be seen, the reaction chamber is surrounded by a blanket, in which the Tritium is bred. The input is only Deuterium (e.g. from water) and Lithium (e.g. from the Earth's mantle), and the output is Helium. Some neutron multiplication materials (e.g. Lead or Beryllium) are also needed to replace the neutrons inevitably lost in nuclear reactions with the structural materials. Some nuclear waste would be generated in these latter reactions, but their amount can be limited by a careful choice of the structural materials and their radioactivity would completely decay in about 100 years. Therefore fusion reactors would be environmentally benign, since *both their fuel and their waste are not radioactive*.
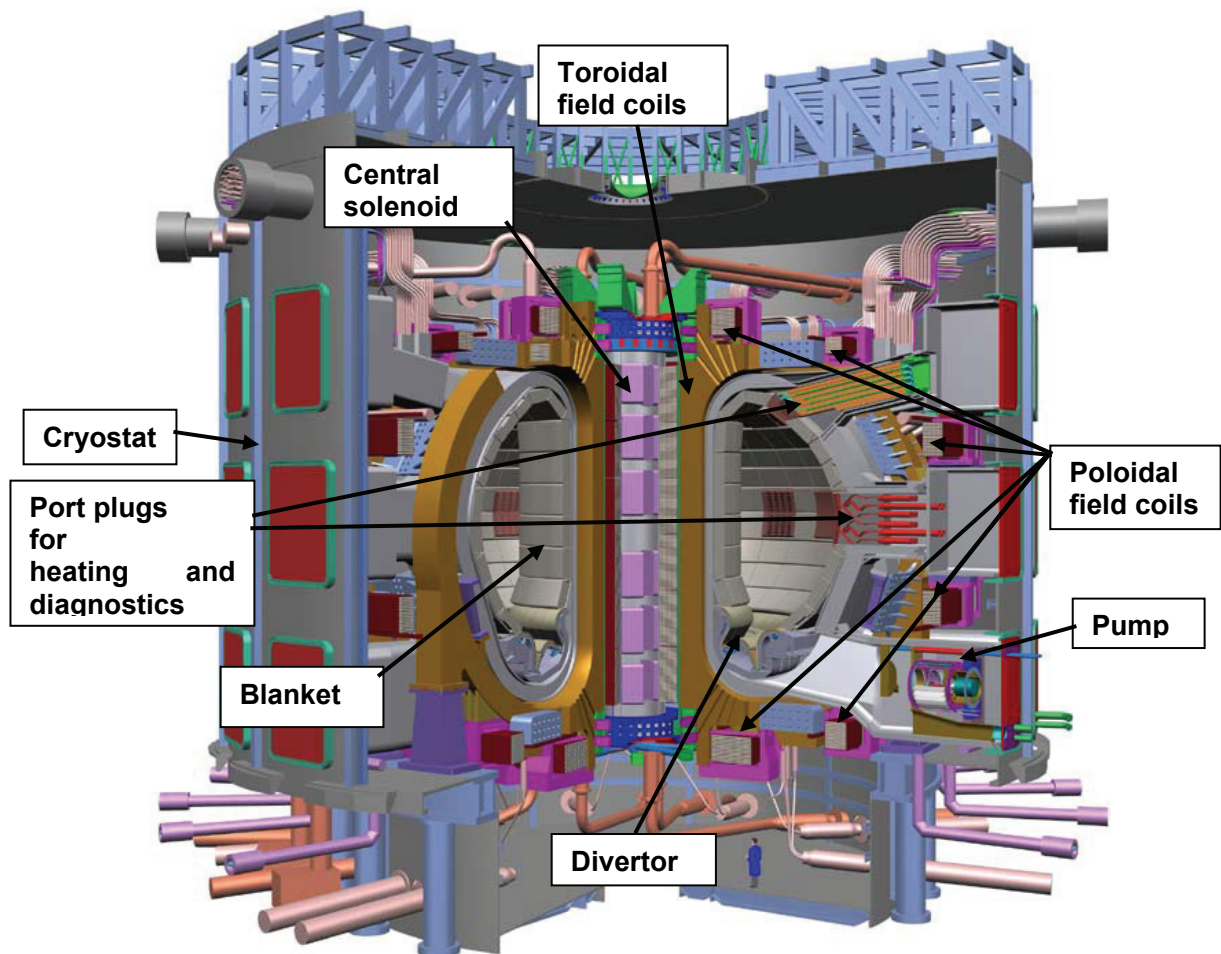
Figure 2: Design of ITER. The key components of a tokamak device are also shown

2. THE STATUS OF FUSION RESEARCH

The above theoretical scheme can be implemented in two ways: by micro-explosions in the chamber (inertial confinement, ICF) or by magnetic confinement of the hot plasma. During the past 5 decades both concepts went through a series of detailed physical experiments, and reached the status where the next generation devices are expected to show positive energy balance. In the case of ICF this will be achieved in single explosions, releasing little energy and the extrapolation to high-frequency (about 10 Hz) repetitive operation still demands substantial technical development for the cost effective and reliable manufacturing of the capsules and operation of driver devices like laser or particle beams.

Magnetic confinement fusion reached a state, where the design of a reactor scale experiment became possible. The International Experimental Thermonuclear Reactor (ITER, ITER 1999) is expected to deliver 500 MW fusion power in discharges up to 1000 s in length. This device will not be a power plant, but rather an experiment and will hopefully provide the physical and technical basis for one or more demonstration fusion reactors, like DEMO (EFDA 2005). Due to their large size and extreme complexity, construction of each of these devices will last at least 10

years and experiments will also take decades. To shorten this time and increase the reliability of the experiments, modeling of all structural components is an absolute necessity.

Compared to previous magnetic fusion devices, ITER represents a major challenge and calls for a cultural change. Up to now experiments were mostly designed and operated by physicists and the major design aspect was flexibility. Although reliable operation of major components was necessary, it was always possible to react to unforeseen problems and new ideas in a flexible way. In ITER the situation will be quite different. The device will be a nuclear installation, therefore safety, licensing, reliability and remote handling becomes critical. Once built, any change to the device (even smaller diagnostic elements) becomes extremely cumbersome and time-consuming. The necessary reliability calls for quality assurance and detailed modeling of all possible operating scenarios. Magnetic fusion research faces a transition from physics research to industrial technology development.

3. STATUS OF COMPONENTS AND KEY TECHNOLOGIES

3.1. Plasma confinement

At 100 million °C temperature the mean kinetic energy of the gas atoms is well above the binding energy of the

electrons, therefore the atoms are completely ionized, i.e. the ions and the electrons form a plasma. The basis of confinement is the fact that a magnetic field forces the particles onto spiral trajectories around the magnetic field lines. As movement along the magnetic field is free, such a configuration must be used in which field lines are not leaving the plasma. The only topological arrangement fulfilling this criterion is a torus. For a good confinement, the field lines not only need to lie on toroidal surfaces, but also need to cover it sufficiently dense. The resulted configuration is a set of nested magnetic surfaces with helical magnetic field.
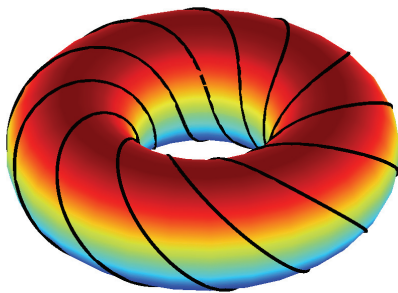
Figure 2: Illustration of field lines and magnetic surfaces in a toroidal confinement device.

There are several ways to create this plasma configuration, the most advanced concept being the tokamak as shown in Fig. 2. The main (toroidal) component of the magnetic field is provided by external coils (toroidal field coils), while winding of the field lines onto tori is done by the magnetic field of a strong current in the plasma itself. In magnetic confinement the kinetic pressure of the plasma is compensated by the magnetic pressure. Since the field strength is limited by cost and technology and the plasma temperature requirement is also given, the plasma density must be limited to values about 10^{-5} times the atmospheric density.

Plasma force equilibrium is very well described by magneto-hydrodynamics (MHD) if the transport coefficients (heat conduction, particle diffusion, etc) are known. For the plasma equilibrium some additional magnetic field components are required, and these are generated by the poloidal field coils.

There is a large variety of waves in plasmas, which – besides MHD – often involve kinetic or even single particle motion effects. Although these waves can be rather complex, they can be individually described with reasonable accuracy. The energy and particle losses of the plasma through radiation and charge exchange can also be well calculated. Despite of these encouraging results, transport losses (e.g. heat and particle diffusion) are not fully understood. So called neoclassical transport theory, describing transport via single particle motion and collisions agrees with experiments neither

in the order of magnitude nor in trends. It is believed that effective transport is driven by micro-turbulence in the plasma involving a spatial scale of 2 3 orders of magnitude below the machine size, and a temporal scale of 4 5 orders of magnitude below the global transport timescale. Most recent experiments and theories show, that turbulence is highly nonlinear and involves interaction between different structures, waves and profiles.

Bifurcations are often observed in the experiments providing different plasma operation modes. Although simulations and experiments have agreed on some basic features and mechanisms in micro-turbulence (PPCF 2006) it is currently not possible, and will most probably not be possible in a decade, to describe plasma transport in a predictive way.

To overcome the lack of a predictive theory, plasma confinement scaling laws have been established. These describe global confinement properties and other global parameters in the form of a power function of various machine and plasma parameters. Constants of the scaling laws are fitted by regression to data from a carefully selected set of comparable measurements on various devices. It is encouraging, that e.g. the effective transport loss in various experiments can be fitted over 2 orders of magnitudes with an accuracy of a factor of two. These scaling laws give the basis of the ITER design with a good reliability.

It has to be noted that besides the most advanced tokamak concept, some alternative toroidal devices are also being developed. The compact tokamak promises higher plasma pressures and increased plasma volume compared to the device size. Various stellarator concepts provide the helical twist of the magnetic configuration via a helically distorted geometry. This solution does not need plasma current and avoids all problems associated with it (see a next section), therefore the stellarator concept can easily be extrapolated to continuous operation. Modern numerical simulation techniques enabled optimization of the inherently three-dimensional stellarator configurations according to various parameters and a wide variety of configurations are being tested in various laboratories, including two large superconducting machines: the already operating Large Helical Device in Japan and the innovative Wendelstein 7 X device under construction in Germany. For an overview of fusion research see Braams and Stott 2002.

3.2. Magnets

In today's fusion devices magnets are usually made of copper conductors. Although they fit to the short pulse lengths (< 1 minute) needed for physics experiments their excessive power dissipation makes them unsuitable for continuous reactor operation. Tokamak devices have been built with superconducting magnets since the 1970s and two new devices started operation very recently in China and South Korea. The technology is mature to be used in ITER, but the size, the high magnetic field and the associated mechanical load imposes some problems which need careful design.

Prototypes of some ITER coils have been built and tested, but some development is still needed to increase the operational margin.

3.3. Heating, fuelling, current drive

In tokamak devices the current circulating in the plasma provides sufficient heating power to reach keV temperatures. To provide the 10 keV temperature for reactor conditions, additional heating has to be applied. Such techniques using particle beams and different radio-frequency waves were already developed in the 1980s and are routinely applied in today's devices. In ITER and DEMO the fusion power will be well above the external heating power. As described earlier, 20% of the energy released in the DT reaction is carried by the α particle (^{4}He nucleus), which is a charged particle, thus confined by the magnetic field. It is expected, that in ITER and DEMO alpha particles will provide most of the plasma heating.

Fusion plasmas can be fuelled by simple gas inlet, but the gas is ionized at the very edge of the plasma and the core can be fuelled only by diffusion. An alternative technique is to freeze Deuterium (or later Deuterium-Tritium mixture) and compress the resulting "snow" into a Deuterium ice pellet. This can be accelerated by various techniques to 0.1-1 km/s velocity and injected into the plasma. During penetration into the plasma these pellets shield themselves by a gas blanket (Leidenfrost effect) and thus they are capable of reaching deeper layers of the plasma. Pellet injection is a well developed technique, injection up to 100 Hz during several seconds is routinely applied on modern tokamaks.

The plasma current in tokamak devices is a necessary component of the magnetic configuration. For about 100-1000 s this current can be driven inductively using the central solenoid but for continuous operation additional techniques are needed. Some of the heating techniques are also suitable and tested for current drive. It is an exceptional feature that the pressure gradient in the plasma across the magnetic surfaces can automatically drive plasma current. There is now sufficient experimental evidence, that this so called bootstrap current exists, and it can provide up to 50% of the plasma current in future devices. Only the rest should be driven by heating particle beams and microwaves.

3.4. Plasma-wall interaction

As stated above, a fusion device must operate at much lower particle densities than found under normal atmospheric conditions. This requires a clean vacuum system capable of producing base pressures below 10^{-6} torr. Additionally, the plasma-wall interaction must be carefully designed to avoid impurities in the Hydrogen isotope plasma.

In today's fusion devices a special instrument, called a divertor, is used for controlling the plasma-wall interaction and particle exhaust. Magnetic field lines from the edge of the plasma are diverted into this region by additional magnetic coils. The closed geometry and the high power handling capacity of the plasma facing components (typically Carbon-Fiber Composites, CFC) prevent melting and impurity reflux into the plasma. Experiments showed that the particle exhaust capability of a well-designed divertor is sufficient for removing the expected Helium produced in a fusion reactor. Unfortunately it became clear that CFC is an unsuitable material for ITER and DEMO as Carbon has an unexpectedly high chemical sputtering yield in the low temperature edge Hydrogen plasma. The resulting Carbon and Hydrocarbon molecules are transported in the edge plasma and deposited in certain regions. Hydrogen isotopes, including radioactive Tritium, are buried in these layers and later often break off as small dust particles (Philipps 2003). This Tritium contaminated Carbon dust would increase the radioactivity of the device above allowed levels.

To overcome this problem a development program has been started on two major tokamak devices to test Tungsten as a divertor material. For less demanding regions both Tungsten and Beryllium are being tested. First results are encouraging but a final answer will be obtained only in ITER, where a combination of Beryllium, Tungsten and optionally CFC is foreseen. The power load and the power handling capacity of the divertor is a critical issue for ITER and DEMO.

3.5. Power extraction and tritium breeding

80% of the power from the DT reaction is carried by the neutron. Being chargeless, this particle, after leaving the magnetic field, will be slowed down in the blanket of ITER and DEMO. The blanket will be cooled by either water (in case of ITER) or by He gas (in some DEMO concepts). Electric power will be generated using conventional power generation schemes. Blanket modules will be installed inside the vacuum chamber of the device, so as to protect the vacuum vessel and superconducting coils from neutron damage and power input.

The same blanket modules must also be used for Tritium production in a fusion power plant. In today's tokamaks the fusion reactions are scarce and tritium breeding technologies cannot be tested. It is one of the aims of ITER to develop and test various tritium breeder concepts without providing all the tritium for its operation. Several concepts exist for this purpose, but a liquid Lithium-Lead based and a Beryllium-Lithium ceramics system are the best explored. According to calculations, both concepts can provide the necessary Tritium for DEMO but technical issues like Helium gas cooling, corrosion, material properties must be carefully checked in ITER.

The output of all pumping stations of a fusion reactor must be connected to a tritium plant, where isotopic separation techniques will be used to extract tritium. Such techniques have been developed, and will be extensively tested under operation conditions in ITER. It is worth noting that the particle confinement time in the plasma is much shorter than the tritium fusing rate, therefore tritium will have to be recycled several times.

3.6. Diagnostics and plasma control

Measurements (called diagnostics) on 100 million °C plasmas cannot be made by conventional industrial techniques. Except the very edge, no material probes can be inserted into the plasma, therefore a broad range of new technologies have been developed in the past decades (Hutchinson 2002). These use an extremely wide range of physical processes and incorporate the use of powerful lasers, atomic beams, microwaves, particle detectors but also simple pickup coils and current collector probes. A modern fusion device relies on typically 30-40 diagnostic systems which measure all important parameters of the plasma (density, temperature, magnetic field, etc) with the necessary spatial and temporal resolution. The diagnostic systems are well developed, but the high neutron dose and plasma load on plasma facing diagnostic components will be a challenge for ITER.

A special collection of diagnostics have been developed to diagnose plasma turbulence, which manifests itself in the fluctuation of density, temperature and other quantities. The measurements are extremely difficult and further development of turbulence diagnostics are necessary.

As up to now, fusion devices are typically operated in 5-10 s long pulses (shots), and control is done by trying to force preset time evolution for certain parameters, collect as much data as possible and evaluate them offline. The amount of data from one discharge rose into the Gbyte range but still can be handled. With the new generation of devices reaching 1000 s pulse lengths this approach will have to be changed. Pulses will be cut into segments and control systems will have decision making algorithms. The strategy of researchers to collect all data will also have to change as the amount of data from one experiment will jump into the several Tbyte range.

3.7. Structural materials

Neutron damage of structural materials is expected to be one of the major problems in fusion power plants. In a fission power plant the neutrons are slowed down in the coolant and/or the moderator before reaching the reactor vessel. In a fusion reactor this is not possible. Due to the low plasma density the full flux of 14 MeV fusion neutrons will reach the vessel wall. In one collision with an atom in the structural material the neutron transfers much more energy than the binding energy of the atom in the crystal lattice. The fast atom will collide with other atoms and finally a full cascade of displaced atoms is produced.

Additionally to elastic collisions, neutrons also trigger nuclear reactions in the structural materials. Some of the resulting nuclei are radioactive, therefore the structure of the device is activated. Helium nuclei (α particles) are often generated in these reactions, which means that irradiation produces Helium gas inside the materials. The gas forms small bubbles which later fuse to larger ones and produce voids in the materials.

Small material samples are irradiated in the core of fission research reactors and their mechanical properties are tested after irradiation. A special low-activation fusion steel material called EUROFER has been developed and characterized in this way, which is now the first candidate for structural material of the DEMO reactor. For ITER, the irradiation damage is not critical as the fusion power is 5-6 times below the requirement for DEMO and ITER will only be operated in pulses.

The irradiation studies in fission reactors are lengthy and cannot give final answer because neither the spectrum nor the dose is identical to what is expected in DEMO. A special test facility called IFMIF (International Fusion Material Irradiation Facility) should be built in parallel with ITER to test materials up to DEMO irradiation doses. IFMIF will be an accelerator driven device, where high-power deuterium beams will fired onto a flow of liquid Lithium target and produce DEMO relevant dose and neutron energy in a small volume.

Besides irradiation damage, structural materials should also withstand high temperatures in some regions. The critical point is inside the blanket, where the power is taken away by the coolant. The higher the coolant temperature, the better is the overall plant efficiency. EUROFER and other steels have a temperature limit at around 500 °C, therefore high efficiency gas cooled reactors require special materials. Some DEMO concepts suppose that advanced alloys or SiC materials can be used which would boost plant efficiency, reduce costs and also activation.

3.8. Maintenance

Unlike in a fission reactor, fuel in a fusion power plant can be replaced continuously in the form of gas inlet or the injection of frozen DT pellets. This means that fuel supply does not require regular maintenance periods. On the other hand, neutron bombardment will change the properties of the structures close to the plasma, the blanket and the divertor. Although the rate of damage for a fusion power plant is not known exactly, from irradiation tests it is expected that the blanket will have to be replaced every 5 years and the divertor even more often. Components of the device behind the blanket (vacuum vessel, coils, etc.) will last for the whole lifetime of the reactor. Blanket and divertor replacement is a major maintenance process, therefore the layout of the reactor must this take in into account. ITER will not have the neutron dose which requires replacement of its components, but due to activation remote maintenance will be a must in that case as well. For DEMO different remote handling concepts are being devised. In one scheme, the blanket is divided into a large number of smaller (~10 tons) modules which can be handled by a robotic arm system. In another concept larger banana shaped modules are handled as one unit. Here the weight is in the 150 t range and special large access ports have to be designed for the tokamak. In both concepts the divertor would be divided into modules rotated in the bottom region of the design on a rail system. These modules could be removed through some of the lower maintenance ports.

3.9. Complexity and safety of the device

Although not an individual component, the complexity of a fusion reactor as a whole poses a critical problem. Based on the experience with today's experiments, it is clear that plasma operation will depend on the simultaneous operation of a large number of critical subsystems based on completely different physical basis: heating, fuelling, current drive, diagnostics. On the other hand, power generation ceases immediately if a problem arises, as the plasma cools down within a couple of seconds at most. After the plasma discharge stops, the device will be heated only by the radioactive decay of isotopes in the activated structural materials. According to simulations (EFDA 2005) even in case of a complete loss of coolant natural convection will keep the temperature of the structure below the melting point, no meltdown or uncontrolled reaction can occur.

Another safety feature of a fusion power plant is, that at any time only very little tritium fuel can be found in the plasma, well below 1 g. About 1 kg is expected to be present in the blanket and tritium plant. Even in case of a serious accident the simulated release of radioactivity will not require evacuation of the surrounding population.

4. SIMULATION AND MODELING NEEDS

After reading the overview about the fusion power plant components, one could anticipate that modeling and simulation are necessary for practically all of them. Some of these can be fulfilled using industrial models and simulation codes, even industrially proven techniques will increase the system's reliability. However, in many cases the problems are so special that new models and techniques have to be developed. This needs insight into the involved physical phenomena and will benefit very much from a combination of physics and engineering approach.

4.1. Plasma confinement

Due to the complexity of phenomena, the plasma is not described with one model but by specialized codes. Plasma equilibrium is addressed by MHD codes, while the edge plasma (Stangeby 2000) can be described by a combination of MHD and Monte-Carlo (MC) codes. This latter models the processes involving neutral particles which are moving in the plasma as independent atoms. An additional important component of the edge plasma description is an ionization-radiation model describing excitation, radiation ionization and recombination of neutrals, thus coupling the MC and MHD codes. Phenomena occurring at the plasma-wall interaction must also be taken into account as, besides the core plasma, they provide sources and sinks. For cross-field transport neoclassical particle and plasma turbulence simulations would be extremely important in order to understand particle and heat transport across the magnetic surfaces. Neoclassical transport seems to be well understood, but turbulence simulations need strong development. Here several similarities have been identified with fluid turbulence (e.g. energy transfer between scales, nonlinear interactions), self organized and chaotic systems (self organized criticality,

bifurcations). A special area is the evaluation of turbulence measurement data, which does not involve physical modeling but often statistical simulation of random phenomena. The fusion community would surely benefit from interaction with other simulation studies.

4.2. Magnets

Superconducting fusion devices are built into a cryostat cooled down to around 5 K. During cooldown the size of structural components changes considerably, therefore the whole structure should be designed and simulated in a way to be able to keep the necessary precision (typically mm) of the magnetic field and allow for movement of components. An associated problem is to design the mechanical construction of the coil system. During operation, and especially during plasma instabilities, huge forces are exerted onto the coils, actually the magnetic field strength is usually limited by the mechanical stability. During fast transients the fast magnetic field change and current dumped from the plasma to the surrounding walls can generate huge forces on plasma-facing tiles and divertor structures. These have to be modeled using combined electromagnetic-mechanical simulations.

4.3. Heating, fuelling, current drive

Special codes already exist to help designing, optimizing and operating heating, current drive and fuelling systems. These models simulate how waves are absorbed in the plasma, how the fast particles are slowed down and how the shielding gas cloud behaves around a pellet. It is expected that with some refinements these, or similar codes will be capable of describing the next generation devices as well.

4.4. Plasma-wall interaction

The influence of the plasma on the solid state walls involves interaction with two plasma particles: electrons and ions. The mean flux of these two should be equilibrated, therefore usually an electrostatic potential drop (sheath) forms in front of the wall to keep faster electrons inside the plasma.

Electrons reaching the surface of the solid material can not kick out any atoms, therefore their effect is mostly thermal heating. On the other hand ions are more massive and have typically 2 orders of magnitude higher energy than the binding energy in the crystal therefore single atoms can be kicked off from the solid state. This is called sputtering and it is the dominating process. Additionally, some incoming ions become attached to the surface, and form layers on it. The behavior of these layers is extremely important, their characteristics depend on the history of the wall as well. As the particles sputtered from the wall also modify properties of the edge plasma, plasma-wall codes should also be coupled to edge plasma codes. Due to the above reasons plasma-wall simulations are extremely difficult. Models should be validated by experiments but even if the elementary processes can be correctly simulated the condition of the wall surface is often not known.

4.5. Power extraction and tritium breeding

Plasma radiation is absorbed at the surface of the blanket of a future reactor while power from neutrons is deposited in the volume of the blanket. This latter can be simulated with standard neutron codes, e.g. MCNP5. The geometry is usually more complicated than in case of a fission reactor, but tools are being developed to be able to transfer engineering models into neutron code input. From the deposited power standard engineering simulation codes can be used to calculate temperatures and heat transfer to the coolant. Steam generators, He gas turbines will be the same as in other power stations therefore models should be applicable.

Tritium breeding is new technology but determination of the amount of generated Tritium can also be done using models developed for fission. The penetration of tritium through breeder materials, its extraction and cleaning process is much less standard, and modeling would need an interdisciplinary approach: chemistry, physics, materials science is involved.

4.6. Diagnostics and plasma control

Diagnostics are an extremely broad field in fusion, with many techniques and physical principles used. This way modeling is also very much dependent on the actual technique in question: optics, microwaves, particle beams, etc, and a large variety of specialized codes exist for different diagnostic systems for current experiments. For future devices it became clear that these have to be complemented by detailed engineering studies on e.g. heat load on components by neutrons, remote handling operations, protection against plasma radiation, etc.

In the past decade data from diagnostic systems are being gradually used as input into plasma control systems. While in the 1980 s only magnetic pickup loops were used to control plasma position and some density diagnostics to control mean plasma density, recently the most advanced devices use on-line processed data from various diagnostics (radiation, temperature, density, magnetic field direction measurement systems) to actively control plasma operation using the heating and current drive systems as actuators. It became possible to actively control plasma instabilities, improved plasma states (so called transport barriers) and other phenomena. As plasma control systems operate on a ms timescale these on-line diagnostic data evaluation systems heavily rely on advanced data evaluation techniques like neural networks, special tomography codes, pattern recognition.

4.7. Structural materials

Up to now the development of structural materials was based largely on an experimental approach. Recently simulations started to reach a level, where they can at least guide practical work. These techniques will be of extreme importance to assess neutron damage in different materials and guide development. Irradiation tests in fission reactors up to a few dpa (displacement per atom) take years, but for characterizing DEMO materials 150 dpa would be needed. This will take even

in the planned IFMIF facility a long time, therefore a validated radiation damage code would be extremely useful.

4.8. Maintenance

As noted above fusion reactors will be activated by the neutron flux and therefore all operations will have to be done remotely. The requirements are much more demanding than in a fission power plant as the geometry is more complex and access is difficult (inside a torus), a broad range of different systems will have to be managed, components can be very heavy (up to 10 100 tonnes) and high precision is needed (mm range). Broken components will have to be taken to a hot cell and repaired there. These requirements can be fulfilled only if remote handling requirements are taken into account form the first moment if the design and operations are modeled.

4.9. Complexity and safety of the device

Complexity of a fusion device and the multitude of physical processes inside it will increase the difficulties very much. In such a machine the number of interfaces can be very high and therefore management becomes extremely important. This applies not only to components, but to simulation codes as well. Codes should adhere to certain standards, should use standardized input and output structures and should be interconnectable. The European fusion program has already started a project to develop or upgrade codes to these standards and create an integrated tokamak modeling.

5. CONCLUSIONS

Magnetic confinement fusion research has reached a maturity where industrial scale experiments are needed. The role of engineering will be extremely important in the construction and operation of these devices, and therefore industrial techniques should be used wherever it is possible. There is a broad range of simulation techniques which are needed and can be applied in future fusion experiments: finite element mechanical, thermal and electromagnetic simulations, computational fluid code, advanced data evaluation and control techniques, maintenance simulations.

Additionally to standard industrial simulations there is a range of modeling techniques which are common with fission and are needed for next generation fusion experiments: neutron transport and activation calculations, simulations for high temperature gas cooling, irradiation damage models, environmental impact and safety models.

Additionally to the above there is abroad range of modeling work which is specific to fusion. These need in-depth understanding of fusion devices and plasma physics. Some of these (especially turbulence models) are in the forefront of physics research and will probably not only use results from other fields, but will also have an impact on them.

Modeling and simulation will be a key component for the development and safe operation of next generation fusion devices, therefore the modeling

community is encouraged to take part in this promising undertaking.

REFERENCES

Braams C M, Stott P E, 2002. *Nuclear Fusion – Half a Century of Magnetic Confinement Fusion Research*. OP Publishing, Bristol, ISBN 0-750-0705-6.

EFDA, 2005. A conceptual study of commercial fusion power plants. EFDA-RP-RE-5.0, available from http://www.efda.org [accessed 10 July 2008].

Hutchinson I H, *Principles of Plasma diagnostics*.Cambridge University Press, Cambridge, ISBN 0521803896 (2002).

ITER Physics Basis, 1999. *Nuclear Fusion* 39, 2137 (complete issue).

Philipps V, 2003. *Plasma Physics and Controlled Fusion* 45, A17.

PPCF, 2006, *Plasma Physics and Controlled Fusion* Special section on Experimental studies of zonal flow and turbulence. 48, S1-S508.

Stangeby P C, 2000 *The plasma boundary of magnetic fusion devices*. IOP Publishing, Bristol, ISBN 0750305592.

AUTHORS BIOGRAPHY

Gabor Veres (born 1965) is a senior research fellow at the KFKI Research Institute for Particle and Nuclear Physics in Budapest, Hungary. He got his PhD in plasma physics in 1992. He has been working on several European fusion devices like JET (UK), ASDEX-U (Germany), TCV (Switzerland). His main research interest is magnetic confinement fusion, more closely magneto-hydrodynamics, spectroscopy and radiation processes in fusion grade plasmas.

REAL TIME COLLISION DETECTION FOR COMPLEX SIMULATIONS BASED ON HYBRID MULTI RESOLUTION APPROXIMATION OF CAD MODELS

I.F. Ceruti[a], G. Dal Maso[b], D. Rovere[c], P. Pedrazzoli[d], C. R. Boër[e]

[a] Technology Transfer System s.r.l. - TTS
[b] [c] [d] [e] University of Applied Science of Southern Switzerland - SUPSI

ceruti@ttsnetwork.com; giovanni.dalmaso@supsi.ch; diego.rovere@supsi.ch;
paolo.pedrazzoli@supsi.ch; claudio.boer@supsi.ch

ABSTRACT

This paper presents models and ideas at the foundation of a new collision detection engine for manufacturing systems simulation. The fast and reliable collision detection method proposed is based on an hybrid approach (usage of mixed type of bounding volumes shape and positioning) capable to satisfy the requirement of a quick and consistent detection while dealing with huge 3D models.

Keywords: complex simulation environment, collision detection, real time, multi resolution approximation.

1. INTRODUCTION

The importance of computer animated simulations is ever increasing both in the design and in the production phase of manufacturing systems.

A comprehensive simulation, whether applied to a single machining centre or to an entire plant, must be capable to represent various aspects of the reality. In a modular 3D simulation library (as the one this paper refers to, described in (Pedrazzoli 06)) these aspects are taken in charge by dedicated engines.

Among those, the collision detection engine deals with the interactions between moving objects. A collision detection engine for real time simulation of manufacturing environments must satisfy the constraints of short response time, low resource usage and the capability to deal with huge simulation models. These models are mostly exported directly from the constructive drawings coming out from CAD software in order to cut the simulation set-up costs and increase the reliability of the simulation results. Thus, the resulting meshes are characterized by an high level of detail and a huge number of triangles.

This paper deals with the design and development of a fast and reliable collision detection engine based on an hybrid approach (mixed type of bounding volumes shape and positioning) capable to satisfy all these requirements.

The article is based on the following structure: after a survey on collision detection state of the art, the advantages and limitations of the different approaches adopted to meet the requirements for collision detection engine will be summarized. In the third part the selected approach at the base of the engine will be extensively explained. Section four details industrial test cases results.

2. STATE OF THE ART

2.1. Problem background

The collision detection engine addressed by this paper is conceived in order to be plugged into a framework for the 3D simulation of complex manufacturing environments, representing either a single machining centre or a complete production plant.

In the specific case, the simulation is connected to a real CNC and acts as a substitution of the real environment receiving the same inputs usually sent to the real machines and emulating their behaviour. This kind of application is mainly used in two different phases of the machine/plant lifecycle:

1. During the development of the machine/plant management logics in order to speed up the debug phase of the CNC and PLC without facing the risk to damage the real structures (mainly to avoid auto collision situations).

2. In production in order to perform an off-line check of complex part programs, thus avoiding possible collisions of the machining heads with the fixtures or with the work piece.

Therefore the collision detection performed during these simulations must be responsive still remaining as adherent as possible to the tested mechanical structures. For this reason, the developed applications based on this framework are required to directly manage the models coming from the constructive drawings.

This approach allows the reuse of machine and plants 3D models from the design phase, cutting down the simulation set up times and the related costs, thus increasing the complexity of the collision check. In fact, the great number of different existing CAD systems requires the models to be exported to a neutral mesh exchange format like VRML (Carey 97), with the result,

in the case of highly detailed models, of big non optimized meshes.

This process, along with the huge number of contiguous moving parts, could dramatically increase the response time of the collision detection engine making it unusable for real-time simulations.

During the development of the here presented collision detection engine several approaches described in literature have been used to automatically create collision structures. Their pros and cons are described in detail hereinafter. Common problems detected in all of these approaches are, on one side the under-exploitation of the knowledge deriving from the kinematic structures involved in the simulation and, on the other, the missing adaptation to the precision level required by each part of the machine/plant.

For this reason a new hybrid, semi automatically defined, multi resolution approximation of CAD models is here proposed; this is based on a mixed structure of primitives which gives the user the opportunity to take advantage of the knowledge deriving from the motion logics, the exclusion rules and the required precision levels for each of the controlled objects. A 3D application has been developed to support the user during the process of collision structures definition.

2.2. Previous work

Collision detection problems have been extensively studied in literature (Kockara 07, Lin 98). (Hubbard 93) introduces a first classification of collision detections: *broad-phase* collision detection allows detecting objects that should be subsequently tested, while and *narrow-phase* collision detection determines the exact collision. This scheme is broadly used in literature, to classify collision detection algorithms (Ericson 05).

Algorithms for the broad-phase are mainly of three kinds: exhaustive, sweep and prune and hash table based. Exhaustive approach simply leads all pairs to be tested against each other. Sweep and prune (Baraff 03, Lin 94) sorts objects through their projection onto the coordinate axes and chooses the ones which projection overlaps. Hierarchical hash tables (Mirtich 96) divides the space into cells and selects objects that overlap the same cell.

Algorithms for narrow phase collision detection can be categorized into four groups: feature-based, simplex-based, image-space based, volume-based and spatial data structures (Moore 98).

Feature-based algorithms, such as Lin-Canny (Lin 91), V-Clip (Mirtich 98), and SWIFT (Ehmann 00a, Ehmann 00b), directly works on the geometric primitives of the object, such as points, vertices and faces. Simplex-based operates on the convex hull of its points: a well known simplex-based algorithm is GJK (Bergen 03).

Image space base (ISB) techniques are computed by image-space occlusion queries which are convenient to be implemented on the graphics hardware, the GPU, and common hardware, the CPU (Benes 05). Cinder (Knott 03) is a well known example of ISB, while

frontiers on this type of algorithms can be found in (Heidelberger 04) and (Govindaraju 04).

Volume based algorithms are conceptually based on the same idea of the ISB techniques; however, they use different methods to compute layered depth images (Heidelberger 04) and distance fields. This group of algorithms are also suitable for GPU implementations. In (Gundelman 03) can be found a volume based algorithm example.

There are two types of spatial data structures for collision detection: spatial subdivision and bounding volume hierarchies (BVH). The first approach recursively divides the space, while the second recursively or iteratively partitions the object itself. With spatial partitioning, splitting of polygons (that is commonly used) causes the increasing of tree depth and the lost of performance. In addition, since cell size of the spatial partitioning cannot cover objects' primitives tightly, when objects are close, contact status determination is difficult.

On the contrary, BVHs provide smaller and tighter hierarchies, and are more suitable for general shapes than simplex based and feature based algorithms. BVH can be called as discrete representation of level of details of objects. At the first level, hierarchy includes one bounding volume which is a very coarse representation of an object. Further levels include more detailed representations of the object. The leaf level (or finest level) of the hierarchy generally includes the object primitives (lines, triangles, or tetrahedra). Bounding volume (BV) does not necessarily enclose its children's bounding volume; instead it must enclose the geometry of an object included in the children BVs.

The selection of the root volume to be descended is called *traversal rule*: generally largest volume is chosen in order to lower the chance of finding overlapping. Non-overlapping BVs are discarded from further consideration (pruning). At last in the traversal, when two leaf nodes from two distinct volumes are reached, then there are two possibilities: testing whether two primitives are colliding (pair-wise test) or testing one primitive with the other's leaf bounding volume (primitive-volume test). If objects' primitives are colliding, a pair-wise test should be performed anyway. Thus, there is a trade-off between the number of iterations and the complexity in the overlap tests.

(Gottschalk 00) states that recursively traversing BVHs is often a bad choice since the number of primitives and the hierarchies can be quite large. This problem can be solved by using iterative traversal technique with FIFO queue (Dingliana 00), with the introduction of a priority on the pair-wise BV tests.

Typical BVs are: axis-aligned bounding boxes (AABB) and spheres, usually chosen for the simplicity of such structure intersection test, OBBs (Oriented Bounding Boxes) and k-DOP (Discrete Orientation Polytopes), used for their fitting approximation of the object, and recently also cylinder (Ketchel 06) for particular geometrical bodies. In addition, with Welzl's algorithm (Fischer 03), finding better fitted spheres,

makes spheres preferable over other topological BVs. In overall, there is a trade-off between BVH complexity and performance (Gottschalk 00). Complex topology choices establish tighter fitted BVs and so fewer overlap tests but causes performance lost. On the other hand, less complex hierarchies provide faster overlap test but less tight BVs.

Recent attempts have been made to summarize different approaches in an unique algorithm. In (Yoon 04), the CHPM, Clustered Hierarchy of Progress Mesh, represents a model with a dual hierarchy: one for a coarse-grained selective refinement, the second for fine-grained local refinement. This technique also exploits the idea of error bounded detection, for approximate but fast intersection spotting.

3. ADOPTED APPROACHES

The here-described technique is the result of an evolution: different approaches have been adopted and developed to face the problem described in the previous sections. Some of them are already known in literature, while in the present work others are studied and developed to face the described problem. The aim of this section is to present an overview of these algorithms, highlighting the advantages and problems they showed when concretely used. These problems, strictly bonded to the selected application field (the real time simulation environments for big models), leaded up to the replacement of the algorithm with a more suitable one.

3.1. Literature Techniques

3.1.1. Triangles

The first idea was to use directly the triangles coming from the simulation model, surrounded by an OBB as approximation. This allowed achieving a unchallenged precision of the response without having any support structure.

Despite this, triangle technique has clearly visible problems: apart from requiring a huge number of test to be performed (each triangle have to be tested against each other triangle), the triangle-triangle test is one of the most time-consuming test, bringing the collision test far from being real time.

3.1.2. BVT (Bounding Volume Tree)

The usage of BVH was considered the solution to be adopted to speed up the collision tests. The model meshes were surrounded by binary trees of both spheres and OBBs: this approach allows performing fast rejection tests when objects were far apart, preserving the precision detail of triangles. The BVT structure is built automatically from the model mesh, assigning triangles to tree leaves, with a controlled tree depth.

Also this approach shows intrinsic lacks. The collision detection performances heavily downgrade when objects are close to each other: almost all leaves triangles need to be tested: in these cases, very frequent in considered applications, the bounding volume

structure is just an undesired overhead. An accurate analysis, as a matter of fact, showed that the efficiency of the test was strictly linked to mesh layout. Most of them generate bounding volumes trees with leaves load balancing problems: many triangles are encapsulated with few leaves or, on the other hand, many triangles are shared between leaves.

This brings redundant triangle-triangle test and pointless tree branches, with children having the same informative value of their parents. As a consequence, also the control parameter has been shown to be completely useless, because of its inability to tune efficiently the created structure. The resulting collision detection time was unpredictable, resulting, in most cases, unacceptably long.

3.1.3. Example

The following figures show an example of problems described in the previous section.
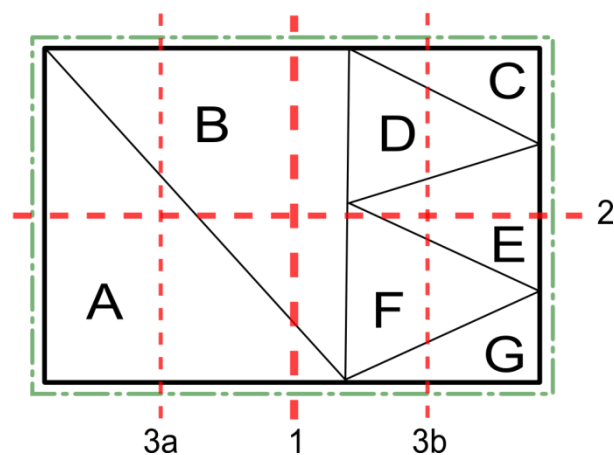

Figure 1: A possible triangle mesh, composed by triangles A B C D E F G, which BV (line-dot) is subsequently cut along the maximum extension. The second cut is common for both first level children, and have been indicated once for simplicity.

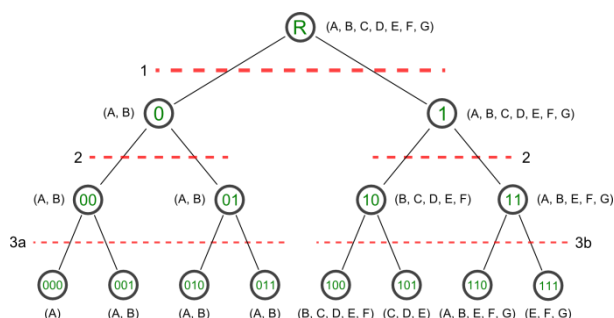Figure 2: A hierarchical structure representing the mesh in [Figure 1]. For each node the related triangles and the generating cut are indicated.

In [Figure 1] a mesh of an object has been taken into account. The OBB has been subsequently divided into children using a simple strategy: the longest extension axis is split in two parts. Moreover a tree with a depth of four is considered enough for the example,

thus three cuts have been performed (the second cut is common for both first level children).

The hierarchy shown in [Figure 2] contains examples of all problems affecting hierarchical structures built on mesh, as for example BVT. The resulting structure is an heavily unbalanced tree: the right branch (1) of the root node (R), as a matter of fact, contains much more primitives for node than the left branch (0). Moreover, considering the leaves of this tree, many primitives are shared between nodes (e.g. B for all nodes except 000, 101, and 111). Node 01 and its children are a typical case of useless branch: it replicates the informative content of its parent referring triangles A and B.

3.2. Developed Techniques

3.2.1. BVMix (Bounding Volume Mix)
As a possible solution to trees with uselessly long branches, a fixed hierarchy of different bounding volumes has been adopted, also including space partition concepts.

This technique, titled BVMix, gives a multi resolution approximation of the given mesh. The model is represented, as a first approximation, by a sphere and by its OBB: this couple of bounding volumes assures a fast rejection when objects are far apart. The OBB is then subdivided in twenty-seven children called *clusters* obtained cutting the main OBB homogeneously on each of its axes. Each cluster contains a grid of spheres (*leaf spheres*) obtained by subdivision, in a tuneable number, of the shortest side: each of these spheres is associated with the triangles of the mesh concerning to the sphere volume. At the end each empty volume, both cluster and sphere, is removed from the hierarchy.

BVMix, preserving the triangle precision and the BVT hierarchical idea and fast rejection idea, is characterised by a control parameter (number of cluster side subdivision) able to fit the complexity of the model. This algorithm gave also the possibility to approximate the test, ending it at the sphere level: a fast response, especially in machine debug preliminary phases, is in most cases a desired feature for a simulation collision detection algorithm. Despite the new approximation capabilities, and the performances improvement in general cases, compared to other techniques adopted, also BVMix suffers some problems.

As for BVT, also for BVMix efficiency depends on mesh layout because triangles are still the final actors of the collision test, and the problem of triangles shared between leaves (spheres) is still unresolved. Last but not least, also in case of usage for approximated collision test, the user has no control on the approximation generated by the grid of spheres, not even tuning properly the control parameter.

3.2.2. Fitting Boxes Algorithm
Analysing the up-to-now described approaches, it's noticeable that problems for achieving goals of collision detection engine, with simulations mesh model, comes from the model mesh layout. Triangles, apart from being the most precise way for detecting collision, are the real bottleneck of every algorithm for collision detection. Thus the hereinafter described approach is based on the idea to sacrifice the precision of triangles in favour of a good approximation able to give a fast and conservative response to collision test.

The last approach described in this section is based on an approximation of each model with a two level hierarchy of boxes: we called it the Fitting Boxes Algorithm. The first level of FBA hierarchy is simply the model OBB, called *master box*. The second level is a set of OBB (aligned with the master box) called sub boxes: each sub box is obtained by refinement of the master box or of another sub box.

The refinement operation substitutes a box with a set of boxes sharing orientation, and whose dimension is a percentage of the generating box, containing at least one triangle of the model. The FBA building process can be made both by the user (who indicates which box has to be refined and the refinement percentages) or automatically (just indicating a percentage the algorithm refines the master box): this process leads to a controllable resolution of the model.

This structure is then used from the collision detection engines for the test phase: excluding the triangle from test phase considerably enhance test response time in each respective situation of objects involved. Moreover FBA evidently did not suffer of problems such as load balancing of the hierarchy (also because it exploit a two level hierarchy), shared triangles, or dependence of its efficiency from the level of detail of the mesh of the model involved in test.

Notwithstanding his effective features, that made FBA more usable simulations than previous approaches performed, also FBA has a main disadvantage: automatic built structures does not always perform a good approximation of the starting model. Also "hand-made" structure approximation can be either poor or needs a great number of boxes to reach a good approximation: curvatures on object are a clear example of this lack, and are intrinsic in the choice of box as approximating structure.

4. HMR APPROACH
None of our previous approaches took into account the peculiarities of the models used in the simulation environment. Machine CAD drawings are composed of a great number of shapes, whose geometry is different the ones available in classic 3D applications (e.g. trees, people, animals, vehicles). The presence of shapes with a lot of rough edges represent a problem when trying to apply automatic mesh reduction algorithms like edge collapse (Cignoni 00).

The approach we are describing exploits a key factor in simulation environments: the presence of physical laws that constraints the movements of the parts composing the machine model. The knowledge of the kinematics allows defining the degrees of freedom

of each part. For example, the limitation of one part movement can easily determine its bounding box in each allowed position. This can reduce the complexity of the approximating model (at least in a coarse-grained level of detail).

A very important aspect in machine simulation is the possibility to define a tolerance in the collision detection precision: some critical parts don't need to get too much close to each other for safety reasons (e.g. to avoid contacts for vibrating parts): in this case the simulation must report a collision even if the model meshes do not geometrically intersect.

4.1. Description

The exploitation of the simulation model knowledge and the inability to use automatic mesh reduction algorithms lead to an approach that defines and exploits a hierarchical multi resolution (HMR) approximating description of the models. This section describes how to define the approximating volumes, the hierarchical relationships and, for moving parts, the association with their degrees of freedom.

An object occupied volume can be defined using a set of geometric primitives optimized for intersection test: spheres, boxes and cylinders. These primitives are well suited for approximating most of the shapes found in machine CAD models. For example a screw can be approximated with two cylinders (one for the head and one for the shaft) or with a single cylinder, depending on the precision required for that part.

Models could eventually contain part that, even if made by simple shapes, cannot be well described with the available primitives. For this reason the here-described approach also includes the possibility to define a triangle mesh starting from the object convex hull. Given that triangle intersection tests are the most expensive, the number of approximating triangles mesh has to be low. The impact of the presence of triangle meshes can moreover be limited using an enclosing simple volume, like a sphere or a box, for early exclusion.

The advantages of hierarchical techniques can be achieved, in this approach, through the organization of these geometries in trees. This structure is traversed with the classical collision detection mode: test fails when a child fails, while succeeds when a leaf intersects.

The key differentiating point, with respect to traditional approaches, is the possibility of mixing different types of geometries allowing choosing the best fitting shapes for each node composing the tree. It is also possible to balance the depth of the tree depending of the required precision weighted against the complexity of the model.

The position and orientation of the geometry of each node is associated with that of a node of the simulation environment scene graph. This association is a key point in the flexibility of the tree structure: a node can be associated with a different transformation from the one linked to his parent or children one. This allows the independent motion of each tree node.

The resulting structure can be seen as a scene graph used for collision detection rather than for rendering. The fact that this structure is explicitly built gives the possibility to access any node and modify its properties. For example the enclosing bounding volume of a multiple drill head can be adjusted depending of the status of the drill tools: when the simulation control logic receives the command to extend or retract the tools, it can also expand or shrink the enclosing bounding volume to take into account tools positions.

Moreover, with this flexible hierarchy is also possible to realize culling and Boolean operations on geometries: a parent-child relation defines, as a matter of fact, the volume resulting from the difference between their approximating volumes. A scene graph like structure allows verifying and fine tuning the resolution of the approximation: where a greater detail is needed the structure can be expanded, leaving the representation coarse-grained in the zone of low interest. In this way it is possible to balance the precision and the speed of the collision detection.

Finally, a great advantage of this structure is the low memory footprint, as each node is the best fitting shape and the depth and the trees don't exceed the necessary.

4.2. Collision structure editor

A collision structure editor has been developed to support the user during the definition of the collision structures. In this 3D application it is possible to load models from VRML files: keeping the hierarchical structure of shapes of this format, it is possible to easily indentify and select model parts. Once the user has selected one or more parts, he can create one of the available bounding volumes (box, sphere and cylinder). The user can also define a triangle mesh as bounding volume. The application computes the best fitting bounding volume for the selection and, in the case of the triangle mesh, creates its convex hull.
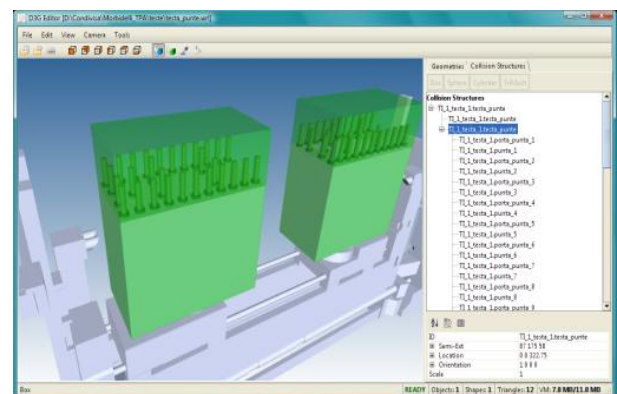

Figure 3: Collision structure editor

The bounding volumes created are the nodes of HMR hierarchy. These nodes can be arranged in the

desired structure using drag and drop operations on the tree view on the right side of the application.

The user can manually edit the properties of the primitives, such as extensions, location and orientation, to adjust the automatically computed values. That means that it is thus possible to define a bounding volume bigger that the fitting one: this is particularly useful when tolerances on movements are to be considered. Objects vibrations and dilatations are only simple examples of requirements bringing to slight enlargement of automatic fitting bounding volumes.

5. TEST CASES

In this section the results obtained by the described algorithms in two different applicative scenarios are reported. In both cases simulations were made on a common desktop PC (Pentium 4 3GHz, 2GB RAM, NVIDIA GeForce 6800 128MB video card) under a Windows XP SP3 operating system. The collision detection engine and algorithms are included, as pluggable modules, in a Java 1.6 simulation environment: thus all algorithms are implemented in Java. Performances measures, reported below, refers to this hardware and software configuration.

5.1. Machine scenario

The described test case concerns a real industrial scenario of a configurable hole punching machine.

5.1.1. Scenario description

This kind of tool-machine actually has up to 24 heads (12 for each of the two sides of the wood panel) and each head has up to 29 tools. The machine also has 5 bridges and 5 supporting structures to keep the panel in place, and 2 machine moving structures. The heads are mounted in pair on an axis that can move along the **x** direction, while each head can move along the **y** and **z** directions and rotate ±90 degrees around its axis. On the head each tool can be extended or retracted: thus the overall shape of the moving head is complex and dynamic: a schema of the machine is reported in [Figure 4].
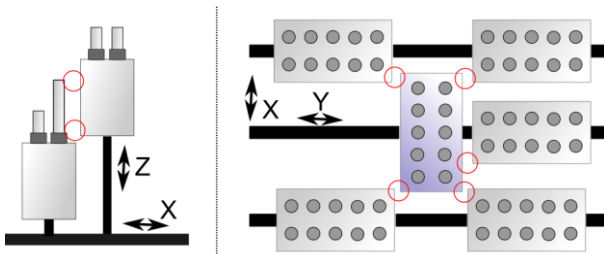

Figure 4: Test case machine heads schema. The red circles highlight possible collision in moving objects in the scene.

This scenario is particularly challenging for each collision algorithm because it has 732 different objects to be checked (267546 combinations of object to be tested), for a total of about 200 thousands triangles.

This amount of colliding pairs can be divided in collision groups (groups with objects worth to be checked for collision) to speed-up the overall test. For example the tools of a head cannot enter in contact with each other as they are firmly mounted on the head, and can thus belong to different collision groups.

A single head, as the blue one in [Figure 4], can collide with each of the heads of the other side of the panel, with the head on the same **y** axis and with the adjacent heads on the **x** axis. Tools can enter in contact with the heads and tools of the other side and with the adjacent tools and heads on their side, bringing the same number of objects to be checked as for the head. Moreover, the first and last heads have only one adjacent head on the **x** axis. The number of pairs of objects to be tested under these hypotheses is around 152 thousands.

Using the previously described method, the knowledge that heads and tools are moving together can be exploited. HMR allows seeing tools as children of a common parent bounding volume: these and the head are thus children of a common root bounding volume. This root can collide with the five adjacent ones and with the ones acting on the other side of the panel. Thus the number of objects to be checked in the first level of detection is around 400: only if collision is detected at first level HMR further checks each child of one hierarchy against each one of the other. Considering that most of the time we don't have collisions don't occur (as the CNC is able to generate safe paths in most of the cases), having a low first level number of objects, can strongly reduce the total number of checks.

The described machine, and its amount of colliding objects, has been used as reference environment for the performed test campaign and as benchmark for all described algorithms. Moreover, a common simulation has been set up: all the machine axes have been moved with a triangular speed motion law, bringing all the objects involved to collide at least once with the other movable objects.

5.1.2. Results

Benchmark simulation is formed of many steps including the update of the environment after the axis motion and a collision detection check. Number of primitive tests and time spent while testing are collected for each algorithm: results are in [Table 1].

Table 1: Results of tested methods for hole punching machine test case: triangles method results are omitted since not significant.

	BVT	BVMix	FBA	HMR
Box - Box	16565K	30707K	18654K	1675K
Sph - Sph	/	2115M	/	/
Tri - Tri	5115M	9279M	/	/
Collision	34099	34099	42062	32185
Time (s)	852,26	1673,69	5,81	4,47
FPS	0,12	0,06	17,21	22,39

Results obtained with Triangles algorithm are here omitted because not significant: time spent for collision isn't comparable with any other method tested.

Results show, as previously stated, that HMR including knowledge on machine in its collision structure is capable to perform better than other methods for collision detection. In particular, its frame rate makes HMR an *interactive* method for collision detection. It is worth notice that this result it is not reached with exact algorithms (Triangles, BVT or BVMix).

As a matter of fact BVMix, in this test case, performs worse than BVT: this is understandable considering that for the rough-edge geometries used BVT (using boxes as base bounding volumes) is capable to better approximate geometries than BVMix (that has spheres as hierarchy leaves). BVMix thus being more suitable for general geometries, as already explained, demotes its performance with squared bodies.

It is moreover useful to notice the number of detected collisions: HMR is supposed to detect more collisions, because it doesn't use the exactness of model triangles.

Grouping tools under a common bounding volume allows HMR to handle the case that when tools are retracted inside the head body, they cannot collide with anything outside the heads (parent bounding volume is sized to include only tool outside volume). For other methods, instead, they always collide with anything they are checked with. This, once again, makes HMR faster thanks to included machine knowledge and explaining the difference in the number of detected collisions.

5.2. Robot arms scenario

This test case concerns a common industrial application: the usage of robot arms acting simultaneously on the environment, for example in an assembly cell like the one in the simulation pictured in [Figure 5].
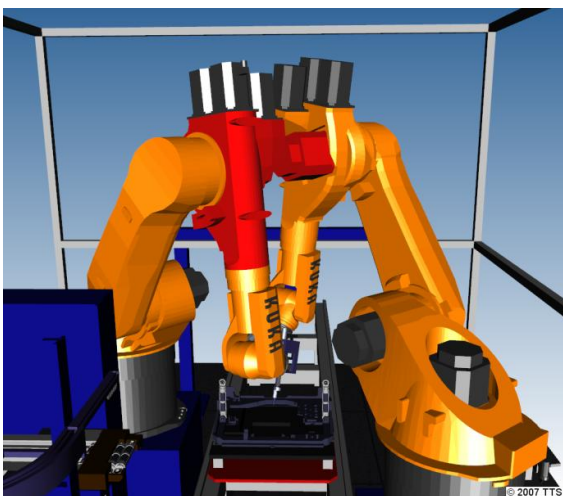

Figure 5: Robot arm scenario. Two robotic arms (KUKA KR 30 HA) colliding during an assembly task.

5.2.1. Scenario description

In this test case two robot arms are assembling a piece, acting on the two side of a conveyor system.

While assembling, it is possible that any of the links of each arm collide both with another of the same arm (*self-collision*) and with links of the other arm. A first kind of collisions is usually avoided thanks to physical joint limits or through properly writing each arm task; the other kind is usually managed in the simulation phase, especially when the number of arms involved increases.

This robotic scenario, even if common in industrial applications, is not particularly challenging for collision detection algorithms: each arm is formed by 6 links and 14 pieces to be checked for collision for a total of about 34 thousands triangles.

5.2.2. Results

Each simulation step requires the motion of all the joints in a task, forcing each arm link to collide with each other many times. Also for this test case, results obtained with Triangles algorithm are omitted because not significant.

Table 2: Results of tested methods for robot arms test case. Triangles method results are omitted since not significant.

	BVT	*BVMix*	*FBA*	**HMR**
Box - Box	307K	3026K	766K	197K
Sph - Sph	/	22M	/	/
Tri - Tri	954M	178M	/	/
Collision	6284	6284	7328	8060
Time (s)	160,501	33,451	0,222	0,109
FPS	6,23	29,89	4,5K	9,2K

This tests case, besides confirming what previously stated, shows also how BVT performances heavily can downgrade depending on the model, also if compared with another exact algorithm as BVMix.

The arm model is composed by many small triangles coupled with some long ones: when a BVT leaf is selected for primitive test, it participates with many elements, many of which are useless for detecting collisions. BVMix leaf spheres generally enclose much less primitives to be tested only when needed: being small are more punctual bounding volumes.

6. CONCLUSION

In this paper a hierarchical fast and reliable collision detection method based on an hybrid approach is presented. This novel technique, called HMR, is capable both to take into account the geometry structure of a model (using different kinds of bounding volumes), and to include knowledge on the model kinematics peculiarities and physical constraints.

The capability to include model knowledge, rising from the organization of bounding volumes in

hierarchies, allows HMR to achieve interactive collision detection with a low memory usage.

The stratification of bounding volumes represents for HMR a smart multi resolution approximation of moving objects: this representation comes directly from the human experience. Objects representation and organization in parent-child hierarchies is created by the user with the help of a visual editor.

Further investigations will be directed on exploiting HMR abilities to use more knowledge, e.g. reusing knowledge related to a single part of the model, to speed up the collision detection on other model parts. Another possible improvement will allow HMR to deal with other bounding volumes types.

REFERENCES

Baraff, D., Witkin, A., Anderson, J., Kass, M., 2003. Physical based modelling. *SIGGRAPH Course Notes*.

Benes, B., Villanueva, N. G., 2005. GI_COLLIDE-Collision Detection with Geometry Images. *Proceedings of the Spring Conference on Computer Graphics*, 95-102

Bergen, G., 2003. Collision Detection. *Interactive 3D Environments. Interactive 3D Technology Series*, Morgan Kaufmann ed.

Carey, R. & Gavin, B., 1997. *The Annotated VRML 2.0 Reference Manual.* Addison-Wesley

Cignoni, P., Costanza, D., Montani, C., Rocchini, C., R. Scopigno, 2000. Simplification of Tetrahedral Meshes with Accurate Error Evaluation. *Proceedings of the conference on Visualization,* 85-92

Dingliana, J., O'Sullivan, C., 2000. Graceful Degradation of Collision Handling in Physically Based Animation. Proceedings Eurographics Computer Graphics Forum, 19: 239-247

Ehmann, S. A., Lin, M., 2000. Accelerated Proximity Queries between Convex Polyhedra by Multi-level Voronoi Marching. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3 : 2101-2106

Ehmann, S. A., Lin, M., 2000. *Swift: Accelerated Proximity Queries between Convex Polyhedra by Multi-level Voronoi Marching*. Tech report. Department of Computer Science, University of North Carolina at Chapel Hill.

Ericson, C., 2005. *Real-Time Collision Detection.* Morgan Kaufman.

Fischer, K., Gartner, B., 2003. The Smallest Enclosing Ball of Balls: Combinatorial Structure and Algorithms. *Proceedings of 19th Annual Symposium on Computational Geometry (SCG)*, 291-301

Gottschalk, S., 2000. *Collision Queries using Oriented Bounding Boxes*. PhD Thesis. Department of Computer Science, University of North Carolina.

Govindaraju, N., Lin, M., Manocha:, D., 2004. *Fast and Reliable Collision Detection Using Graphics Hardware*. Tech report. Department of Computer Science, University of North Carolina at Chapel Hill.

Gundelman, E., Bridson, R., Fedkiw, R, 2003. Nonconvex Rigid Bodies with Stacking. *Proceedings of ACM SIGGRAPH*

Heidelberger, B., Teschner, M., Gross, M., 2004. Detection of collisions and self-collisions using image-space techniques. *Journal of WSCG*, 12 : 1-3

Hubbard, P. M., 1993. Interactive Collision Detection. *Proceedings of IEEE Symposium on Research Frontiers in Virtual Reality*, 24-31.

Ketchel, J. S., Larochelle, P. M., 2006. Collision Detection of Cylindrical Rigid Bodies for Motion Planning. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, 1530-1535.

Knott, D., Pai, D., 2003. Cinder: Collision and Interference Detection in Real-Time Using Graphics Hardware. *Proceeding of Graphics Interface*, 73-80.

Kockara, S., Halic, T., Iqbal, K., Bayrak, C., Rowe, R., 2007. Collision detection: a survey. *IEEE International Conference on Systems, Man and Cybernetics*, 4046-4051

Lin, M., Canny, J. A, 1991. Fast algorithm for incremental distance calculation. *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*

Lin, M., Gottschalk, S., 1998. Collision Detection between Geometric Models: A Survey. *Proceedings of IMA Conference on Mathematics of Surfaces*, 37-56

Lin, M., Manocha, D., 1994. *Efficient Contact Determination between Geometric Models*. Tech report. Department of Computer Science, University of North Carolina at Chapel Hill.

Mirtich, B., 1996. *Impulse Based Dynamic Simulation of Rigid Body Systems*. PhD Thesis. University of California, Berkley

Mirtich, B., 1998. V-Clip: Fast and Robust Polyhedral Collision Detection. *ACM Transactions on Graphics*, 17 : 177-208

Moore, M., Williams, J., 1998. Collision Detection and Response for Computer Animation. *Computer Graphics*, 22 : 289-298

Pedrazzoli, P., Sacco, M.; Jönsson, A., Boër, C.R., 2006. Virtual Factory Framework: key enabler for future manufacturing. *Digital enterprise technology: perspectives and future challenges*, Springer, 1, 83-90

Yoon, S., Salomon, B., Lin, M., Manocha, D., 2004. Fast collision detection between massive models using dynamic simplification. *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, 71 : 136 -146

OPTIMIZING A HIGHLY FLEXIBLE SHOE PRODUCTION PLANT USING SIMULATION

F.A. Voorhorst [a], A. Avai [b], C. R. Boër[c]

[a]HUGO BOSS Ticino SA, Coldrerio, Switzerland
[b]Technology Transfer System, Milano, Italy
[c] CIM Institute for Sustainable Innovation, Lugano, Switzerland

[a]Fred_Voorhorst@hugoboss.com, [b]Antonio.Avai@ttsnetwork.com, [c]Claudio.Boër@icimsi.ch

ABSTRACT
This paper explores the use of simulation for the optimization of highly flexible production plants. Basis for this work is a model of a real shoe production plant that produces up to 13 different styles concurrently, resulting in maximum 11 different production sequences. The flexibility of the plant is ensured by organizing the process in a sequence of so-called work islands, using trolleys to move shoes between them. Depending on production needs one third of the operators are reallocated. The model considers the full complexity of allocation rules, assembly flows and production mix. Analyses were performed by running use cases, from very simple (providing an insight in basic dynamics) up to complex (supporting the identification of interaction effects and validation against reality). Analysis gave insight in bottlenecks and dependencies between parameters. Experiences gained distilled in guidelines on how simulation can support the improvement of highly flexibly organized production plants.

Keywords: shoe plant simulation, production mix, labour allocation

1. INTRODUCTION
Discrete event simulation has been widely used to model production line (Roser et al. 2003) and to analyse its overall performances as well as its behaviour (Boër et all. 1993). For the most part, past models have concentrated on the mechanical aspects of assembly line design and largely ignored the human or operator component. (Baines et all. 2003). The simulation model, presented in this paper, was developed in Arena (Kelton et all. 2003) and it augments the standard production system model to include labour movements and its dynamic allocation many times per shift.

This paper describes the experiences and findings in using discrete event simulation as tool to better understand a plants dynamic behaviour prior to optimization and further improvements

The remainder of this paper is organized as follows: in section 2 a short description of the problem is presented and section 3 gives an overview about the actual system to produce men shoes. Section 4 provides a description of all the modelling and implementation issues to be faced in order to get a simulation model with a correct detail level. In section 5 the results are presented and conclusions follow.

2. PROBLEM DESCRIPTION
The challenge we face is to better understand the dynamic behaviour of the shoe production plant in order to be able to predict the daily volume and as basis for improvements to obtain a more fluent production. Actually there are many factors influencing these aspects, such as labour availability and allocation of operators, availability of lasts and, clearly, the composition of the daily production plan, the so-called production mix. The production process has almost 40 different operations, grouped in work islands, to which approximately 70 operators are allocated. The production plant can work on more then 100 shoe variants, each one different in production routing and/or cycle times for operations.

The main goal of this project is to identify the scenarios under which the system breaks down (production target is not achieved) in order to evaluate the impact of key factors such as production mix and labour allocation on the overall performances. The theoretical target productivity is about 1.700 pairs of shoes per day. However in the real system, daily through-put is not constant and shows large variations, sometimes 25% below target value.

3. SYSTEM DESCRIPTION
The actual production plant assembles high quality man shoes in various colours, mainly of 3 different families:
1. Shoes with glued leather sole
2. Shoes with stitched leather sole
3. Shoes with rubber sole

From the 3 families the production processes of 50 shoe styles were modelled, amounting in 11 different process sequences (differences due to colour are not included).

The organization in work islands makes the production a very flexible system, both in terms of product types and capacity allowing the possibility to

maximize through-put while minimizing investment such as the total number of lasts per style needed. At any point in time there are up to 13 different styles in production, each needing a specific last model with a significantly different (between families) or a slightly different (within one family) production sequence. In addition, shoes of the same style can have different colours, such as black, ebony, brown, grey, white, etc. which have an additional impact on the production sequence.

The production plant, organized in a circular fashion, is split in 2 main departments:

- The assembly department where shoes are assembled by means of last starting from upper, sole and insole, as it is displayed in Figure 1.
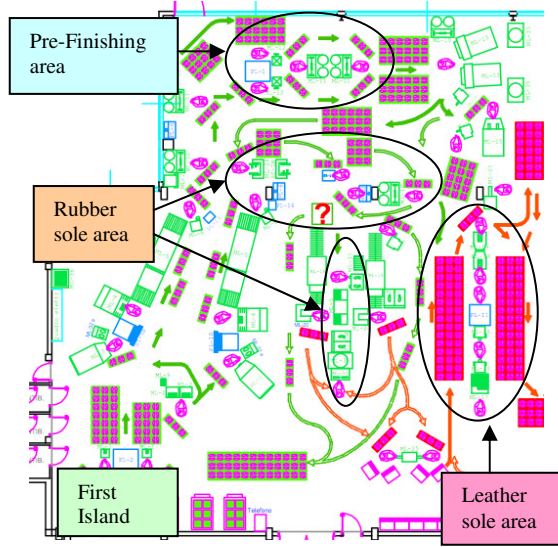

Figure 1: Layout of assembly department

- The finishing department, see Figure 2, where shoes are creamed, brushed, finished and packaged.
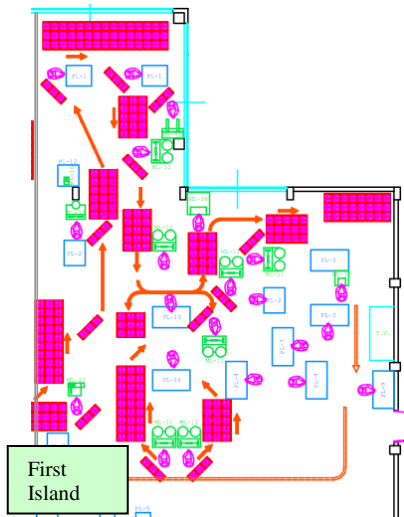

Figure 2: Layout of finishing department

Each department is organized in different working islands, by grouping one or more machines and working positions. Furthermore, as shown in Figure 1, in the assembly department, 3 macro areas, to which a single team is assigned, can be identified:

1. The pre-finishing area, composed by 3 islands, where the leather upper can be aged, daubed of cream and brushed.
2. The rubber sole area, formed by 7 islands where rubber soles are glued and coupled with shoes.
3. The leather sole area, composed by 4 islands where shoes, with leather soles, are stitched.

The rubber and leather sole areas are crossed only by some shoe articles, so workers are allocated only when some trolleys are waiting to be worked.

Shoes move from one island to the other by means of trolleys, moved by workers. In general an operator takes a waiting trolley, performs an operation to each shoe on the trolley and pushes the processed trolley to the waiting area for the next island. There are 2 trolley types:

- Assembly trolley: each one holds uppers, with the respective lasts, soles and insoles. They are used only in the assembly department.
- Finishing trolley: it transports shoes trough the finishing department.

The number of assembly and finishing trolleys is limited in order to keep constant the flow of shoes but, on the other hand, it can have a negative impact on the through-put. If many trolleys are stacked up in different positions, there are none available to be loaded with new shoes. Better production fluency is achieved when the length of trolley queues are minimal.

4. MODELLING ISSUE

This paragraph describes the simulation architecture as well as all the relevant aspects analyzed during the modelling and simulation model deployment phases.

The applied methodology follows a top down approach: first, the flow of shoes in the production plant has been simulated, and it has been refined adding details and rules by means of several meetings and interviews with foreman and production responsible. Then, the rules dealing with the production batches composition and dispatching have been modelled and tested. Last, the dynamic behaviour of labour allocation between different islands and inside the 3 macro areas has been simulated. It was assumed that operators have equal skills and are interchangeable. Furthermore, an extensive campaign to measure cycle times by direct observations was carried on.

4.1. Simulation architecture and input data analysis

The simulation model is driven by 3 Excel files with the following input data:

234

1. The production mix, in terms of shoe articles, quantity and colour to be produced
2. The assembly sequence per style along with stochastic cycle times for each operation
3. Several parameters related to the process together with the distances between islands

All these data are automatically imported in the simulation model at the beginning of each run.

For the stochastic cycle times a triangular distribution was used. (Chung C. A. 2004)

4.2. Simulation of shoes flow

A particular attention was kept to simulate the following issues that are described in the next 2 paragraphs:

- The input buffer policy in each island
- The trolley selection and dispatching rules at the roughing island

4.2.1. Input buffer policy

Every island has an input buffer where trolleys are stacked up if they cannot be processed immediately. These buffers are simulated as queues following the same policy except for the last removing island.

The defined policy for a queue is as following: each coming trolley is ranked based on its order number and, then, it is released following the FIFO rule (first in – first out) when the machine is free. In this way, each island tries working together all trolleys with the same order number.

At the last removing island, lasts are taken out from the shoe and put back into baskets. To minimize the number of baskets being filled in parallel, the last removing island does not follow the FIFO rule. Instead, trolleys are worked on last codes. This ensures a minimal change in baskets as large numbers of the same last are processed in one batch.

4.2.2. Trolley selection at roughing island

All worked shoes have to be roughed in the roughing island then they pass trough a reactivation oven where the cement is reactivated and, eventually, sole is applied to shoe bottom and pressed. There are 2 reactivation ovens for shoes with leather soles and one for rubber soles. In order to reach the productivity target and to keep the number of workers involved in the mentioned processes as small as possible, the worker at roughing island follows some rules of thumb to decide which trolley to take out from his/her queue, work it and move to the right reactivation oven.

The main issue in the modelling phase was just to understand the basic lines followed in this decision process and then to clearly define the several rules of thumb.

By means of direct observations and interviews with foreman and workers staffing the roughing island as well as reactivation ovens, it was found out the second reactivation oven for leather sole is switched on when

- The amount of stacked up trolleys at first oven for reactivating leather sole is greater then a certain threshold
- The oven for activating rubber sole is switched off.

Once it's switched on, it should work for about an hour and then it is switched off again.

Generally, more then 10 trolleys with different shoe articles are staked up at roughing island. Many times during a shift, the worker in this island has to decide when the second oven for leather sole has to be switched on, which and how many trolleys sent to it, or, vice versa, when the oven for rubber sole has to be activated.

The selection process is triggered by 2 events:

1. If some trolleys, holding shoes with rubber sole, are waiting at the roughing machine, they will be worked if the queue at oven for rubber sole is very short. This kind of process goes on until the queue at first oven for leather sole is long enough to avoid its stopping.
2. If no trolleys, holding rubber sole, are waiting and the queue at first oven for leather sole is too long then the selection process is a little bit complex. The basic idea is to try to work at roughing machine a certain amount of trolleys holding the same last in order to reduce the number of set up at roughing machine and to keep on the second oven for leather sole for an hour, at least. This area could become a candidate to be investigated by means of simulation to improve system performances.

Furthermore, when too many trolleys are staked up at this island, another manual roughing machine is activated for about an hour staffed by an operator to reduce the length queue of waiting trolleys.

4.3. Production batches composition

A *production batch* represents a single lot put in production at the same time in order to use the available lasts efficiently. It can be composed by one or several orders of different shoes but using the same last code to be produced. The *batch size* represents the amount of lasts used for each production batch.

At the very beginning of the simulation, the whole production plan is examined to aggregate sequential items with the same last code and to disaggregate items with ordered quantities greater than the number of available lasts. In the first case, the aggregation mechanism is mainly based on *homogeneous batch concept*: the basic idea is to create batches, using the same last code, with a similar size. In the latter case, orders with big quantity are split based on

- Available last
- Homogeneous batch as mentioned before.

A split order is put in production again when, at least, there is a certain percentage of available last in the stock again compared to the batch size.

4.4. Simulation of dynamic labour reallocation

Workers are re-allocated many times during a shift mainly because:

- The amount of available labour is less than the actual working positions
- Some shoe articles have long cycle times for some operations/islands and the number of workers allocated to these islands have to be increased to avoid queues

The decision on how to allocate labour takes into account many factors such as:

- Batch size
- Assembly sequence and cycle times
- Work already in process
- Last availability
- Labour availability
- Skill of each worker

By changing the schedule it is possible to influence the labour need. In the real system, the production responsible can modify the schedule based on the actual situation in production. This is done in order to increase flexibility in labour management, and to avoid trolleys being stacked up in front of some islands. This supervisory behaviour is discarded as it is beyond the scope of this project and the simulation strictly follows the schedule.

The first step to simulate the dynamic labour reallocation was to understand the general principles and rules applied by the production responsible and model them in a formal way. In particular, the following items were defined:

- The decision events: when decisions on labour reallocation have to be taken
- The worker allocation or de-allocation rules for each decision moment

In general, labour allocation rules can be applied during these four specific decision moments:

1. When a new item arrives to an island with no worker available
2. When a queue of an island is getting too long
3. When an island has no item to be worked
4. When a worker has completed a certain number of trolleys

In the first two moments, an available worker has to be moved to the needed island, in the third case, an operator becomes available to be moved and in the last case a worker is eligible for transferring.

4.5. Labour allocation modelling

About 65% of available workers have a fixed position. In both assembly and finishing some work islands are continuously staffed whereas others are not. The remaining flexible workers are assigned depending on the production needs. In the simulation this is modelled by grouping the flexible workers in a single pool, and allocating them according to rules reacting to the first or second event, as mentioned before.

An operator, if available, is taken from the pool immediately when an island 'requests' an operator, for example when the number of waiting trolleys exceeds a specific amount. When there are no workers available in the pool, 2 different situations have been simulated:

1. If the requiring island belongs to a macro area, as mentioned in the paragraph 3, an operator, working in the same area of the empty island, can be shared: he/she can work in 2 different positions alternatively.
2. If the requiring island does not belong to a macro area, it has to make a "*reservation*". This mechanism will be described in the next paragraph.

4.5.1. Reservation mechanism

The reservation mechanism simulates the request of labour dynamic reallocation when all available workers are busy and some trolleys are waiting for being worked in, at least, one island. In the actual system this mechanism represents the moment when some trolleys reach an empty island and the foreman has to wait until, at least, a worker can be moved.

A reservation is triggered when some trolleys are staked up in an empty island and when no workers can be moved to this position. This situation can become critical because many trolleys could pile up. In order to avoid this scenario, a worker has to start working in this empty island as soon as possible.

When a reservation is made, the first worker becoming available (either free or candidate for transferring) is reallocated. The simulation model calculates the travelling time based on the starting and arrival positions.

5. SIMULATION RESULTS AND PERFORMANCES EVALUATION

After having concluded the validation, the simulation model was ready to make several runs and analyse of production performance and through-put under different conditions, aiming to identify bottlenecks and main important process drivers. To help the validation and analysis an animation was provided as shown in Figure 3.

The simulation was tested against different production mixes. Production mixes defines the combination of shoe families produced and for each family the quantities (batch sizes) produced. Both the combinations of families as well as the batch sizes were systematically changed.

The following variables were measured:

- The overall performances, mainly daily through-put
- The labour utilization
- The production fluency indicated by the staking trolley in some key islands.
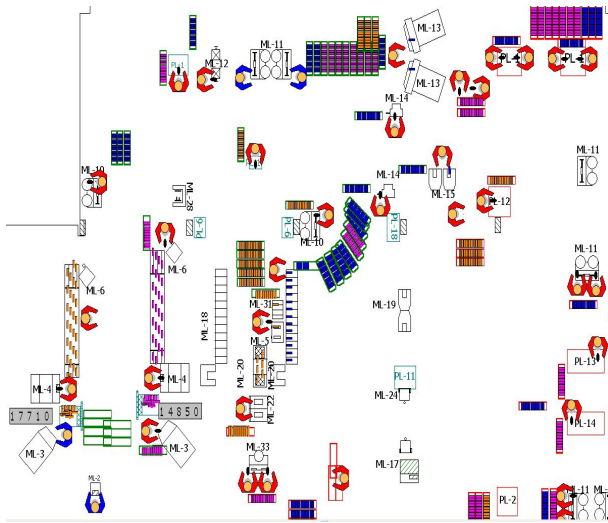
Figure 3 Screenshot of simulation model animation

To obtain a good understanding of the production dynamics the analysis was based on use-cases of different complexity, first simulating simple production plan composed by only one shoe family and then adding the other two families changing the mix and the batch size, and finally using production mixes composed of three types of shoes. Furthermore, first the performances of the two departments were assessed separately and then the whole production system was analyzed. In addition, a specific analysis was carried out to investigate some input parameters dealing with labour management.

5.1. Use-case one for assembly area: producing only one family of shoes

In this first use-case the production mix is composed only by a single shoe family to identify the family specific bottlenecks. Figure 4 shows an example of the trough-put for the different shoe families in the assembly area. Such a results such demonstrate the large difference in produced quantities depends on the shoe family. Similar differences were found for resource allocation and production fluency. As expected, through-put is determined by the produced shoe family and not influenced by the batch size.
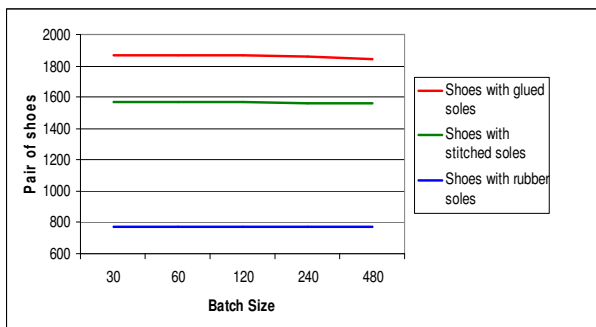

Figure 4 Daily through-put vs batch size for each shoe family in the assembly area

5.2. Use-case two: producing two shoes families

In the second use-case the production mix is composed of two shoe types to identify main interaction effects between shoe families. Figure 5 and 6 show an example of through-put when combining two shoe families in the assembly area. Results show how produced quantities are impacted by the production mix i.e. the combination of families/styles produced, and not influenced by the batch sizes. As expected, through-put is determined by the production-mix.
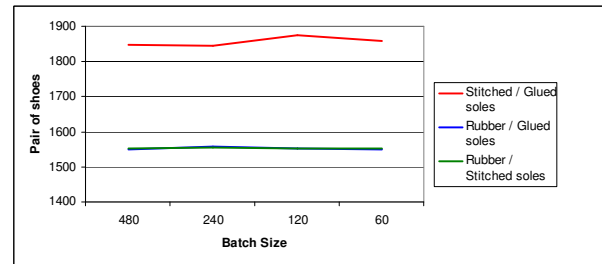

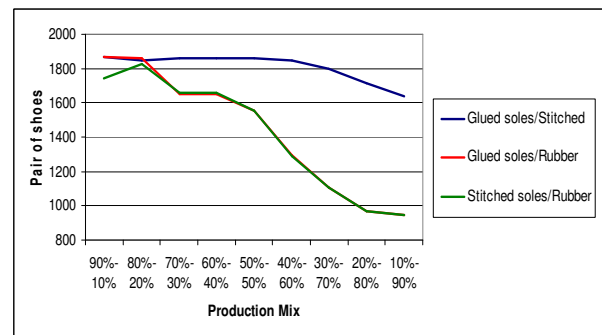Figure 5: Through-put vs batch sizes when combining two shoe families

Figure 6 Through-put vs production mix when combining two shoe families

Similar differences were found for resource saturation (see Figure 7) and production fluency.
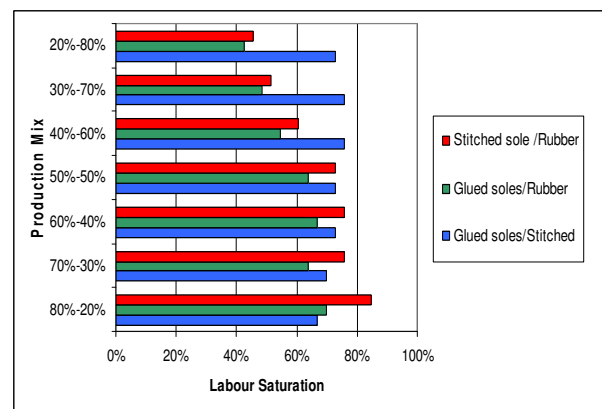

Figure 7 Labour saturation VS production mix

237

5.3. Use-case three for assembly area: producing three shoes families

In the third use-case the production mix is composed of three shoe types. To limit the number of simulation runs the production mixes followed the strategy used in production. Typically, half of the production capacity is assigned to one shoe family while the second half is shared by the remaining families.

Looking at Figure 8, the productivity is influenced minimally by the batch size if between 120 and 240 and if the ratio of stitched/rubber shoe families is between 1 and 2. Rubber shoes have a significant impact on productivity that is lower about 5% then the target if the daily percentage of produced shoes with rubber soles is bigger then 30%. Currently, the annual demand for rubber sole is close to 20-25%, although demand changes with every year and/or season.
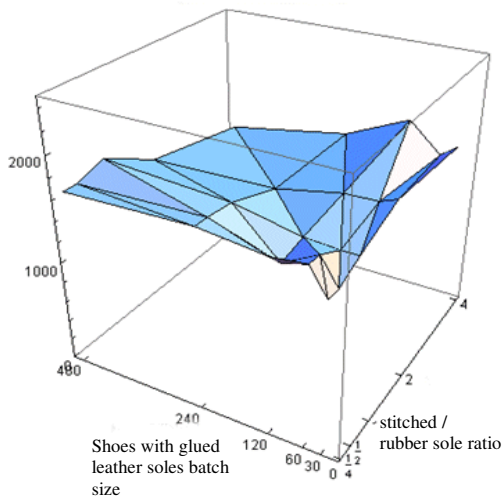

Figure 8 Through-put vs batch size when combining three shoe families, for the assembly area

As far as concerning the labour utilization under not critical production mixes, its overall saturation ranges from 64% up to 76% for the assembly area and most variations were found at the following areas:

- The cream island, its utilization increases by about 30% rising the quantity of shoes with stitched leather sole in the production plan
- The reactivation oven for rubber sole and the last removing island, their utilization is largely influenced by the batch size of shoes with rubber sole

Although these more complex production-mixes allowed for a validation of the simulation against the real production, we did not find a clear relationship between the production mix and through-put.

5.4. Finishing area overall performances

The finishing area performances are not directly related with shoe families/styles, but with finishing sequence as well as cycle time of each shoe article. Based on this consideration, all the shoe articles were grouped in three

macro categories i.e. easy, normal and difficult to finishing, and specific production mixes were defined with different compositions.
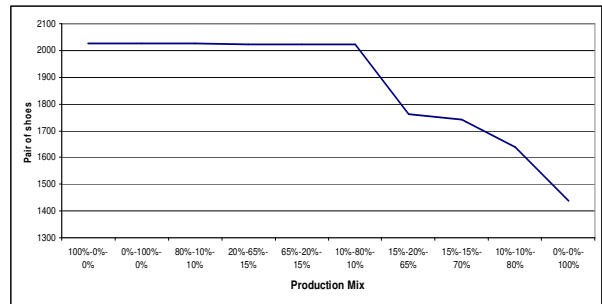

Figure 9 Through-put vs production mix

The daily through-put considering only finishing department, see Figure 9, ranges from about 1400 up to 2050 pairs of shoes and it is not influenced by batch size. Brushing and cream islands are the main bottlenecks and most of the finishing trolleys are staked up in these key position.

As far as concerning labour utilization, its saturation ranges from 72% to 95%, as shown in Figure 10, simulating only the finishing area.
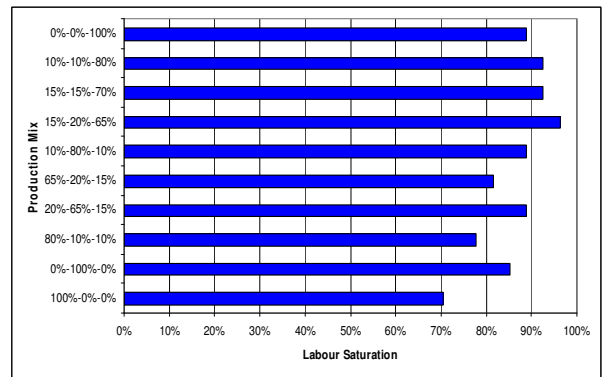
Figure 10 Labour saturation in the finishing area vs production mix

5.5. Production plant overall performances

Based on the previous results, the target productivity for the plant can be reached under scenarios with the following constraints:

- The rubber sole percentage in the daily production mix is lower then 30%
- The percentage of shoe articles with long cycle times in the finishing area is lower then 65%

In the first case, the assembly area is the bottleneck for the production plant, while in the second case the finishing department cut down the productivity.

Finally, a real production mix of two weeks was tested simulating the whole production plant as well as only the finishing area. In the first case the through-put is about 1863 pair of shoes per day, while in the latter, it is 2020 pair of shoes, indicating room for optimization.

A similar result was found analysing labour utilization through sensitivity analysis. Hourly productivity of the all production system is decreased by 10% when the number of available operators for the assembly area is reduced from 33 to 27. As expected for this production mix, decreasing the labor availability in the finishing area has no impact on the overall performances.

Some what if analysis were carried on some input parameters managing labour allocation, showing that a some potentials to increase through-put by a fine tuning activity.

6. CONCLUSION

This paper explored the use of simulation to better understand production dynamics as basis for determining an optimization strategy. The real shoe production plant provided a challenging example of a highly flexible production process operating on diverse production mixes.

Through a combination of analysing simple and complex scenarios, full picture of the production's dynamic was obtained. Simple use-cases were instrumental in identifying basic dynamics and understand the system response of more complex use-cases. The more complex use-cases, although difficult to interpret, had the advantage that they supported the validation of simulation results against real production.

Further research will concentrate on combining detailed modelling such as described in this paper with 'modelling the model' technologies, for overall optimization (testing against realistic use-cases) (Merkureyeva et all, 2008). We expect a combined approach of a time consuming detailed model and a less detailed but faster model enables to find concrete solutions for optimal sets of process parameters while reducing analysis time.

REFERENCES

Baiens, T., Hadfield, L., Mason, S., Ladbrook, J., 2003. Using empirical evidence of variations in worker performance to extend the capabilities of discrete event simulations in manufacturing. *In Proceedings of the 2003 Winter Simulation Conference*. 1210-1216

Boër, C.R., Avai, A., El-Chaar, 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 Systems*. Beijing

Chung, C.A, 2004. "Simulation modelling handbook", Crc Press

Kelton, W.D., Sadowski, R.P., and Sturrock, D.T., 2003. Simulation with Arena, 3rd Edition, WCB/McGraw-Hill, New York, NY.

Merkureyeva, G. Metamodelling for simulating applications in production and logistics. Available from: http://www.sim-serv.com [accessed 16 June 2008]

Merkureyeva, G, Brezhinska S., Brezhinskis J., Response surface-based simulation metamodelling methods. Available from: http://www.sim-serv.com [accessed 16 June 2008].

Roser, C., Nakano, M., and Tanaka, M., 2003. Buffer allocation model based on a single simulation. *In Proceedings of the 2003 Winter Simulation Conference*. 1230-1246.

AUTHORS BIOGRAPHIES

Fred Voorhorst is managing innovation at HUGO BOSS Ticino SA, a department for Product Development and Operation Management for five product groups, one of which is Shoes. He has more than ten years of experience in managing (business) innovation projects, both in industrial as well as academic context.

Antonio Avai is a partner and technical director of Technology Transfer System, an IT company located in Milan. He has managed research programs with 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.

Claudio Roberto Boër is Director of the ICIMSI Institute CIM for Sustainable Innovation of the University of Applied Science of Southern Switzerland. He has more than 16 years of industrial experience and research in implementation of computer aided design and manufacturing as well as design and setting up manufacturing and assembly flexible systems. He is author of the book on Mass Customization in the Footwear based on the European funded project EUROShoE that dealt, among several issues, with the complexity and optimization of footwear assembly systems.

SIMULATION CALCULATION OF TRACTOR-POTATO PLANTER COMBINATION MODEL

Jan Szczepaniak[a]

[a]Industrial Institute of Agriculture Engineering, Starołęcka 31, 60-963 Poznań, Poland

[a]janek@pimr.poznan.pl

ABSTRACT

The paper presents a concept of mathematical model, which may be used for analyzing the behavior of tractor – agricultural machine combinations at the preliminary design stages. The model is based on 2D so called bicycle models (two-wheeled) that are often referred to in the literature, but it is significantly modified because third axle was added. The model has been applied to a tractor – potato planter combination. The measurements of behavior tractor - potato planter combination during traffic were conducted to model parameters identification. Obtained results permit evaluate influence value of combination tractor – potato planter parameters on its traffic stability.

Keywords: agricultural machine, simulation, motion stability

1. INTRODUCTION

It is evident from statistical data that many serious accidents take place during the transfer of agricultural machines to the working site and during their movement while working. Nearly every third fatal accident in agriculture is caused by driving over, hitting or catching by vehicles and moving machines. The most serious effects occur in accidents during transport and those related to manoeuvring with tractors and machines (nearly 30% of all fatal accidents). One may assume that at least some of these accidents are caused indirectly by the design features of agricultural machines which unfavourably influence their traction properties.

Research concerned vehicles steerability and stability simulation were conducted for a long time. However the results were rarely transferred into the agricultural machine field. It demands taking into consideration specificity of agricultural machine construction and its combination with tractor on the one hand. On the other hand it demands analysis of machine behavior on diverse grounds (field ways, fields) possesses complex and changeable properties (Saarilahti 2002)

2. MODEL DESCRIPTION

As the result of increasing of computer possibility very complex, nonlinear models of vehicle start occur. But important disadvantage of a model with many degrees of freedom is great deal of data needed to circumscribe vehicle features. It is especially essential as the model is appropriated to using on preliminary design stage, when many data lack. Lacked and uncertain data decrease accuracy of results obtained by simulation and put usefulness of expenditure of work connected with model building in question.

In that case we made an attempt construct comparatively simple model with few degrees of freedom. Such models can be met in various publications concerned control of vehicle dynamics both transverse and longitudinal (Grzegożek 2000; Ślaski, Mac, and Szczepaniak 2000; Harada and Harada 1999).

The model is based on 2D so called bicycle models (two-wheeled) that are often referred to in the literature (Feng and He 2005; Börner and Isermann 2006), but it is significantly modified because third axle was added.

Attempt is made to apply the model to the tractor – potato planter combination (Fig. 1). Semimounted potato planter is attached to tractor by lower pull rod of 3 point linkage. Potato planter is equipped with castoring wheels. It necessary to emphasize that potato planter revolution with reference to tractor round Z axis normal to road surface is impossible.

Figure 1: Tractor – potato planter combination during field works

During deriving the model following assumptions were made:

- combination movement has constant longitudinal speed v_x,
- flexibility of engine mounting and seats are passed over,
- articulated joints clearances are passed over,
- turning angles, slip angles of wheels and of vehicle are so little to formulas $\sin(\Psi) \approx \Psi$ and $\cos(\Psi) \approx 1$ can be accepted.
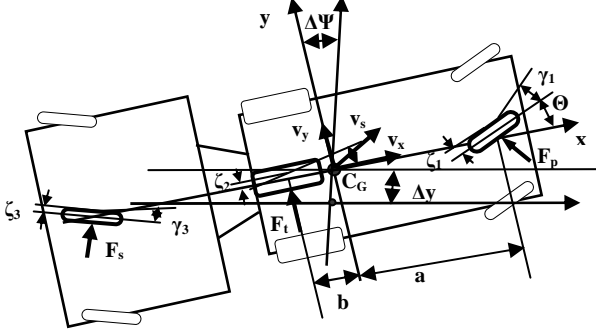

Figure 2: Three wheel 2D model of tractor – potato planter combination; symbols are described under model system of equations (1)

Proposed model has seven degrees of freedom: lateral displacement of combination center of gravity, combination rotation angle, lateral displacements of front and rear tractor wheels, slip angle of tractor's front and rear wheels, turning angle of castoring wheels. The longitudinal speed and tractor driven wheel turning angle are the input data.

System of equations (1) was obtained on the basis of Newton equation of dynamics and Kieldysz theory (Chatschaturov, 1976, Karnopp 2004) describing rolling of pneumatic wheel (Szczepaniak and Grzechowiak 2006).

$$\begin{cases} \dot{v}_y = \dfrac{c_{\xi_1}}{M}\zeta_1 + \dfrac{c_{\xi_2}}{M}\zeta_2 + \dfrac{m(\kappa^2 + a_k^2)}{M(a_k + b_k)}\ddot{\theta}_k - v_x\dot{\Psi} \\[2mm] \ddot{\Psi} = \dfrac{ac_{\xi_1}}{I_{zz}}*\zeta_1 - \dfrac{bc_{\xi_2}}{I_{zz}}*\zeta_2 - \dfrac{c}{I_{zz}}\dfrac{m(\kappa^2 + a_k^2)}{(a_k + b_k)}\ddot{\theta}_k \\[2mm] \dot{\zeta}_1 = -v_y - a\dot{\Psi} + v_x\Theta + v_x\gamma_1 \\[2mm] \dot{\zeta}_2 = -v_y + b\dot{\Psi} + v_x\gamma_2 \\[2mm] \dot{\gamma}_1 = -\dot{\Psi} - \dot{\Theta} - v_x a_1\left(\zeta_1 + \dfrac{K_{\delta_1}}{c_{\xi_1}}\gamma_1\right) \\[2mm] \dot{\gamma}_2 = -\dot{\Psi} - v_x a_2\left(\zeta_2 + \dfrac{K_{\delta_2}}{c_{\xi_2}}\gamma_2\right) \\[2mm] \ddot{\theta}_k = -\dfrac{v_x k_k}{K_{\delta_k}}\ddot{\theta}_k - k_k\dfrac{(a_k + b_k)^2}{m(\kappa^2 + a_k^2)}\dot{\theta}_k - k_k v_x\dfrac{(a_k + b_k)}{m(\kappa^2 + a_k^2)}\theta_k \end{cases} \qquad (1)$$

where Ψ – combination rotation angle [rad], v_y and v_x – components (lateral and longitudinal) of velocity of combination center of gravity [m/s], ζ_1, ζ_2 – side displacement front and rear tractor wheels [m], γ_1, γ_2 – slipping angle front and rear tractor wheels [rad]. Θ – mean turning angle of driven wheels [rad], θ_k – turning

angle of castoring wheels [rad], a – distance from front tractor axle to combination center of gravity [m], b – distance from rear tractor axle to combination center of gravity [m], c – distance from potato planter axle to combination center of gravity [m], M – combination mass [kg], m – castoring wheels mass [kg], I_{zz} – combination moment of inertia with respect to vertical axis [kg*m^2], c_{ξ_1}, c_{ξ_2} – coefficients of lateral stiffness of tractor tires [N/m], K_{δ_1}, K_{δ_2}, K_{δ_k} – cornering coefficients of tractor and potato planter wheels [N/rad], a_1, a_2 – coefficients of deflected tires curvature [m^{-2}], a_k – distance from castoring wheel pivot axis to its center of gravity [m], b_k – distance from the castoring wheel pivot axis to its center of gravity [m], k_k – coefficient of castoring wheel sidewall stiffness [kg/rad].

3. CONDUCTED TEST AND SIMULATION

The identification experiment was carried out to estimate exact values of combination model parameters. The investigation was aimed at getting data about combination behavior during execute various maneuvers (change traffic lane, pass, turning with various radii) and at loads acting on combination wheels. During the measurements ISO recommendations were applied.

The modern, specially worked out apparatus kit was using during test (figure 3). It contains among other things sensor for simultaneous measurement of wheel position and orientation and microwave sensors for speed and distance measurement (Sparrer 2003). All loads (forces and moments) acting on tractor wheels and its angle velocity were measured too.

Figure 3: Testing equipment on the tractor front wheel: a) velocity sensor, b) system for measure of driven wheel position

Tractor tires property investigation was made too. The series of measurements of lateral, longitudinal, side and cornering stiffness of tires were carried out.

Obtained data was used to parameters identification of tractor - potato planter combination model.

4. MODEL IDENTIFICATION

The identification was made by minimization of loss function describing estimation error of model output (Szczepaniak 2008). Identification problem was reduced

to single criterion minimization task. It can be solved by means of many procedures available for example in Matlab system (Coleman, Branch, and Grace 1999). It was assumed, that objective function describing estimation error should be non-negative function of estimated parameters:

$$F_{critt}(par) = \frac{1}{n}\sum_{k=1}^{m}\sum_{i=1}^{n}\alpha_k\left(y_{mk}(t_i) - y_{pk}(t_i)\right)^2 \qquad (2)$$

where n – number of simulation steps, *par* – vector of identified parameters of model, t_i – moments of simulation and measurement steps, y_m – m dimensional vector of model state parameters obtained out of simulation, y_p – m dimensional vector of model state parameters obtained out of identification experiment, α_k – weights of state parameters.

Properties of objective function (2) are difficult to determine because to it's calculate system of model equations (1) should be solved. In that case three different algorithms available in Matlab were tested:

- Nelder-Mead method (*fminsearch* function in Matlab) – finds minimum of unconstrained multivariable function using derivative-free method,
- *fmincon* – finds minimum of constrained nonlinear multivariable function,
- *lsqnonlin* – solves nonlinear least-squares (nonlinear data-fitting) problems:

$$\min_x f(x) = \frac{1}{2}\left(f_1^2(x) + f_2^2(x) + \cdots + f_m^2(x)\right) = \frac{1}{2}\|F(x)\|_2^2 \qquad (3)$$

Test of selected optimization methods effectiveness was conducted by means of numerical experiment, on the basis results obtained out of computer simulation. Optimal values of identified parameters were known owing to using the same model as standard model and identified model. Beside the problems connected with measurement inaccuracy and model simplification were avoided. Used approach makes possible minimize identification error to zero during the test. It allows to comparison and evaluation of proposed optimization methods efficiency.

As the result of carried out numerical experiment it turned out that the procedure *fmincon* is the worst. The best is the procedure *fminsearch* (Nelder Mead method).

After numerical test and optimization method selection proper identification of model parameters was started.

The identification was carried out by means of minimization of function (2) on the basis comparison values of model state parameters such as lateral velocity or combination rotation angle obtained from simulation and from measurements. Vector of identified parameters included: I_{zz}, $K_{\delta 1}$, $K_{\delta 2}$, $K_{\delta k}$, κ, k_s, $c_{\xi 1}$, $c_{\xi 2}$, a_1, a_2.
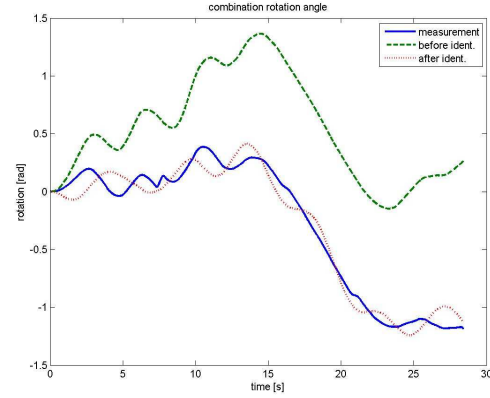

Figure 4: Charts of combination rotation angle (Ψ) obtained from measurement and from simulation before and after identification

5. FORECAST SIMULATION

After getting correct parameters values of model tractor – potato planter, model can be used to forecast simulation conducting. It enables constructors to describe influence of the parameters on combination behaviour.

In the real construction we can change one of the parameters only with simultaneously change another one. For example: we can change location of combination axle only together with change centre of gravity location and moments of inertia. Using the model we can model such changes too of course. But we can estimate influence of only single parameter on machine behaviour. It facilitates constructor decision concerning selection of changed parameters.

Sampled analyses of tractor – potato planter combination behaviour are presented. The legends enclosed to the charts show values of the changed parameters.

On the figures 5 and 6 we can see changes of combination trajectories caused to changes of distance between combination center of gravity and tractor front axle.
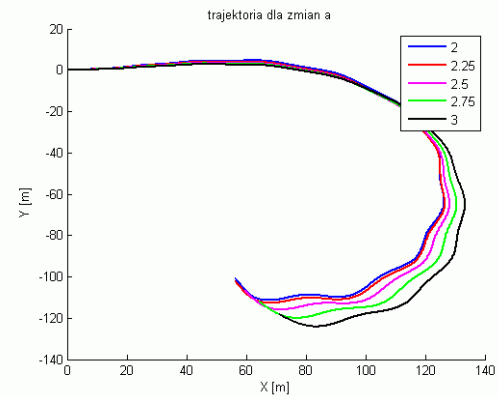

Figure 5: Changes of combination trajectory for various distances from front tractor axle to combination center of gravity (the distances values are shown on the legend [m]) during ride along a circle
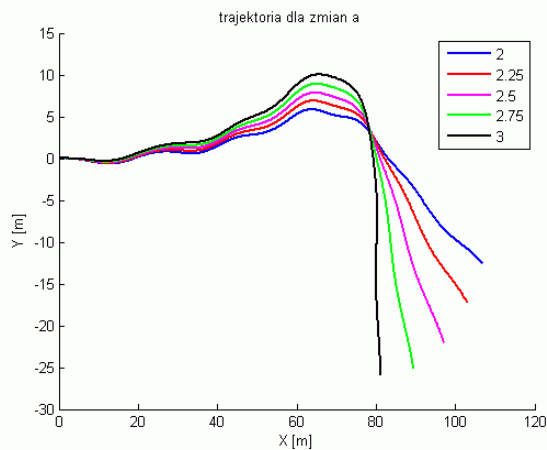
Figure 6: Changes of combination trajectory for various distances from front tractor axle to combination center of gravity (the distances values are shown on the legend [m]) for cornering

We can check, that according to expectation (potato planter wheels are castoring wheels) change only distance betweeen axle of potato planter and combination center of gravity doesn't influence on motion of the vehicle significantly (figure 7).
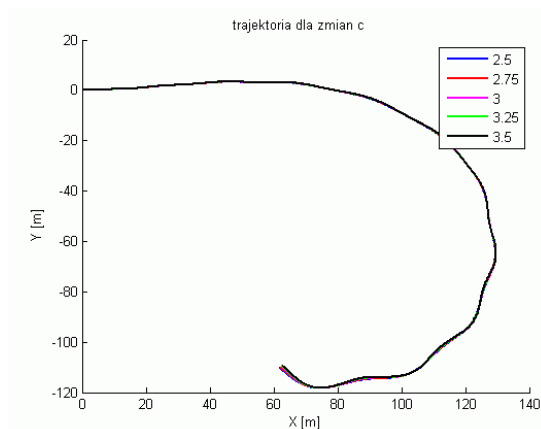

Figure 7: Changes of combination trajectory for various distances from potato planter axle to combination center of gravity (the distances values are shown on the legend [m]) during ride along a circle

The location of centre of gravity has especially influence on combination behaviour. Its shift means simultaneous change of distances from both tractor axles and planter axle. The positive value showed on the legend means forward shift (figure 8).
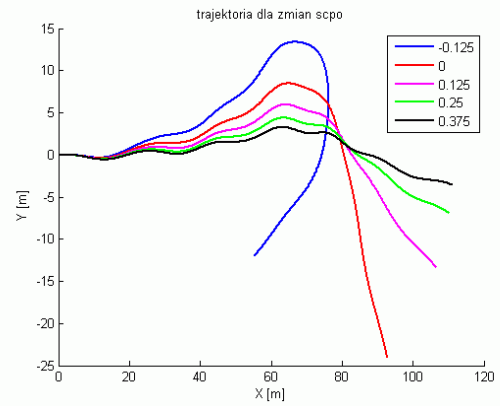
Figure 8: Changes of combination trajectory for various combination center of gravity position (the shift values are shown on the legend [m], positive values mean forward shift) for cornering

We can see that back shift center of gravity is able to result in stability loss (blue line).

Test of interactions between individual model parameters is important too. The figures 9 and 10 show the influence of tractor front axle tires lateral stiffness on adequately front and back tractor tires side displacement.

Figure 9: The changes of lateral displacement of tractor front axle tires (model variable ζ_1 [m]) for various values of lateral stiffness of tractor front axle tires $c_{\xi 1}$ [N/m] (see legend)

According to expectations less displacement occurred for bigger tires stiffness. It is visible especially during changes of tires displacement direction.

243

Figure 10: The changes of lateral displacement of tractor back axle tyres (model variable ζ_2 [m]) for various values of lateral stiffness of tractor front axle tires $c_{\xi 1}$ [N/m] (see legend)

Figure 10 shows that change of tractor front axle tires stiffness doesnn't influence significantly on behaviour of back axle tires.

Similarly, change of lateral stiffness of back axle tires causes changes of dislacements only for back axle tires (figures 11 and 12).
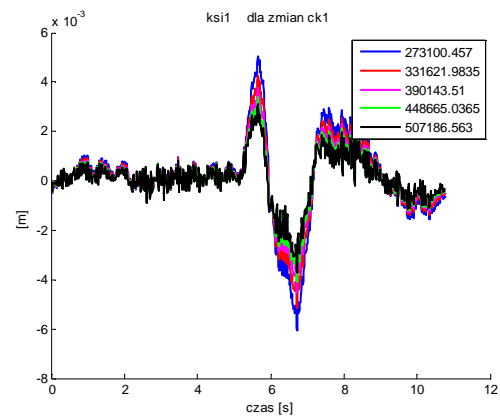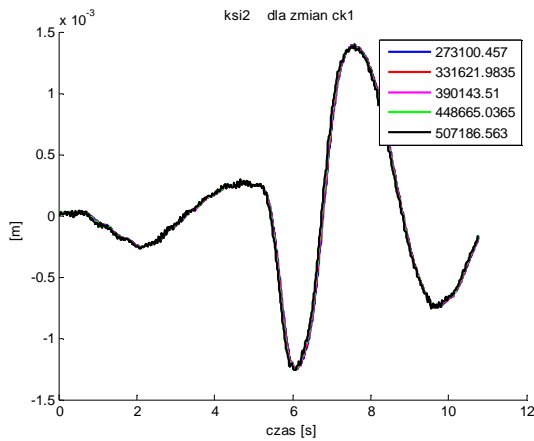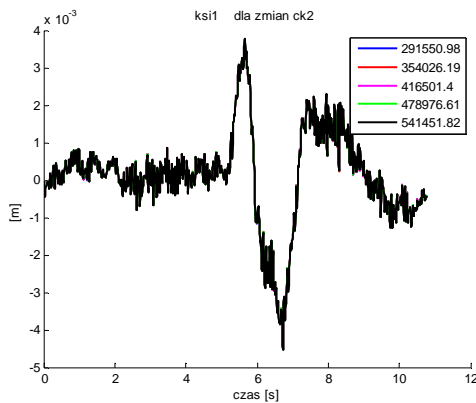

Figure 11: The changes of lateral displacement of tractor front axle tires (model variable ζ_1 [m]) for various values of lateral stiffness of tractor back axle tires $c_{\xi 1}$ [N/m] (see legend)
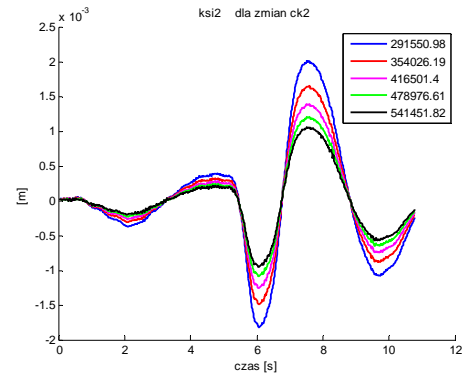

Figure 12: The changes of lateral displacement of tractor back axle tires (model variable ζ_2 [m]) for various values of lateral stiffness of tractor back axle tires $c_{\xi 1}$ [N/m] (see legend)

On figure 12 we can see that bigger stiffness of back axle tires cause less lateral displacement of this axle tires, The differences are less than for tires of front axle. It is caused probably by bigger pressure on tractor back axle.

6. CONCLUSIONS

Our long-term goal of presented work is providing agricultural machine constructors with tools (adequate models) making possible machine traffic on the diverse ground simulation on the relatively early stage of designing, before prototype construction. It will allow taking into consideration construction properties influence on agricultural machine way of motion, particularly on motion stability.

REFERENCES

Börner, M., Isermann, R., 2006. Model-based detection of critical driving situations with fuzzy logic decision making. *Control Engineering Practice,* 14,527-536.

Chatschaturov, A.A., 1976. *Dinamika sistiemy doroga-szina-awtomobil-woditiel.* Moscow: Mashino-stroienie.

Coleman, T., Branch, M. A., Grace, A., 1999*: Optimization Toolbox For Use with Matlab. User Guide version 2.* Natick: The Math Works Inc.

Feng, L. He, Y.;2005. Study on dynamic model of tractor system for automated navigation applications. *Journal of Zheijang University Science,* 6A(4), 270-275.

Grzegożek, W., 2000. *Modelowanie dynamiki samochodu przy stabilizującym sterowaniu siłami hamowania.* Kraków: Wydawnictwo Politechniki Krakowskiej.

Harada, M., Harada, H., 1999. Analysis of lateral stability with integrated control of suspension and steering system. *Society of Automotive Engineers of Japan Review*, 20(4), 465-470.

Karnopp, D., 2004. *Vehicle Stability*, New York: Marcel Dekker, Inc.

Lugner, P., Pacejka, H. B., Plöchl, M. 2005. Recent advances in tyre models and testing procedures. *Vehicle System Dynamics*, 43, 413-436.

Pacejka, H. B. 2006. *Tyre and Vehicle Dynamics*. Amsterdam: Butterworth Heinemann.

Saarilahti, M., 2002. *Soil Interaction Model. Quality of Life and Management of Living Resources*. Quality of Life and Management of Living Resources Contract No. QLK5-1999-00991. Available from: http://ethesis.helsinki.fi/ julkaisut/maa/mvaro/publications/31/soilinte.html [accesed 20 September 2005]

Sparrer. B. 2003. *TurboLab 5. User Manual*. MDZ Bührer & Partner GbR. Available from: http://www.corrsys-datron.com/Support/Manuals/ c_Software/cds-m_Turbolab_Analysis_e.pdf [accessed 10 February 2008]

Szczepaniak, J., 2008. Identyfikacja parametrów modelu agregatu rolniczego ciągnik – sadzarka do ziemniaków. *Journal of Research and Application in Agricultural Engineering*, 53(2), 16-20.

Szczepaniak, J., Grzechowiak, R., 2006. Modelling of a Tractor – Potarto Planter Combination for Use in Traffic Simulation. *Machine Dynamics Problems*, vol. 30(4), 140-145.

Ślaski, G., Mac, J., Szczepaniak, J., 2000. Komputerowa analiza obciążeń dynamicznych elementów zawieszenia i ramy nośnej na przykładzie wielozadaniowej naczepy rolniczej. *Problemy Budowy oraz Eksploatacji Maszyn i Urządzeń Rolniczych*, 2, 251-258.

AUTHORS BIOGRAPHY

Jan Szczepaniak received his MS (1977) and the Ph.D. (1999) in Engineering from Poznan University of Technology. He is head of Modern Method of Designing and Testing Department in Industrial Institute of Agriculture Engineering. His main research areas include agricultural engineering, modelling, simulation, signal processing and finite element modeling. During his career, Dr. Szczepaniak is author or co-author over 100 technical papers. He is also a designer of agricultural machines awarded in international fairs. For example, for device for opening the pair of side arms of the multi-segment land of roller he achieved a gold medal of The Belgian and International Trade Fair for Technological Innovation in 2007 and a silver medal of The 107th International Invention Salon "Concours – Lepine" in 2008.

THE WEB WINDING SYSTEM CONTROL BY THE BACKSTEPPING METHOD

Nabila RABBAH[(a)], Bahloul BENSASSI[(b)]

[(a,b)]University Hassan II, Faculty of Sciences, Aïn-Chock, Casablanca – Morocco.

[(a)]nabila_rabbah@yahoo.fr, [(b)]bahloul_bensassi@yahoo.fr

ABSTRACT

In the industry of flat rolled sheet product, strip crown and shape has been one of the most important factors for quality assurance, productivity improvement, cost effectiveness, and customer satisfaction. The transient work roll profile influences load distribution and imprints an undesirable profile on rolled strip. We designate by a winding process any system applying the cycles of unwinding, transport, treatment and winding to various flat products. This system knows several constraints such as the thermal and mechanical effects that generate dysfunctions due to the influence of the process conditions. For this installations type, the various automatisms functions, often very advanced, are realized in modular systems with distributed architecture. In the present paper, a backstepping control technique is proposed. The control variables are velocities and tractions forces along the web winding system. The proposed control law and Lyapunov function guarantee asymptotic stability from all initial values.

Keywords: Web Winding System (WWS), Nonlinear Control, Lyapunov Functions, Backstepping method.

1. INTRODUCTION

In the early days of control theory investigation, most of concepts such as stability, optimality and uncertainty were descriptive rather than constructive. In the recent two decades, a number of new methods have been developed for designing controllers to control nonlinear dynamic systems. These are mainly recursive methods, such as backstepping, forwarding, and various combinations of them. A common concept of the above named basic recursive methods is the design of a globally stable control system, having a cascade structure, for a class of nonlinear dynamic systems. In particular, the backstepping method is based on Lyapunov function theory (La Salle and Lefschetz 1961), but its origin can be found in some theories of linear control, such as the feedback linearisation method or the LQR method.

The beginning of the development of the backstepping method applied to nonlinear control systems design dates back to the end of the 1980s. A list and a discussion of publications issued at that time can be found in the overview by Sontag and Sussmann, (1989), as well as by Kokotovic and Arcak (2001). The backstepping method is based directly on the mathematical model of the examined system, introducing new variables into it in a form depending on the state variables, controlling parameters, and stabilising functions. The controlled system may be in the state equations with a triangular form. The design of the controller pass by several step, in the first step we consider a Lyapounov function for the first error state, then, the virtual control is calculated in the order to guarantee the negativity of the Lyapounov function proposed. For this virtual control, we associate a second error sate, between the second state and the virtual control calculated in first step, then we consider the augmented joint Lyapounov function whose the first function and the second error are appear. The second virtual control is calculate with the same reasoning. The exact control will calculate in the last step by using the virtual control laws calculated in the past steps. We can interpret this method by the addition of the integrators after each step (Krstic, Kanellakopulos and Kokotovic 1995).

This paper presents a new concept of web winding system in which control velocities and tensions are derived for nonlinear controllers designed with the Backstepping method. The dynamics of a Web Winding System (WWS) is described by its strongly nonlinear behavior. In all cases of rolling up or unfolding of a web material, the flatness difficulty arises. Considering the complexity of the system due to nonlinearity and the strong coupling between the web velocity and the web tension, it is more convenient to linearize this WWS. However, this model remains very depend on the set point considered and especially on the variation rate of nonlinearities. This situation pushed the researchers to be directed more and more towards the techniques of the nonlinear control based on the Backstepping's technique. The paper is organized as follows: The web winding model is described in Section 2, with a brief description of the WWS. A detailed description of winding control design methodology is in Section 4. In Section 5 the control system performances are evaluated in simulation of the web winding model. The last section concludes the paper.

2. PLANT MODEL

2.1. WWS description

The web system is very important in a rolling mill, because its parameters determine the strip quality. Among its parameters, we quote:

- The entry and exit of the traction forces.
- The entry and exit velocities ensured by the winders motors, and the work rolls velocity (Fig.1).
- The pressure force or the variations between the work rolls and their parallelism.

The variation of the exit strip flatness evolves because of the thermal dilation of the cylinders (Rabbah and Bensassi, 2006; Rabbah and bensassi 2007a), but also due to the elasticity forces (Schmitz and Herman 1995). To avoid this phenomenon, the traction forces are applied to limit the elasticity of the rolled material.

The thickness control is ensured by programmable automats, which are called AGC (Automatic Gauge Control system), (Ueno and Sorao 2004). Their goal is to maintain the strip thickness uniform in spite of the acting factors to change it. Considering the complexity of the Cold Rolling System (CRM), the modeling and the control of the WWS should be studied to minimize the flatness defaults. With this intention, we start with the development of a mathematical model describing the dynamic behavior of the system.
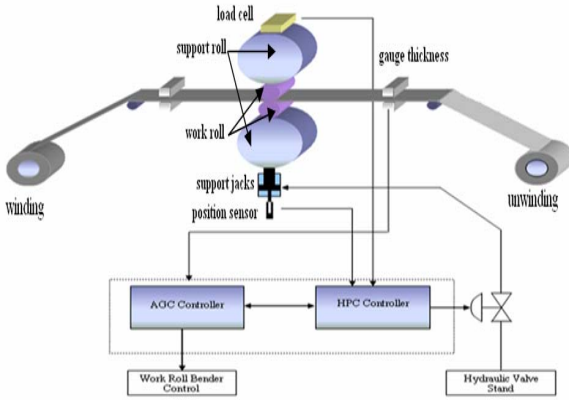

Figure 1: Interactions between the components of the cold rolling system.

2.2. Global model

Let us consider the nonlinear model of the wws (Rabbah and Bensassi 2007b; Rabbah and bensassi 2008) defined by the state representation (1) which can be put in the general form of the nonlinear affine control system:

$$\begin{cases} \dot{x} = f(x,u) = f(x) + g(x)u \\ y = h(x) \end{cases} \quad (1)$$

with:

$$x = \begin{bmatrix} V_1 & T_1 & V_2 & T_3 & V_3 \end{bmatrix}^T = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix}^T,$$

$$y = \begin{bmatrix} T_1 & V_2 & T_3 \end{bmatrix}^T \text{ et } u = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix}^T$$

$$f(x) = \begin{bmatrix} \dfrac{gr_1^2}{J_1} x_2 - \dfrac{1}{\tau_{em1}} x_1 \\[2mm] -\dfrac{1}{L} x_2 x_3 - \dfrac{ES}{L} x_1 + \dfrac{ES}{L} x_3 \\[2mm] -\dfrac{gr_2^2}{J_2} x_2 + \dfrac{gr_2^2}{J_2} x_4 - \dfrac{1}{\tau_{em2}} x_3 \\[2mm] -\dfrac{1}{L} x_3 x_4 - \dfrac{ES}{L} x_3 + \dfrac{ES}{L} x_5 \\[2mm] -\dfrac{gr_3^2}{J_3} x_4 - \dfrac{1}{\tau_{em3}} x_5 \end{bmatrix} \quad h(x) = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$g(x) = \begin{bmatrix} g1, g2, g3 \end{bmatrix} = \begin{bmatrix} \dfrac{r_1}{\tau_{em1}.k_1} & 0 & 0 \\[2mm] 0 & 0 & 0 \\[2mm] 0 & \dfrac{r_2}{\tau_{em2}.k_2} & 0 \\[2mm] 0 & 0 & 0 \\[2mm] 0 & 0 & \dfrac{r_3}{\tau_{em3}.k_3} \end{bmatrix}$$

These parameters characterize the system: it is multivariable, strongly coupled, nonlinear and time-varying.

The model (1) is composed of three subsystems: The first one is of the state vector $[V_1, T_1]^T$, controlled by the tension U_1, the second has as a state vector $[V_2]$, is controlled by U_2 and the third subsystem has as a state vector $[T_2, V_3]^T$, is controlled by the tension U_3.

Table 1 contains the operating condition used in this analysis.

Table 1: parameters of the operating conditions

Web length between winder and unwinder	1.15 m
Young's modulus	0.16 109 kg/m²
Web section	0.19 mm²
Sliding coefficient	0.8
Strip or web thickness	1.6 mm
Strip or web width (largeur)	600 mm
Diameter or work rolls	0.45 m
Nominal torque of unwinding/winding motors	700 kN.m
Nominal velocity of unwinding/ winding motors	120 Rpm
Rolling speed	1400 mpm

3. THE BACKSTEPPING CONTROLLERS

The control problem considered consists in forcing the web velocities and the downstream/upstream web tensions to follow the reference signals given, noted

respectively V_2^{ref}, T_1^{ref} and T_3^{ref}. This suggests the following errors:

$$e_1 = T_1^{ref} - T_1 \tag{2}$$

$$e_2 = V_2^{ref} - V_2 \tag{3}$$

$$e_3 = T_3^{ref} - T_3 \tag{4}$$

The regulator synthesis will be done in two steps. In the first step, we will put the virtual controls and the stabilising functions associated. In the second step, we determine the control laws able to ensure convergence towards zero of the difference between the virtual orders and the stabilising functions associated.

Step1: The dynamic of the tracking errors e_1, e_2 and e_3 are given by:

$$\dot{e}_1 = \dot{T}_1^{ref} - \frac{ES}{L}(V_2 - V_1) + \frac{V_2}{L}T_1 \tag{5}$$

$$\dot{e}_2 = \dot{V}_2^{ref} + \frac{gr_2^2}{J_2}(T_1 - T_3) + \frac{1}{\tau_{em2}}V_2 - \frac{r_2}{\tau_{em2}\cdot k_2}U_2 \tag{6}$$

$$\dot{e}_3 = \dot{T}_3^{ref} - \frac{ES}{L}(V_3 - V_2) + \frac{V_2}{L}T_3 \tag{7}$$

The tracking error e_2 tends asymptotically towards zero if; in this case, the virtual control is selected as follows:

$$\alpha_2 = \dot{V}_2^{ref} + \frac{gr_2^2}{J_2}(T_1 - T_3) + \frac{1}{\tau_{em2}}V_2 \tag{8a}$$

Thus, the dynamic error e_2 can be rewritten as follows:

$$\dot{e}_2 = \alpha_2 - \frac{r_2}{\tau_{em2}k_2}U_2 = -c_2 e_2 \tag{8b}$$

The quantities $\frac{ES}{L}(V_2 - V_1)$ and $\frac{ES}{L}(V_3 - V_2)$ are posed like virtual control inputs for the system (5) and (7). It follows that the tracking errors e_1 and e_3 tend asymptotically towards zero if these virtual controls are selected such as:

$$\frac{ES}{L}(V_2 - V_1) = \alpha_1 \quad \text{with} \quad \alpha_1 = c_1 e_1 + \dot{T}_1^{ref} + \frac{V_2}{L}T_1 \tag{9}$$

$$\frac{ES}{L}(V_3 - V_2) = \alpha_3 \quad \text{with} \quad \alpha_3 = c_3 e_3 + \dot{T}_3^{ref} + \frac{V_2}{L}T_3 \tag{10}$$

where c_1, c_2 and c_3 are the positive real constants unspecified.
Indeed, by doing this we obtain:
$\dot{e}_1 = -c_1 e_1$, $\dot{e}_2 = -c_2 e_2$ and $\dot{e}_3 = -c_3 e_3$

The α_1, α_2 and α_3 quantities are called stabilising functions. Then, the first Lyapunov candidate function is defined as:

$$v_1 = \frac{1}{2}\left(e_1^2 + e_2^2 + e_3^2\right) \tag{11}$$

The derivative of the first Lyapunov function takes the form:

$$\dot{v}_1 = -c_1 e_1^2 - c_2 e_2^2 - c_3 e_3^2 \tag{12}$$

who shows the influence of the controlling parameters c_1, c_2 and c_3 on the convergence of the e_1, e_2 and e_3 errors.

For the equations (9) and (10), the found result supposes that $\frac{ES}{L}(V_2 - V_1)$ and $\frac{ES}{L}(V_3 - V_2)$ are effective controls. As such is not the case, we cannot impose the equalities (9) and (10). We will only try to tend these controls towards their ideal trajectories which are precisely the stabilising functions α_1 and α_3. For this purpose, we introduce the errors:

$$z_1 = \alpha_1 - \frac{ES}{L}(V_2 - V_1) \tag{13}$$

$$z_3 = \alpha_3 - \frac{ES}{L}(V_3 - V_2) \tag{14}$$

The dynamic of the e_1 and e_3 errors are expressed in function of z_1 and z_3 as follows:

$$\dot{e}_1 = -c_1 e_1 + z_1 \tag{15}$$

$$\dot{e}_3 = -c_3 e_3 + z_3 \tag{16}$$

The second step in the regulator synthesis consists in forcing all the $(e_1, e_2, e_3, z_1, z_3)$ errors to converge towards zero with a suitable choice of the effective controls U_1, U_2 and U_3.

Step 2: The dynamic of the e_1 error is given by:

$$\begin{aligned}
\dot{z}_1 &= \dot{\alpha}_1 - \frac{ES}{L}(\dot{V}_2 - \dot{V}_1) \\
&= c_1 \dot{e}_1 + \ddot{T}_1^{ref} + \frac{1}{L}T_1 \cdot \dot{V}_2 + \frac{1}{L}V_2 \cdot \dot{T}_1 - \frac{ES}{L}(\dot{V}_2 - \dot{V}_1)
\end{aligned} \tag{17}$$

248

Taking into account (13) and (9), we obtain:

$$\dot{z}_1 = c_1(-c_1 e_1 + z_1) + \ddot{T}_1^{ref} + \frac{1}{L}T_1 \cdot \left(-\frac{g r_2^2}{J_2}T_1 + \frac{g r_2^2}{J_2}T_3 - \frac{1}{\tau_{em2}}V_2 + \frac{r_2}{\tau_{em2}k_2}U_2\right)$$
$$+ \frac{1}{L}V_2 \cdot \left(-\frac{1}{L}T_1 \cdot V_2 - \frac{ES}{L}V_1 + \frac{ES}{L}V_2\right)$$
$$- \frac{ES}{L}\left(\frac{g r_2^2}{J_2}T_1 + \frac{g r_2^2}{J_2}T_3 - \frac{1}{\tau_{em2}}V_2 + \frac{r_2}{\tau_{em2}k_2}U_2 - \frac{g r_1^2}{J_1}T_1 + \frac{1}{\tau_{em1}}V_1 - \frac{r_1}{\tau_{em1}k_1}U_1\right) \quad (18)$$
$$= \beta_1 - \left(\frac{ES}{L}\left(\frac{r_2}{\tau_{em2}k_2}U_2 - \frac{r_1}{\tau_{em1}k_1}U_1\right) - \frac{T_1}{L}\cdot\frac{r_2}{\tau_{em2}k_2}U_2\right)$$

where β_1 includes the measurable terms on the right of the first equality, that is to say:

$$\beta_1 = c_1(-c_1 e_1 + z_1) + \ddot{T}_1^{ref} + \frac{1}{L}T_1\left(\frac{g r_2^2}{J_2}(T_3 - T_1) - \frac{1}{\tau_{em2}}V_2\right)$$
$$+ \frac{1}{L}V_2 \cdot \left(-\frac{1}{L}T_1 \cdot V_2 + \frac{ES}{L}(V_2 - V_1)\right) \quad (19)$$
$$- \frac{ES}{L}\left(\frac{g r_2^2}{J_2}(T_3 - T_1) - \frac{g r_1^2}{J_1}T_1 + \frac{1}{\tau_{em1}}V_1 - \frac{1}{\tau_{em2}}V_2\right)$$

In the same way, the dynamics of the e_3 error is:

$$\dot{z}_3 = \dot{\alpha}_3 - \frac{ES}{L}(\dot{V}_3 - \dot{V}_2)$$
$$= c_3 \dot{e}_3 + \ddot{T}_3^{ref} + \frac{1}{L}T_3 \cdot \dot{V}_2 + \frac{1}{L}V_2 \cdot \dot{T}_3 - \frac{ES}{L}(\dot{V}_3 - \dot{V}_2) \quad (20)$$

While using (16) and (10), the preceding equation becomes:

$$\dot{z}_3 = c_3(-c_3 e_3 + z_3) + \ddot{T}_3^{ref} + \frac{1}{L}T_3 \cdot \left(\frac{g r_2^2}{J_2}T_1 + \frac{g r_2^2}{J_2}T_3 - \frac{1}{\tau_{em2}}V_2 + \frac{r_2}{\tau_{em2}k_2}U_2\right)$$
$$+ \frac{1}{L}V_2 \cdot \left(\frac{1}{L}V_2 \cdot T_3 - \frac{ES}{L}V_2 + \frac{ES}{L}V_3\right)$$
$$- \frac{ES}{L}\left(-\frac{g r_3^2}{J_3}T_3 - \frac{1}{\tau_{em3}}V_3 + \frac{r_3}{\tau_{em3}k_3}U_3 + \frac{g r_2^2}{J_2}T_1 - \frac{g r_2^2}{J_2}T_3 + \frac{1}{\tau_{em2}}V_2 - \frac{r_2}{\tau_{em2}k_2}U_2\right) \quad (21)$$
$$= \beta_3 - \left(\left(\frac{ES}{L} - \frac{T_3}{L}\right)\frac{r_2}{\tau_{em2}k_2}U_2 + \frac{ES}{L}\cdot\frac{r_3}{\tau_{em3}k_3}U_3\right)$$

where β_3 includes the measurable terms on the right of the first equality, that is to say:

$$\beta_3 = c_3(-c_3 e_3 + z_3) + \ddot{T}_3^{ref} + \frac{1}{L}T_3 \cdot \left(\frac{g r_2^2}{J_2}(T_3 - T_1) - \frac{1}{\tau_{em2}}V_2\right)$$
$$+ \frac{1}{L}V_2 \cdot \left(-\frac{1}{L}V_2 \cdot T_3 + \frac{ES}{L}(V_3 - V_2)\right) \quad (22)$$
$$- \frac{ES}{L}\left(\frac{g r_2^2}{J_2}(T_1 - T_3) - \frac{g r_3^2}{J_3}T_3 + \frac{1}{\tau_{em2}}V_2 - \frac{1}{\tau_{em3}}V_3\right)$$

To study the stability of the system (8b), (15), (16), (18) and (21), of state vector $(e_1, e_2, e_3, z_1, z_3)$, we consider the Lyapunov candidate function increased:

$$v_2 = v_1 + \frac{1}{2}(z_1^2 + z_3^2) \quad (23)$$

Its derivative with respect to time:

$$\dot{v}_2 = e_1 \dot{e}_1 + e_2 \dot{e}_2 + e_3 \dot{e}_3 + z_1 \dot{z}_1 + z_3 \dot{z}_3 \quad (24)$$

Taking into account (8b), (15), (16), (18) and (21), we obtain:

$$\dot{v}_2 = -c_1 e_1^2 - c_2 e_2^2 - c_3 e_3^2 - d_1 z_1^2 - d_3 z_3^2$$
$$+ z_1\left(e_1 + d_1 z_1 + \beta_1 - \left(\frac{ES}{L}\left(\frac{r_2}{\tau_{em2}k_2}U_2 - \frac{r_1}{\tau_{em1}k_1}U_1\right) - \frac{T_1}{L}\cdot\frac{r_2}{\tau_{em2}k_2}U_2\right)\right)$$
$$+ z_3\left(e_3 + d_3 z_3 + \beta_3 - \left(\left(\frac{ES}{L} - \frac{T_3}{L}\right)\cdot\frac{r_2}{\tau_{em2}k_2}U_2 + \frac{ES}{L}\cdot\frac{r_3}{\tau_{em3}k_3}U_3\right)\right) \quad (25)$$
$$+ e_2\left(c_2 e_2 + \alpha_2 - \frac{r_2}{\tau_{em2}k_2}U_2\right)$$

where d_1 et d_3 are the positive real constants unspecified.

The preceding equation suggests choosing the U_1, U_2 and U_3 controls such as:

$$\begin{cases} e_1 + d_1 z_1 + \beta_1 - \left(\frac{ES}{L}\left(\frac{r_2}{\tau_{em2}k_2}U_2 - \frac{r_1}{\tau_{em1}k_1}U_1\right) - \frac{T_1}{L}\cdot\frac{r_2}{\tau_{em2}k_2}U_2\right) = 0 \\[2mm] e_3 + d_3 z_3 + \beta_3 - \left(\left(\frac{ES}{L} - \frac{T_3}{L}\right)\cdot\frac{r_2}{\tau_{em2}k_2}U_2 + \frac{ES}{L}\cdot\frac{r_3}{\tau_{em3}k_3}U_3\right) = 0 \\[2mm] c_2 e_2 + \alpha_2 - \frac{r_2}{\tau_{em2}k_2}U_2 = 0 \end{cases} \quad (26)$$

we can deduce the three laws control, there forms are:

$$\begin{cases} U_1 = \frac{L}{ES}\cdot\frac{\tau_{em1}k_1}{r_1}\cdot\left(e_1 + d_1 z_1 + \beta_1 - \left(\frac{ES}{L} - \frac{T_1}{L}\right)\left(\frac{r_2}{\tau_{em2}k_2}\right)\left(\frac{\tau_{em2}k_2}{r_2}(c_2 e_2 + \alpha_2)\right)\right) \\[2mm] U_2 = \frac{\tau_{em2}k_2}{r_2}(c_2 e_2 + \alpha_2) \\[2mm] U_3 = \left(\frac{L}{ES}\cdot\frac{\tau_{em3}k_3}{r_3}\right)\left(e_3 + d_3 z_3 + \beta_3 + \frac{r_2}{\tau_{em2}k_2}\left(\frac{ES}{L} - \frac{T_3}{L}\right)\left(\frac{\tau_{em2}k_2}{r_2}(c_2 e_2 + \alpha_2)\right)\right) \end{cases} \quad (27)$$

The derivative $\dot{v}_2$ becomes:

$$\dot{v}_2 = -c_1 e_1^2 - c_2 e_2^2 - c_3 e_3^2 - d_1 z_1^2 - d_3 z_3^2 \quad (28)$$

It is a negative definite function of the vector $(e_1, e_2, e_3, z_1, z_3)$. It follows that the system of the state vector $(e_1, e_2, e_3, z_1, z_3)$ has a equilibrium point globally asymptotically stable on the position $(e_1, e_2, e_3, z_1, z_3) = (0, 0, 0, 0, 0)$. That means in particular that the tracking errors (for the winding velocity and the upstream/downstream tensions) tend towards zero whatever the initial conditions.

4. SIMULATION AND RESULTS DESCUSSION

The performances of the regulator worked out in the preceding paragraph will be illustrated now by simulation. We will use the www model quoted in section 2. We took as references V_2^{ref}, T_1^{ref} and T_3^{ref}, which respectively have as a value 29m/s, 100N and 120N.

The values chosen for the regulator parameters of of Backstepping during simulations are:

$(c_1, c_2, c_3, d_1, d_3) = (60, 32, 78, 110, 224)$.

The simulations results are presented by figures 2, 3 and 4. We note that the continuation objectives are achieved, as well for velocity as for the upstream and downstream tensions. These results show that the continuation errors corresponding to the each parameter are cancelled after 30 seconds for the tensions and 15 seconds for the web velocity.
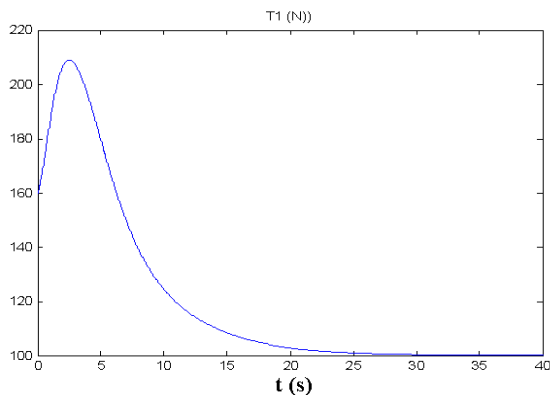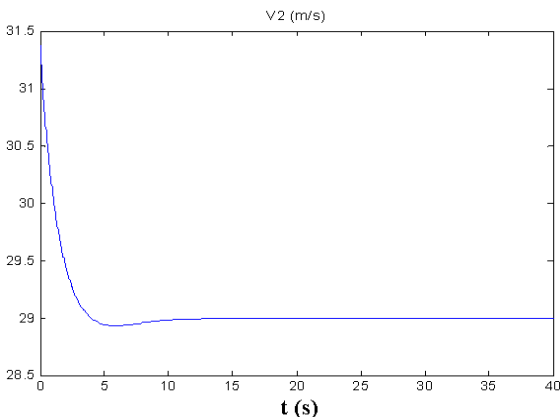

Figure 2: Upstream tension control by Backstepping.

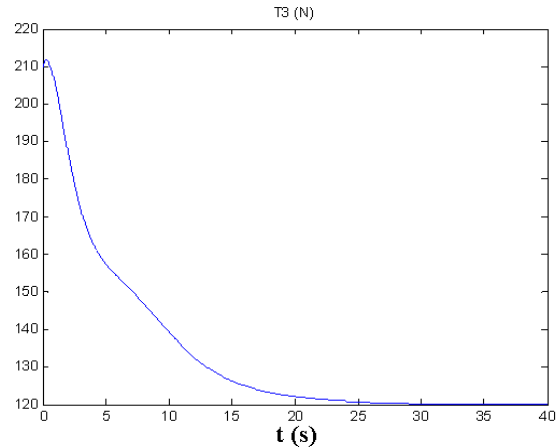

Figure 3: Web velocity control by Backstepping.

Figure 4: Downstream tension control by Backstepping.

5. CONCLUSION

In this paper, we approached the control problem of the web velocity and the web tensions of a web winding system, using a regulator worked out by the Backstepping technique. The synthesis rested on the standard model, which holds account owing to the fact that all the state variables were supposed to be available. We formally established that the closed loops made up of this regulator and the model from which it is resulting, are overall asymptotically stable. Moreover, the regulator by Backstepping asymptotically ensures a perfect trajectories tracking of the web velocity and tensions references. This result was confirmed by simulation way.

REFERENCES

Kokotovic, P., Arcak, M., 2001. Constructive nonlinear control:A historical perspective. *Automatica*, Vol. 37, No. 5, pp. 637–662.

Krstic M., Kanellakopulos, I., Kokotovic, P. V., 1995. *Nonlinear and Adaptive Control Design*. Wiley, New York.

La Salle, J., Lefschetz, S., 1961. *Stability by Liapunov's direct method with applications*. Academic Press, New York.

Rabbah, N., Bensassi, B., 2008. Control Lyapunov Function for a Nonlinear web winding system. *International Conference on Control and Applications (CA 2008)*. May 26-28, Quebec, Canada.

Rabbah, N., Bensassi, B., 2006. Modélisation et simulation d'un système de laminage à froid quarto réversible. *Rencontre Nationale des Jeunes Chercheurs de Physique (RNJCP06)*, 18-19 décembre, Casablanca, Morocco.

Rabbah, N., Bensassi, B., 2007a. Modelling and simulation of web winding system of a reversible rolling mill. *International Conference on Applied Simulation and Modelling (ASM2007)*, pp. 474-478. August 29-31, Palma de Mallorca, Spain.

Rabbah, N., Bensassi, B., 2007b. Commande d'un système d'entraînement de bande. *Manifestation*

des Jeunes Chercheurs en Sciences et Technologies de l'Information et de la Communication (MajecSTIC 2007). pp. 251-255, 29, 30 et 31 octobre, Caen- France.

Schmitz, A., Herman, J.C., 1995. Modélisation du laminage à froid des aciers. Etude des aspects métallurgiques. Analyse par la méthode des tranches. *CRM Faculté Polytechnique de Mons*, pp. 37-43, May 16-17, Belgique.

Sontag, E. D., Sussmann, H. J., 1989. Further comments on the stabilizability on the angular velocity of a rigid body. *Systems and Control Letters,* pp. 213-217.

Ueno, T., Sorao, K., 2004. Improvement of Accuracy for Gauge and Elongation Control by Dynamic Process Control Simulator. *Proceeding of the IEEE International Conference on Control Applications*, pp. 896- 901. September 2-4, Taipei, Taiwan.

AUTHORS BIOGRAPHY

Bahloul Bensassi (bahloul_bensassi@yahoo.fr) is Professor of the higher education, researcher teacher in Faculty of sciences, University Hassan II Aïn-Chock, Casablanca-Morocco. He is responsible of professional Licence and of Logistic engineering Master, expert in Industrial engineering, expert of the public higher education, expert of the private higher education, and inspector of the private higher education.

Nabila Rabbah (nabila_rabbah@yahoo.fr) is currently a doctoral student in Faculty of Science, University Hassan II Aïn-Chock, Casablanca-Morocco. Her primary research interest is in automation of dynamic nonlinear systems and control.

SIMULATION STUDY OF AN AUTOMATED AIR CARGO TERMINAL

Bill KP Chan[a], Henry YK Lau[b], Steve KK Chan[c]

[a]Industrial and Manufacturing Systems Engineering Department, University of Hong Kong
[b]Industrial and Manufacturing Systems Engineering Department, University of Hong Kong
[c]Hong Kong Air Cargo Terminals Limited

[a]billchan@hku.hk, [b]hkylau@hku.hk, [c]steve.chan@hactl.com

ABSTRACT

SuperTerminal 1, the main terminal of Hong Kong Air Cargo Terminals Limited (Hactl) located at the south east corner of the Hong Kong International Airport, is a US$1 billion investment and it is the largest air cargo terminal of the busiest airport in the world. This 6-storey infrastructure consist a complex array of interconnected computer controlled material handling systems. This paper illustrates the use of simulation modeling as a planning tool for service improvement of this world-class automated air cargo terminal. The objective of the simulation is to substantiate the value of service improvement initiatives. The results from the simulation model were used to quantify and assess the impact before implementation of these service improvement initiatives. A simulation model of the existing operation was built and validated. Experiments were performed by transforming operational orders of typical days into input orders under assumptions of different "what-if" scenarios

Keywords: air cargo terminal, simulation, airport, material handling system

1. BACKGROUND

Hong Kong International Airport has been ranked as the busiest airport for international air cargo since 1996. In 2006, Hong Kong International Airport handled 3.6 million tonnes of freight in which 80 percent are handled by the Hong Kong Air Cargo Terminals Limited (Hactl). Hactl also serves over 90 airlines and 1,000 freight forwarders. At the heart of Hactl that handle the air cargos is the main terminal called SuperTerminal 1 (ST1) which basically is a giant 6-storey infrastructure consisting of interconnected automatic storage and retrieval systems, powered roller decks, automatic cargo hoists, transfer vehicles and other material handling systems centrally controlled by a computer system called Logistics Control System (LCS).

ST1 is a multi-storey automatic warehouse consists of 6 levels (Figure 1). Air cargos are loaded and unloaded at the truck docks located in the G/F and the 1/F. The second floor is a circulating conveyor system called the Unitized Cargo Distribution System (UCDS) for moving cargo units throughout the terminal. Workstations are located at the 3/F and the 4/F for cargo build-up and break-down. Empty containers are stored at the top level.

ST1 is divided into three key operational areas namely landside service area (Landside), bulk cargo service area (BCS), and airside service area (Airside). Landside is responsible for the cargo release and cargo acceptance to shippers, consignee, and transportation companies in accordance with the service instruction from Airlines, shippers, consignees or transportation companies. Landside is also responsible to the truck management in truck park and truck dock. BCS is responsible for the cargo break-down, build-up and consolidation in accordance with airline service instruction. Airside is responsible for the ULD transfer-in and transfer-out operation in accordance with stipulate service requirement. Airside is also responsible for the empty ULD management and overall CSS management for frontline.

The ST1 terminal capacity is over 3.5 million tones per annum and it operates with two giant cargo handling systems: the Container Storage System (CSS) and the Bulk Storage System (BSS). BSS is located in the middle of ST1. Air cargos, not containerized in unit load devices (ULD), are stored in BSS. In addition, there is also a system of automated guided vehicles (AGV), called the Bulk Cargo Distribution System (BCDS), located as ring structures around the entries and exits of BSS. The BCDS is tightly integrated to the BSS to handle import and export air cargos moving in and out of the BSS. Besides, loaded and empty ULDs are handled in CSS. The main focus of this paper is on the simulation study of the CSS.

Located on both the east and west wings of the ST1, CSS is a cargo handling facility offering storage of ULDs and circulating them across the terminal. CSS consists of 12 computer controlled automatic stacker cranes (SC) which service a matrix of compartment spaces with more than 3,500 storage locations that are divided into 12 zones. These zones are located in two wings (east and west wings). Each wing has 4 sets of conveyors behind the outer CSS compartment served as by-pass links to connect CSS zones of the same wing. The east and west wings are separated by the BSS, the

Landside and the BCS. Landside is located on the G/F and 1/F. There are about 200 truck docks in the Landside for cargos acceptance or release. BCS is located on the 3/F and 4/F where over 300 workstations are installed. The workstations are ULD holding places for cargo build-up and break-down. The CSS is connected to the workstations by 72 link bridges with 40 Automated Transfer Vehicles (ATV) as the interfacing equipment. There is also a network of interconnecting conveyors, hoists and transfer vehicles that connect all 12 zones together. The UCDS is the main conveyor structure that connect the east and west wings. Altogether, there are 24 cargo hoists (CH) located in east and west wings for transferring ULDs vertically across the CSS including 14 building hoists located near the center structure of ST1 and 10 CSS inner hoists located between the ULD storage compartments. There are also 4 CHs in the north wing for transporting perishable cargo units. Moreover, there are over 140 System Entry / Exit Points (SEP) located at different locations for direct man-machine interface for ULD storage, transfer and retrieval.

In order to assess the performance of various activities, processes, resources and equipment in ST1, the simulation team of the Industrial and Manufacturing Systems (IMSE) department of the University of Hong Kong (HKU) was commissioned by Hactl to undergo a simulation study on the CSS of ST1. The result of the study had been successfully delivered to Hactl in mid-2007. This paper describes how the simulation model can be used to help Hactl in substantiating the value of service improvement initiatives and quantifying the impacts under different "what-if" scenarios.

The use of simulation in facility planning and process improvement is not new to Hactl. Luk (1990) explained how Hactl use simulation to evaluate system features before final design and construction prior to the relocation of the Hong Kong International Airport from Kai Tak to Chap Lap Kok in 1998. Since ST1 of Hactl is such a complex and highly automated structure, ST1 has been a subject of a number of researches. Lau and Zhao (2006) used ST1 as a typical case in the development of an integrated scheduling methodology for automated cargo handling system. On (2005) proposed a dynamic routing strategy for automatic material handling system and the strategy was benchmarked with the existing routing strategy of ST1 using simulation. In general, because of the complexity of air cargo handling processes and the keen competition between airports, simulation is found to be a valuable tool to maximize performance and push down costs in real life applications. DeLorme et al. (1992) illustrated the use of simulation for evaluation and analysis of air cargo operations by describing the development of a simulation model for the Dallas/Fort Worth Airport cargo hub of AA Cargo. Nsakanda et al. (2004) used simulation to quantitatively evaluate and compare different policies, business practices and

processes within a given set of operational and business constraints before the new facility was put into operation.

This paper is organized as follows: a brief description of the air cargos handling in Hong Kong is described in section 2. The simulation model is discussed in section 3. Results and analysis are reported in section 4. At the end of this paper, the conclusion of the simulation study is presented.

2. OVERVIEW OF OPERATION IN HACTL

In general, the service operations in Hactl can be classified into three major types in all export, transshipment and import flow cycles. These operations are: prepacked cargo handling, bulk cargo handling and perishable cargo handling.

2.1. Export Flow

Figure 2 shows the flow processes in export cargo handling.

2.1.1. Step 1: Cargo Acceptance

Prepacked cargos are accepted at different locations. At the ground floor of ST1, acceptance of prepacked cargo is performed on the Prepacked Cargo Handling Centers and the Perishable Cargo Handling Centers. Prepacked cargo can also be accepted at the Twenty-foot ULD Storage Center and the Hactl ULD Center located at the southern end of the Express Center and the northern end of the East Truck Park respectively.

The acceptance of bulk cargos are performed on both ground floor and first floor of ST1. After cargo acceptance on first floor, the bulk cargos are stored into BSS. For bulk cargo accepted on ground floor, the bulk cargos are stowed into the CSS.

Export perishable cargos are accepted at the Perishable Cargo Handling Center located at the north wing of ST1 and on both G/F and 1/F, with direct interface with the Airside.

2.1.2. Step 2: Cargo Processing

After cargo acceptance, prepacked cargos will be moved through powered roller decks, cargo hoists, ATVs and SCs and then stored in CSS prior to outbound dispatch.

Loose bulk cargos will be loaded onto boxes/bins at the time of acceptance. Loaded boxes/bins will be moved into the BCDS. The loaded boxes/bins will then be stored in BSS through the BCDS. According to each pre-manifest from an airline, related bulk shipments are retrieved from BSS and the cargos are built up with appropriate ULDs on the workstation floors. After cargo build-up, cargo units are then transferred to CSS for temporary storage or directly transported to Airside for immediate outbound dispatch

Perishable cargo will be directly transferred to the cool room after acceptance for temporary storage or immediately outbound dispatched to Airside.

Container Storage System (CSS)

Box Storage System (BSS)

The Bulk Cargo Distribution System transports the boxes/bins between the Box Storage System, Customs Examination Hall, workstations, truck docks, and skid cargo storage area

4/F

Workstations on 3/F & 4/F for cargo build-up and break-down

Pre-Clearance Team Office

Overhead bridges link the Container Storage System with the main terminal building

3/F

Unitised Cargo Refrigeration Centre

Empty unit storage on 2/F

Bulk Cargo Refrigeration Centre

2/F

Perishable Cargo Handling Centre

Truck docks for bulk cargo release and acceptance

Food & Environmental Hygiene Department

Customs Examination Hall

Cargo Survey Room

Agriculture, Fisheries and Conservation Department

Livestock Room

1/F

Onboard Courier Centre

Unitised Cargo Control Office

Truck docks for prepacked and unitised cargo release and acceptance

Livestock Room

Mini-Shipment Centre

Dangerous Goods Store and Radioactive Goods Store

G/F

Tenant Restricted Area
2/F - 4/F are Tenant Restricted Areas
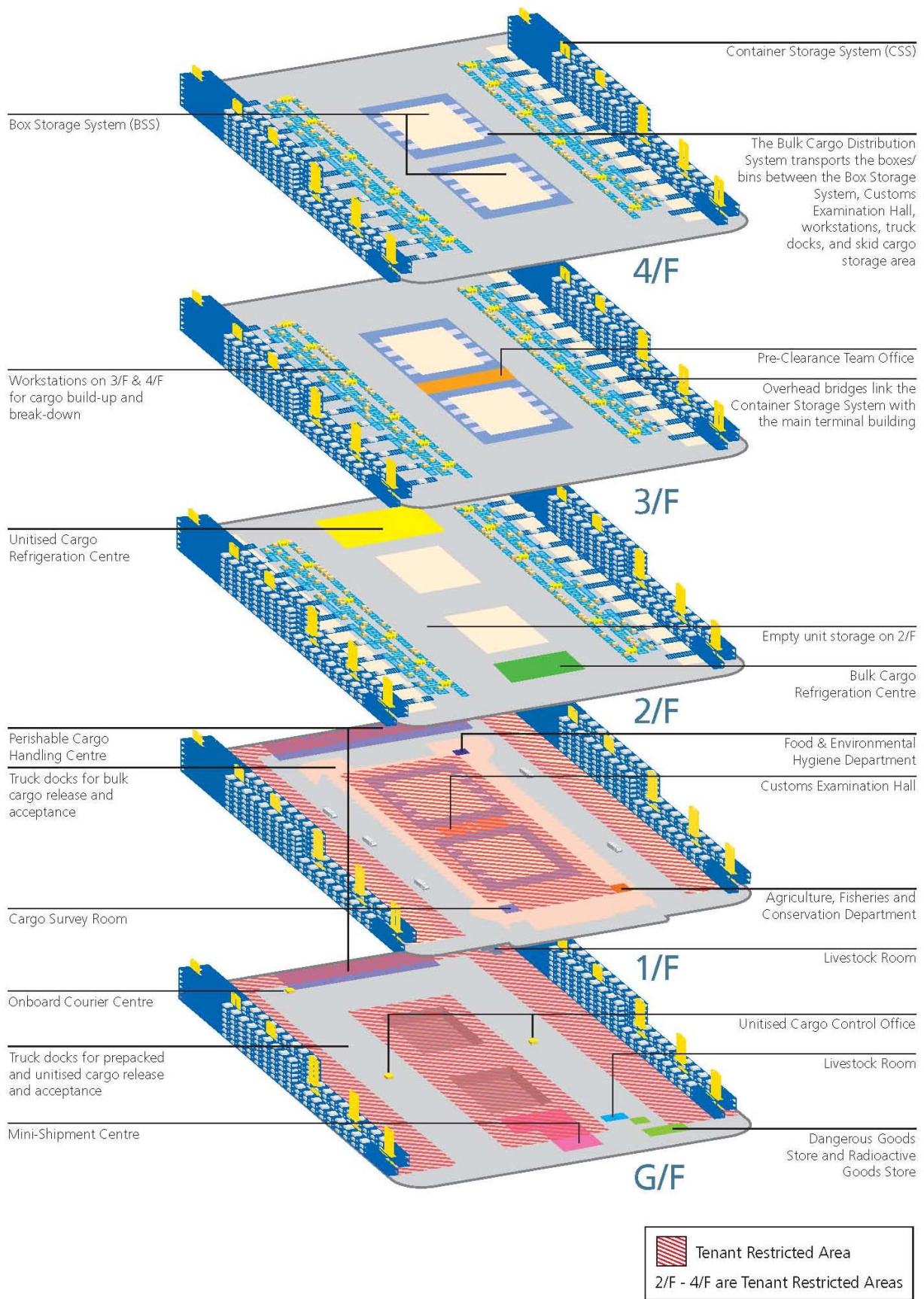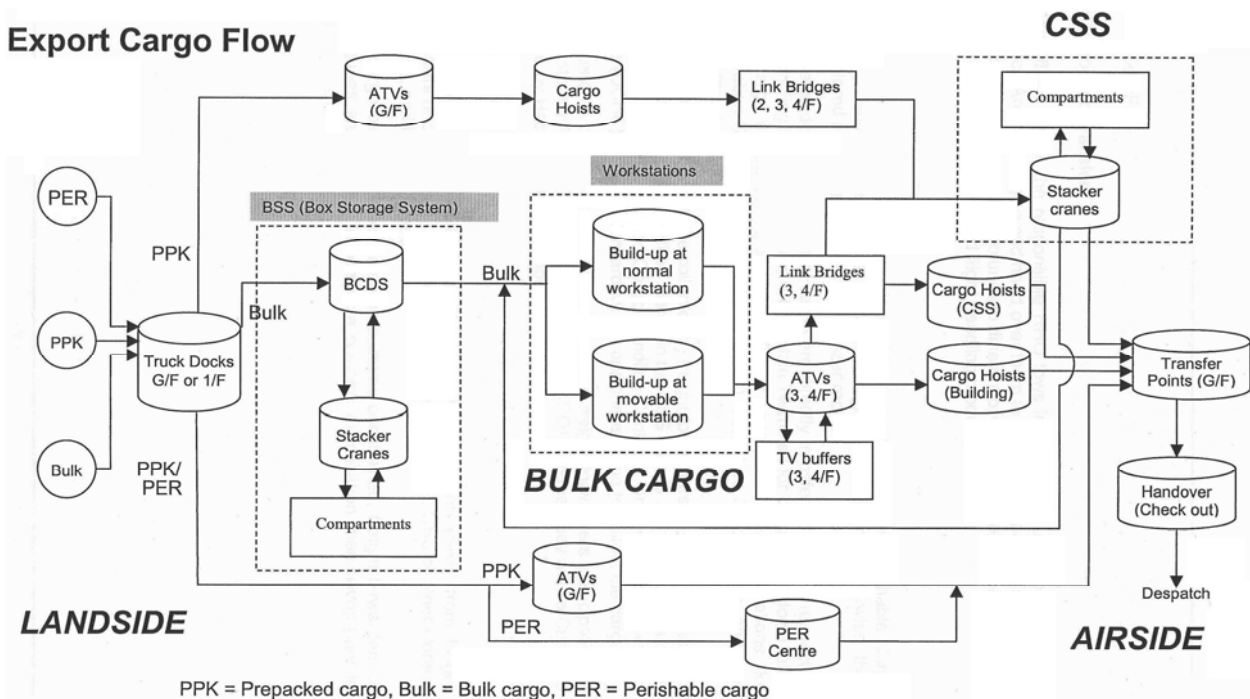
Figure 1 ST1's Cargo Handling Area (Hactl, 2007)
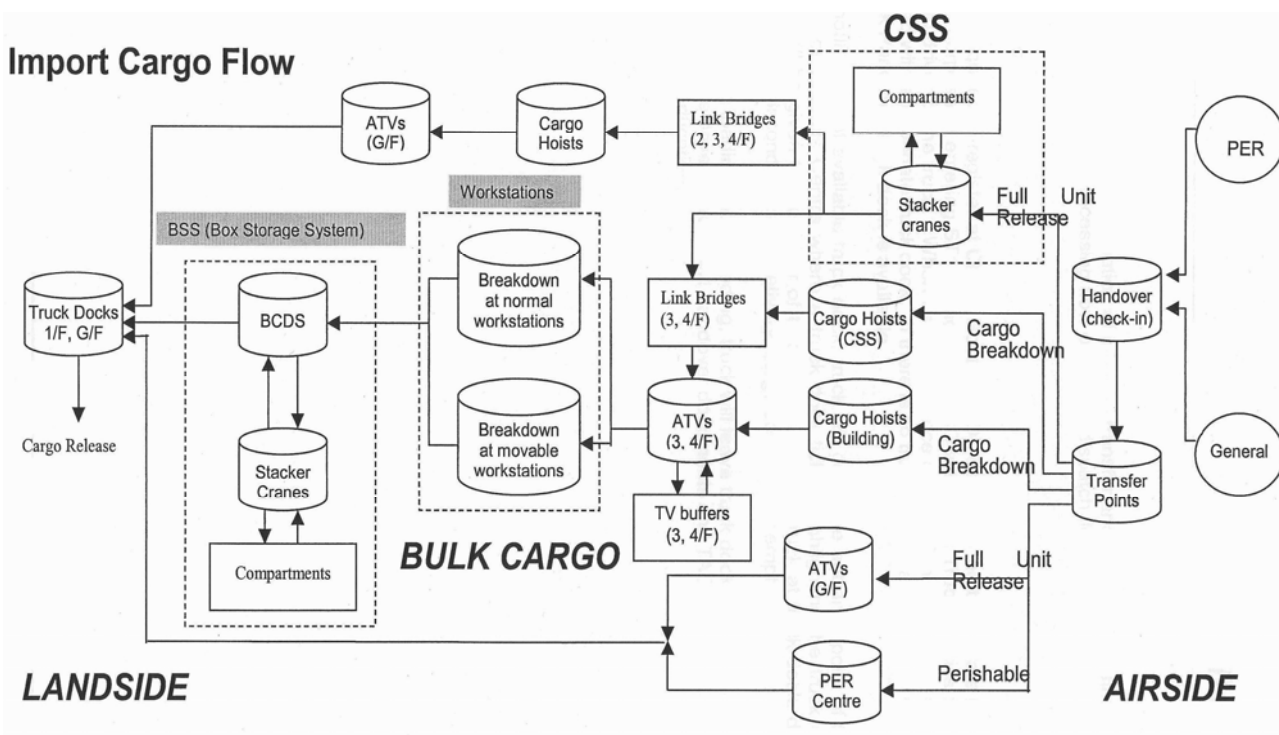
254

Figure 2: Export Cargo Flow (Source: Hactl)

Figure 3: Import Cargo Flow (Source: Hactl)

2.1.3. Step 3: Outbound Dispatch

All export cargos are transported to the Airside Interface Area some time before the Scheduled Times of Departure (STD) of their respective export flights assigned to them according to the pre-agreed airside service standards. The cargo units are then picked up at the CSS outbound transfer points and towed on-dolly to the dolly staging lanes by the Airside staff. Thereafter, the export cargos are handed over to the responsible Ramp Handling Operator (RHO) to complete the outbound dispatch process.

2.2. Import Flow

Figure 3 shows the flow processes in import cargo handling

2.2.1. Step 1: Inbound Transfer

Upon flight arrival, the responsible RHO tows import cargo units to the Airside Interface Area for staging and hand over them to Hactl. These inbound cargo units are then transported to different operational areas according to their respective cargo natures and flight handling modules.

2.2.2. Step 2: Cargo Processing

For perishable cargo, they are released to consignees at the Perishable Cargo Handling Center immediately after cargo handover with RHOs. Otherwise, they will be stored inside the cool room.

Containerized / palletized general cargos are directly transferred to their respective flight handling modules (3/F and 4/F of ST1) for cargo breakdown. After the breakdown process, cargos will be stored in BSS, or appropriate storage locations like Refrigeration Center, Dangerous Goods (DG) room, Live Stock Room, and floor goods area, depending on their cargo nature.

For those general cargos to be collected by freight forwarders or consignees in a full unit or unitized form, they will be stored in the CSS prior to cargo release, and are transported to the Prepacked Cargo Handling Center or Unitized Cargo Handling Center (UCHC) when the forwarders or consignees come to collect the cargo.

2.2.3. Step 3: Cargo Release

Upon receipt of duly authorized Shipment Release Form for freight forwarders or consignees, cargo will be retrieved from BSS, CSS or other storage locations and handled over to them.

If the cargo being collected remains customs constrained, freight forwarders or consignees are required to complete customs examination themselves prior to leaving the terminal.

2.3. Transshipment Flow

Transshipment flow is a combination of flow processes. It combines the inbound transfer and cargo processing of the Import Flow with cargo processing flow outbound dispatch flow in Export Flow

3. THE SIMULATION MODEL

3.1. Key factors to success

There are a number of factors that constitute the success of a simulation project (Shannon, 1998.). In this particular project described in this paper, there are three key conditions needed to be satisfied before starting the construction of the simulation model of such a complex structure. Fulfilling these conditions is the major reason contributing to the success of the project.

The first condition is the readiness of data. Since the ST1 operation is highly computerized, all cargo handling activities in the ST1 are logged automatically. The availability of sufficient amount of accurate and relevant data reduces a lot of uncertainty in model building.

Secondly, the simulation team members' understanding on the ST1 cargo handling processes is also of vital importance to the success of the project. The more the team members know about the operation of the system, the better the model will be built. Since the IMSE department of HKU values highly on the connections with the logistics industry in Hong Kong, an informal relationship has been built up between the simulation team and Hactl even before Hactl's request of the simulation study. The relationship is developed through study tours, informal visits and trial studies. Thus, the simulation team has already had some preliminary knowledge on ST1's operational processes beforehand.

Most importantly, the simulation project is supported by Hactl's management. The management realizes the benefits of simulation study and the value of alternative source of information for decision making before a major investment on service improvement or new facilities. Through meetings, demonstrations of past experiences of the simulation team and the teams' expatiation on their understanding on ST1's operation, confidence and trust on the team has been developed. The support of Hactl's management greatly facilitates the communication of all parties and personnel involved in this project.

3.2. Model Development

The development of the simulation model mainly follows the procedures described in "the steps in a simulation study" of Banks and Carson (2001). Before the project formally started, objective of the simulation was clarified and the boundary, scope and level of details of the model were defined through meetings with all related parties. A formal proposal was submitted to the management of Hactl for approval. The proposal included the objective of the study, a detail planning and development schedule of the project, project cost, assumptions made in the model, well defined model boundary, scope and level of details, expected output

from the model and the information required in building and running the model such as historical data, schematics, process diagrams, control logics of conveyors, AGV, cargo hoists,…etc.. A simplified model prototype had also been built to demonstrate the user interface of input data, the visualization control of the model animation and the graphical presentation of the resulting simulation output data.

After the proposal was accepted, a simulation team was formally formed. The simulation team included a project manager whose main responsibility was to oversee the progress of the project and coordinate the team members in the course of model development. There was also a representative from Hactl in the team to collect all information from Hactl. He also arranged on-site visits or interviews with Hactl's operation staff for the simulation team. Feedback from the representative was also important to ensure the project was moving in a direction align with the management expectation in the progress of the project. Besides, there were altogether four analysts who cooperate within themselves to perform model design, model building, data modeling, programming, verification, validation and scenario analyses.

3.3. Objective, boundary and scope

The objective of the simulation is to study the throughput, utilization and cycle time of the processes, resources and equipment of the CSS. In CSS, either loaded or empty ULDs or unitized general cargo units are processed. Individual air cargos will not be handled. Thus, the smallest unit of entity under this study is defined to be the containers but not the individual cargos.

The boundary of the system can be clearly defined by the SEPs. ULDs are moved into the system or moved out from the system through any one of these SEPs. The SEPs are the interface between the CSS and the airside / landside. ULDs moving into the CSS from airside / landside or moving out to the airside / landside must pass through one of the SEPs. Within CSS, ULDs are either transporting on a material handling facility, such as a powered roller decks or a cargo hoist, or stored on a TV buffer or in a compartment storage location.

3.4. Model Conceptualization

Simulation modeling is both a science and an art (Shannon, 1998) and the essence of the art of modeling is abstraction and simplification. Including too much details of this complex system into the model makes the model not only too inefficient to run but also too expensive to build. It may take a longer time to construct the model and may not be able to deliver the necessary information to the management before the dead line. On the contrary, oversimplified model may have too many assumptions which may not of value to the decision making of the management. The challenges are to identify the essential parts of the problem, choose the right level of details and design the model around the problem accordingly.

The main theme of the study is to analyze the traffic of the CSS which is basically a huge material handling system. Thus, choosing a commercially available simulation software package that is efficient in modeling material handling system is important. Moreover, the traffic of the CSS can be viewed as the aggregation of movement of containers or unitized cargos moving through the system. Thus, the major input to the model should be the generation of transportation orders of the ULDs in the system. However, as described in later part of this section, ULDs can be originated from a large number of locations and they can also have a large number of destination. There are a large number of source / destination combinations and there are also a great number of routes that a ULD can move through. In designing the model, it is important to implement all the routes into the system without getting lost in the complexity.

The Pareto principle, which is also known as the 80-20 rule, states that, for many events, 80% of the effects come from only 20% of the causes. After analyzing the historical data of the CSS, it is found that the 80-20 rule is also true in the traffic of CSS. Thus, different routes with different source / destination combinations are classified into different priorities and the modeling effort is also allocated to different routes accordingly.

The processes to transport a ULD within the CSS, no matter it is an import, export or transshipment flow, can be categorized into four different order types namely storage, transfer, retrieval and reshuffling. Storage and retrieval orders are instructions for ULD storage into compartment and retrieval from compartment respectively. Transfer orders are instructions for sending ULD from one location to another location within the CSS structure outside the compartment storage locations. Reshuffle orders are instructions for sending ULD from one compartment to another compartment. Depending on the size of the ULDs, more than one ULD can be placed in a single compartment storage location. Therefore, a reshuffle order must be issued to relocate the front ULD before the ULD at the back of the same storage location can be retrieved. Although reshuffle orders are non-productive and should be minimized, this type of orders can be hardly avoided especially when the utilization of compartment is high.

Since Hactl is a highly automated material handling system, all issued orders are recorded by the LCS. These recorded orders represent faithfully the operational activities happened in the entire CSS. Therefore, the model is designed with its input similar to the format of an order file and generates loads based on the information of an order. There is an advantage in validation to design the input in this way. Since the order file for each day is readily available, only minor modification and conversion is needed before feeding into the model. Days of order files can be cascaded together to form a single input file and the resulting

behavior of the model can be used for comparison of the log files.

The complexity of the model mainly comes from the vast diversity of the routes traveled for different source / destination combination of the ULD transportation orders. The CSS contains altogether 14 zones, 6 for the east wing, 6 for the west wing and 2 for the north wing. Zones in the east or west wings are subdivided into 9 areas namely Landside (G/F), BCS (3/F), BCS (4/F), Airside (Building), Airside (Inner), Airside (Outer), 2/F, Compartment and Roof Top. Zones in the north wing are subdivided into 8 areas: PCHGK, PCHGU, PCH1K, PCH1U, airside, 2/F, BCS (3/F) and BCS (4/F). Orders can be generated from any area / zone as the source and any other area / zone as its destination. Each area contains a number of end units: workstations, TV buffers, compartment or SEPs. If individual locations are grouped into their corresponding areas, there are approximately (6x9x2+2x8)*(6x9x2+2x8) = 15376 different source / destination combinations. Furthermore, each source and destination combination may have more than one route. Also, there are about 2500 decision points scattered on the paths of different routes. When a ULD arrives at a decision point, the program in the model should be able to instruct the ULD to choose the correct branch based on its destination and other relevant conditions at the moment of the ULD arrival.

To tackle the complexity of the system, routing tables are used. These routing tables are designed to specify all the routing conditions at a decision point. After designing the control mechanism of the routing tables, simulation analysts can work separately in parallel to fill in the routing tables. Programmers can also focus on writing efficient algorithms to decode the routing tables and table look-up for the routing decisions. With the use of routing tables, the implementation of routing decision can be separated from coding. Also, routing decisions of different zone or area can be handled by different analysts. More workforces can then be put into the simulation team to shorten the duration of model development.

Since the CSS is centrally controlled by the LCS computer system, one way to implement the decision rules in the routing tables is to refer directly to the routing rules that have been programmed in the LCS. This approach requires an updated program documentation of the LCS. Reading program documentation is helpful for the simulation analysts to understand the system but relying on the documentation to build the simulation model is risky since the documentation may not be fully updated. The simulation analysts also needed to be granted the right to read the source codes of the LCS control program. However, this approach is not feasible simply because it is too inefficient to understand a program of such a large system by tracing the logics of the source codes written by somebody else.

Another way to determine the decision rules in the routing tables is to draw inferences of routing rules from the observation of ULD movements recorded in the log files and ignore what has been programmed in the LCS. In CSS, whenever an ULD moved into a device, such as a power roller deck or a cargo hoist, the information of the activity will be generated. The information includes the date and the time, the ULD identity number, the device's identity number …etc. The information are recorded, collected and stored in a daily log file. These daily log files faithfully record what has happened in the CSS. Simulation analysts analyze the log files and implement routing rules from observation on the log files without knowing how the LCS controlling the system in detail.

Although there are a large number of source / destination combinations of the ULD transportation orders, not all combinations have the same amount of traffic. It is found that the majority of the traffic is within the same zone and between two adjacent zones. These two types of traffic formed the highest portions in the whole CSS up to about 84% (Figure 4).
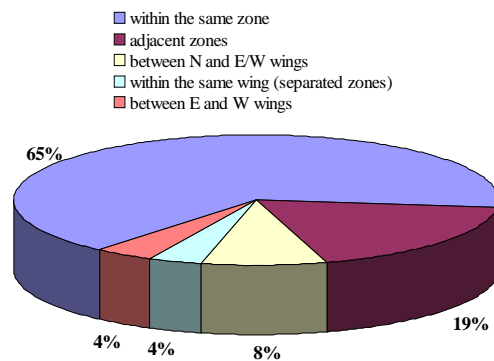
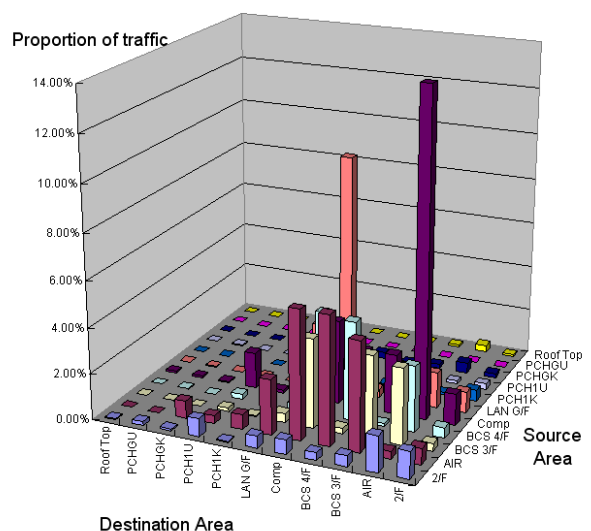

Figure 4: Traffic between different locations in CSS

Figure 5: Source/Destination Analysis

After investigating the log files, it is obvious that the simulation analysts should put more effort in the traffic within the same zone and the traffic between adjacent zones because these paths have the greatest traffic and hence have the largest impact to the system.

Traffic between the North and the East or West wings is the second priority. Then, the traffic within the same wing but between separated zones is the next. Although the East and the West wings had the largest number of source / destination combinations, ULD movements between these two wings had the least traffic. Thus, this should be the last one to be handled.

A better understanding of the traffic in CSS can be reviewed if further analyses into the area level. A sample source /destination analysis of a typical day is shown on Figure 5.

From these analyses, it is found that the highest traffic is the traffic from compartments to the airside and the traffic from the landside ground floor area to the compartments. The majority of the rest of the traffic is the traffic among the compartments, airside and the BCS on 3/F and 4/F.

Apparently, the traffic inside the CSS obeys the Pareto principle. Simulation analysts should focus on the major ULD transportation paths which constitute relatively smaller portions of the total possible paths but with the greatest impacts on the traffic of the system. Analysts should also set a priority of their focus on the rest of the paths according to the amount of traffic of the path shown from the analyses. It does not mean that paths with less traffic can be implemented with less accuracy. In fact, an accurate model depends on whether all paths have been implemented accurately. The advantage of this approach is to keep the analysts alert to which source / destination combinations of ULD movements have the greatest impact on the overall traffic of the system. Setting the implementation priority of the paths can facilitate the credibility on the model to be built up more quickly because there is a higher chance to show to the people involved that an 80% correct model can be constructed within only 20% development time according to the 80-20 rule.

3.5. Data Collection

As the ST1 is highly computerized, activities in the system are automatically recorded as log files. There is no shortage of data. An advantage of using automatically logged data is that the chance of human error is comparatively small. Even though the chance of error is not high, the received data from Hactl is always examined before it is used in the model. Since there is a lot of data available, data need to be carefully selected and filtered so as to reduce redundancy and to extract information from the data to the right level of data abstraction. Furthermore, filtering out unnecessary data reduces the total amount of data processed in the computer and, reduces the amount of i/o handling and hence lowers the amount of memory required, and thus, increases the speed of model running. The primary input to the model is the ULD transportation orders. For each order, only the order initialization time, ULD identification number, order type, source location and destination location in terms of area and zone are needed.

Besides the input ULD transportation orders which can be found in the log files, some parameters in the simulation model, for example, processes involve human handling, needed to be modeled as statistical distributions estimated from historical data. These processes include the times for building up or breaking down a certain type of container and the engagement times of ULDs when it is transferred into the system or out from the system at a SEP. Some parameters, for examples, the speeds of the conveyors, SCs, CHs and ATVs, are taken from the operation specifications and carefully verified with logged data of ULDs passing through these equipments.

3.6. Model Translation

The model was built using Applied Material's AutoMod. Microsoft Excel 2003 programmed with VB scripts was used as the user interface. The whole system was running in Window XP on a workstation with Dual Core 3.2GHz CPUs and 4G RAM. A single run simulating one typical day operation could be completed within 10 minutes.

Statistics are collected at every 1 hour time slot. The output statistics are presented in Excel file in chart and data format. Collected statistics includes the statistic of all kinds of equipment - Automatic Transfer Vehicles, Bypass Lane, Compartment, Link Bridge, Stacker Crane, Transfer Point, Transfer Vehicle Buffer, UCDS, and Workstation. The utilization, throughput and cycle time of these equipments are recorded and shown on an excel interface. Statistic can be displayed at various levels - for individual equipment or average value in different grouping, e.g., by level, by zone, by wing, or by the direction of container movement.

3.7. Verification and Validation

To avoid preconceived understanding of an analyst on a path in the system, another analyst that is not involved in the determination of the routing tables or the program codes of the path is involved in the verification. Artificially made order files for testing of particular paths are created for model verification. Animation is used to allow the analysts to observe how a container moving through the system. Using this technique, most of the mistakes can be discovered.

Animation is also a good tool for validation. After model construction and verification, the model is validated with the assistance of the operation experts from Hactl. With the help of animation, operation experts can easily point out the difference from the real system. For example, at the junctions of conveyors crossing each others in different directions where the traffic is heavy, there are some special handling and control rules which are overlooked at the early stage of model development. Operation experts can point out the different directly on the animation simply because the animation looks different from their daily experiences.

Historical data are also used for comparison with the model output in the validation process. Altogether,

18 days data are used. These data can be separated into two groups. Each group of data is data from consecutive days of at least one week. The two groups are separated by more than a months. There is no exceptional event happens in any one of these selected days, i.e., these days can represent operation of typical days. The aim of validation of a model is to determine whether the model is an accurate representation of the system for the simulation purpose, i.e., the analysis of the traffic in the system. Thus, throughputs of three most critical types of equipment (stacker cranes, cargo hoists and automatic transfer vehicles) are chosen for the comparison. Paired t-test with the level of significance be $\alpha = 0.05$ together with correlation coefficient are calculated for the comparison.

4. USE OF THE SIMULATION MODEL

The reason to build a simulation model for the CSS is to quantify the impact of service improvement initiatives. There are a number of questions that the management would like to get the answers from the model: What happen if some workloads, says all workloads from the workstations of zone D or all workloads to compartment of zone L are suspended for a certain period of time? What is the impact if some equipment or SEP is out of service? Will the operation be resumed to normal after fixing the equipment? Is it possible to maintain the service by temporarily transferring the workloads from one group of SEPs to another group of SEPs for a certain period of time in case there are equipment failures? What are the differences of the impacts if the above situation happens at different times of a day?

All these questions are related to the traffic inside the system or, more precisely, the throughputs, utilizations and cycle times of all equipment in the system. To find out the impacts on different scenarios, historical data files are analyzed and modeled statistically into an input order file that represents typical operation days. This model with typical operation days input order file is referred to be the base scenario. Experiments can be done by feeding input order files which are obtained by transforming the input order file of the base scenario according to the assumptions of different "what-if" scenarios. The result of the experiments will then be compared with the result of the base scenario.

5. CONCLUSION

In this paper, the use of simulation for service improvement of the largest air cargo terminal in Hong Kong Airport has been illustrated and the approach of the simulation modeling is described.

REFERENCES

Hactl, 2007. Hong Kong Air Cargo Terminals Limited SuperTerminal 1 Brochure. Hactl.

M. Luk, 1990. Hong Kong air cargo terminals to work in synch because of simulation applications. *Industrial Engineering,* vol. 11, pp. 42-5.

H. Y. K. Lau and Y. Zhao, 2006. Joint scheduling of material handling equipment in automated air cargo terminals. *Computers in Industry,* vol. 57, pp. 398-411.

W. On, 2005. *Dynamic Routing for Automated Material Handling Systems.* Thesis (MPhil). University of Hong Kong.

P. DeLorme, J. Procter, S. Swaminathan, and T. Tillinghast, 1992. Simulation of a combination carrier air cargo hub. *Proceedings of the 24th conference on Winter simulation,* pp. 1325-1331.

A. L. Nsakanda, M. Turcotte, and M. Diaby, 2004. Air cargo operations evaluation and analysis through simulation. *Proceedings of the 36th conference on Winter simulation,* pp. 1790-1798.

R. E. Shannon, 1998. Introduction to the Art and Science of Simulation. *Proceedings of the 1998 Winter Simulation Conference.*

J. Banks and J. S. Carson, 2001. *Discrete-event system simulation.* Prentice Hall Upper Saddle River, NJ.

AUTHORS BIOGRAPHY

Bill KP Chan is a Senior Research Assistant with the University of Hong Kong, and he is responsible for the research and development of simulation systems for analyzing complex logistics systems including the Central Asia Hub of DHL Aviation (HK) Limited, Terminal 4 simulation system for Hongkong International Terminals (HIT), and studies on the Hactl Cargo Handling System.

Henry YK Lau received a B.A. degree in Engineering Science and D.Phil. in Robotics from the University of Oxford, England in 1991. Prior to joining The University of Hong Kong as an Associate Professor, Dr. Lau was a College Lecturer at Brasenose College, Oxford and has been working in industry as a senior system engineer with AEA Technology plc., UK working on projects involving system integration and telerobotics systems for the nuclear industry. Dr. Lau's research interest includes intelligent automation, automated warehousing, and interactive virtual reality. In addition, Dr. Lau is also interested in the design and evaluation of automated material handling systems such as automated warehouses using simulation and virtual reality techniques.

Steve KK Chan is an Assistant Manager in the Service Improvement Department of the Hong Kong Air Cargo Terminals Limited (Hactl) at the Hong Kong International Airport, in which the company operates the world's largest single automated air cargo terminal – the SuperTerminal 1.

A SEQUENTIAL HEURISTIC PROGRAMMING APPROACH FOR A CORRUGATED BOX FACTORY: TRADEOFF BETWEEN SETUP COSTS AND TRIM WASTE

Pierrette P. Zouein[a], Jessica Diab[b]

[a] [b] Department of Industrial and Mechanical Engineering, Lebanese American University

[a]pzouein@lau.edu.lb, [b]Jessica.diab@lau.edu.lb

ABSTRACT

This paper considers a corrugator trim problem for a cardboard boxes manufacturing plant that produces boxes of various sizes for a finished goods inventory that services known customer demand. We present a suboptimal three-step procedure that considers minimizing both trim waste cost and setup time cost where setup cost is measured as the loss of production resulting from stock rolls size changes. The procedure leads to the least-cost method of combining customer orders on the corrugator and the optimal corrugator width to use over an entire shift. The proposed method was motivated by our analysis of the day-to-day scheduling of the corrugator at the UniPaK facility one of the leading manufacturer of cardboard boxes in Lebanon. This paper concludes with an application of the proposed method to derive a lower cost corrugator schedule at UniPak.

Keywords: Corrugator scheduling, heuristic, decision support system

1. INTRODUCTION

The corrugator trim problem is defined in the literature as the problem of determining the least-cost method of combining customer orders on a corrugator where one of the major costs to avoid is waste or excess trim lost from the material used.

The corrugator trim problem is a well studied problem in the literature. There have been many attempts to solve this problem using computer models, heuristics and optimal methods (Haessler and Talbot 1983). Early work was done by Eismann (1957) and Gilmore and Gomory (1961) which proposed linear programming formulations for the general trim problem in the paper industry. It was recognized later that because of the nature of the production process, the corrugator problem could not be completely modeled by means of linear programming which led to the development of sequential heuristic procedures. Viswanthan and Bagchi (1993) developed a best-first tree search algorithm to solve a constrained two-dimensional cutting stock problem where constraints are set on the number of sheets of given dimensions to be cut using only orthogonal guillotine cuts only.

Minimizing trim loss is only one of several major concerns that arise in drawing the optimal corrugator schedule. Other concerns are corrugator width utilization, cutting pattern changes (order changes), avoidance of split orders, and shutdowns costs. The problem becomes far more complex if the corrugator scheduling problem is integrated with other problems of the multi-stage production process, machine failures and unpredictability of customer behavior to ensure "on-time and in-full" deliveries of customer orders (Darley and Sanders 2004).

Krichagina et. al (1998) considered the cutting-stock problem subject to random customer demand and where the objective is to minimize long run expected average costs related to paper waste, shutdowns, backordering, and holding finished good inventory. They used a 2-step procedure with a linear programming model in the first step and a Brownian control in the second step to generate a suboptimal solution to the problem. Simplifying assumptions such as aggregating machines and dedicating machines to the production of a single grade single color papers, and not explicitly modeling shutdown and startup times were used to be able to find a good solution to the problem.

Given the complexity of the corrugator scheduling problem, it would be impossible to make optimal decisions that will achieve all the desired objectives. Thus a hierarchical heuristic approach has been typically adopted to solve the corrugator scheduling problem and this by decomposing it into smaller problems that are solved sequentially. This is why the corrugator trim problem is still for the most part solved manually.

The approach proposed in this paper was motivated by our observations at UnipaK, one of the largest cardboard manufacturers in Lebanon and UnipaK's interest in developing a method for scheduling jobs on the corrugator that would minimize both trim waste cost and setup cost resulting form roll width and cutting pattern changes on the corrugator. More specifically, this paper looks at the tradeoff between minimizing trim loss and maximizing roll width utilization in order to minimize waste in material and setup times incurred by changing the roll width and the cutting patterns at every order run. The approach proposed in this paper is a sequential heuristic

programming one where a 3-step sequential procedure is used to draw an optimal daily schedule for UniPak.

2. PRACTICAL CONSIDERATIONS IN SCHEDULING JOBS ON CORRUGATORS

The manufacture of corrugated cardboard boxes involves fabricating a continuous strip of corrugated board then cutting it into sheets of customer-specified dimensions. The corrugator forms a linerboard strip into a fluted shape then sandwiches it between two liners to produce a single-wall board. There are a number of common flute styles. The corrugated strip is next passed over a long set of rollers to allow sufficient time for drying. It is here that the first trim waste is incurred because the board edges are rough and irregular, about 1 cm is removed from each side of the strip. Then the corrugator board is cut into smaller strips corresponding to the specified sheet width, and cut-off to yield the sheet length. Cut-off knives and slitting knives are used to make horizontal cuts and vertical cuts respectively. Most trim waste occurs during the slitting/cut-off stage. The amount of waste is determined by the width of the corrugated strip being produced. This is why linerboard rolls exist in a number of different sizes. Changing from a narrower roll width to a larger roll width normally slows down the machine.

Corrugated boxes are bulky and subject to weather damage, hence the manufacturer would want to have low inventory levels and frequent production runs to ensure on-time delivery. To this end, the production scheduler typically goes through the following steps to schedule customer orders on the corrugator in order to minimize excess trim waste.

- *Selection and classification of Jobs:* the planner divides and arranges the factory tickets received from the customer service department first based on the fluting type of each order and on the liner (paper) type and color. The scheduler then arranges the different sets of factory tickets in ascending order of the different layers GSM (gram per square meter) and not the overall GSM and further group them into sets of orders that have the same GSM requirements within a 5% tolerance (difference). It is a common practice in the corrugated cardboard industry to quality-upgrade orders for practical considerations.
- *Combining Jobs:* the planner combines jobs within a group that can be processed simultaneously on the corrugator in order to minimize trim waste. In combining jobs, the planner has to determine the roll width based on the number of sheets that can be produced for each job per horizontal cut. The roll length is determined later based on the total number of sheets to be produced for each job to satisfy the order quantity. The planner can increment or decrement the order quantity of any job in order to adjust the total length of roll needed to meet the demand of both jobs in a combination that minimizes linear trim waste. This is another common practice in the corrugated cardboard industry where orders specify the quantity with overrun and underrun tolerances, typically a 10% tolerance. Also in combining jobs, the planner has to accommodate many practical considerations in combining jobs. For example, a maximum of four sheets can be generated per cut on one cutoff and a maximum of six to seven sheets can be generated as a total on both cutoffs. This is important to control serious bottleneck problem down the line at the finishing stage.
- *Sequencing Jobs on the corrugator:* the last step is to schedule the jobs on the corrugator based on the availability of the finishing machines and the due date priority of the orders.

Setup times are incurred when changing rolls and in axle changeovers. Setup times vary with the type of change; from our observation of Unipak operations, an average 12 minutes are needed to change the linerboard roll width and 10 minutes to change the fluting type to the next thicker or thinner fluting type. As for the triplex or triple axle changeovers, 5 minutes are needed on average to fix and set the next order cutting pattern. In this paper an average of 12 minutes setup time is used for roll-width changes including triplex changeovers if any.

It has also been noted that the corrugator speed varies between a maximum of 160 Mpm (Meters/min) and a minimum of 60 Mpm depending on the order length. The smaller the order length is, the slower the corrugator speed. In this paper and for the purpose of assessing the savings in the proposed scheduling approach, we shall assume an average corrugator speed of 100 Mpm. Consequently the money value of the unit setup time could be estimated by multiplying the corrugator speed by the product of the setup time required and the unit cost of paper which is assumed in this paper to be 0.07$ per meter

3. CORRUGATOR SCHEDULING APPROACH FOR MINIMIZING WASTE AND SETUP TIME

This section presents a three-step suboptimal procedure for drawing the daily corrugator schedule. The procedure returns the least-cost combination of orders to be scheduled on the corrugator over a given shift along with the optimal corrugator roll width to be used to produce all customer orders scheduled for production over that shift at minimal trim loss. Fixing the roll width over the whole shift will reduce setup time incurred by changing roll width for different order runs. The output is a roll width used over an entire production shift and the optimal combination of

customer orders that will be processed during that shift that minimizes trim loss.

The following is a sequential heuristic programming approach for scheduling jobs on a corrugator while minimizing trim loss and maximizing roll width utilization.

1. Solve the *Job Selection problem* to identify the set *S* of all possible pairs of jobs (i, j) that can be combined based on paper type, color, fluting and GSM requirements.

2. Fix the roll width *k* and solve the *Job Matching problem* to find, for each roll width and for each pair of jobs (i, j) in *S*, the optimal number of horizontal and vertical cuts that minimize total trim waste W_{ijk}.

3. For each roll width *k*, solve the *Roll Width Optimization problem* to find the optimal combination of customer orders that minimizes trim waste for roll width *k*, W_k. Then select the best roll width that minimizes geometric waste and setup time based on equivalent dollar values.

3.1. Step 1: Solve Job Selection Problem

The corrugator under consideration has 2 cutoffs and thus can process 2 different orders at the same time. Jobs are first classified as primary or secondary depending on their due date. Primary jobs are jobs that should be produced now in order to meet their delivery date. Secondary jobs are those that can be run along a primary job in order to minimize trim waste. Orders are paired if they can be processed simultaneously on the corrugator. A pair of orders is formed by combining a primary job with another primary or secondary job provided that the second job has:

- The same fluting type and paper color as the first job
- A gram per square meter (GSM) requirement within 5% of the first job GSM requirement.

Figure 1 shows the algorithm used for solving the *Job Selection* problem. The Job Selection problem is solved on Excel. The program returns the set of paired orders *S* that can be processed simultaneously, where $S = \{(i, j); i \in POrders; j \in AOrders; P_{ij} = 1\}$. Each pair (i, j) is a possible solution for the Job Matching problem.

3.2. Step 2: Solve Job Matching Problem

Given *S* the set of paired orders generated in step 1 and *R* the set of roll widths available, where *R* for the case of Unipak is *R* = {1.8, 1.9, 2, 2.05, 2.1, 2.10, 2.15, 2.2} in meters, the Job Matching problem consists of finding, for a set roll width *k*, the optimal number of vertical and horizontal cuts for each combination of

jobs (i, j) in *S*, that minimizes the geometric trim waste W_{ijk}.

```
Job_Selection(POrders, AOrders)
{
    • Get number (Np) of primary orders (POrders),
      and number (N) of all orders (AOrders).
      POrders is a subset of AOrders. AOrders is
      ordered by listing POrders first.
    • Get fluting requirements (F) of AOrders
    • Get paper color (C) of AOrders
    • Get GSM requirements (Gsm) of AOrders
    For i = 1 to Np do
    {
        For (j = i+1 to N; j++)
        {
          if  OR {(F_i ≠ F_j);(C_i ≠ C_j)}
            then P_ij = 0
            else if

              ABS( (gsm_j − gsm_i) / Max(gsm_j, gsm_i) ) ≥ 0.05

              then P_ij = 0
              Else P_ij = 1
        }
    }
}
```
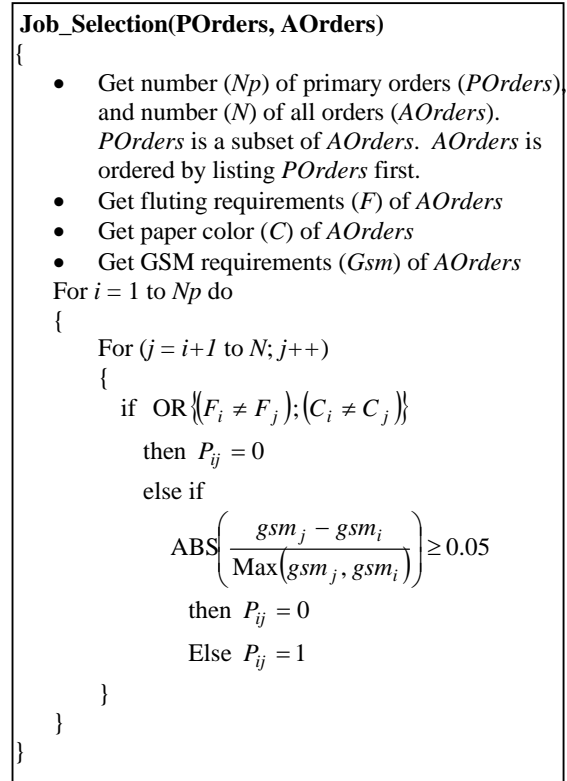
Figure 1: Job_Selection Algorithm

There are a number of practical considerations and machine limitations that should be observed in qualifying paired jobs in *S* for further consideration. Some of these practical considerations and limitations are

- the total width required for combining the two orders should not exceed the corrugator roll width size
- the two orders combined would need the same run length with a ±10% margin.
- the total number of sheets produced on the two cut-offs should not exceed seven sheets for the case considered in this paper
- the number of sheets produced per cut on each of the upper and lower cut-off should not exceed four sheets also for the case considered in this paper.

Figure 2 shows the algorithm used for solving the *Job Matching* problem. The algorithm returns for each roll width *k* in *R* the set S_k of paired jobs (i, j) from *S* that can be processed together on the corrugator using a roll width *k* and the minimum trim waste W_{ijk} resulting from running them together on the corrugator. Note that if the two jobs in a pair cannot be processed

simultaneously on the corrugator for a given roll width because of roll width limitations and thus the number of vertical cuts for either one of the 2 jobs is zero, the pair (i, j) is removed from the set S and thus from further consideration.

Job_Matching (S, R, S$_k$)
{
  Forall $k \in R$, do
  {
    Forall $(i, j) \in S$, do
    {
    Solve **nb_verticalcuts** to find optimal number of vertical cuts per horizontal cut for order i (V_{ij}) and

    order j (V_{ji}) that minimize linear trim waste

    If $V_{ij} \neq 0$ and $V_{ji} \neq 0$

        then
        • Solve nb_horizontalcuts to find total length of roll needed to satisfy demand of primary order i while minimizing geometric trim waste W_{ijk} for the

          combination (i, j, k)
        • Append (i, j, W_{ijk}) to the set S_k
    }
    Forall $i \in POrders$, do
    {
    • Compute number of vertical cuts per horizontal cut for order i by dividing k with the sheet width of order i
    • Compute total length of roll needed to satisfy demand of order i by dividing order quantity of i by the number of vertical cuts found
    • Compute geometric trim waste W_{iik} for the combination (i, i, k)
    • Append (i, i, W_{iik}) to the set S_k
    }
  }
}

Figure 2: Job_Matching Algorithm

3.2.1. Get Optimal Number of Vertical Cuts Per Horizontal Cut (nb_vertical cuts)

Given a roll width k and a given combination of jobs (i, j), let W_i, L_i, and D_i be the width, length, and demand in number of sheets for order i. Similarly let W_j, L_j, and D_j represent the requirements for order j. The following program finds the optimal number of sheets for orders i (V_{ij}) and j (V_{ji}) that can be generated per horizontal cut in order to minimize linear trim waste generated LW_{ijk}.

Min $LW_{ijk} = k - V_{ij}W_i - V_{ji}W_j$

Subject to

$$V_{ij}W_i - V_{ji}W_j \leq K - 0.025$$
$$D_j L_j V_{ij} \geq 0.9 D_i L_i V_{ji} \quad\quad (1)$$
$$D_j L_j V_{ij} \leq 1.1 D_i L_i V_{ji}$$
$$V_{ij} + V_{ji} \leq 7$$
$$V_{ij} \leq 4$$
$$V_{ji} \leq 4$$
$$V_{ji} \text{ and } V_{ij} \text{ integers}$$

The constraints in (1) take into account some practical considerations and limitations of the corrugator machine highlighted earlier. In particular the first and second constraints limit possible job combinations to those that would require the same run length plus or minus a 10% acceptable margin. The above program could be easily adjusted to allow the production of partial orders by relaxing this constraint and updating the list of orders in S by adding the residual orders with updated remaining demand. This program was solved using LINGO and the optimal number sheets, V_{ij} and V_{ji} that minimize the linear trim waste for the combination (i, j, k) is exported to excel for further processing.

3.2.2. Get Optimal Number of Horizontal Cuts (nb_horizontalcuts)

Given a roll width k and a job combination (i, j), and given the optimal number of sheets per horizontal cut V_{ij} and V_{ji} for orders i and j respectively, the following program returns the optimal number of horizontal cuts H_{ij} and H_{ji} for orders i and j respectively that minimize the geometric trim waste W_{ijk} for the combination (i, j, k)

Min $W_{ijk} = (k - V_{ij}W_i - V_{ji}W_j)H_{ij}L_i$

Subject to
$$\quad\quad (2)$$
$$0.9 D_i \leq V_{ij} H_{ij} \leq 1.1 D_i$$
$$0.9 D_i L_i \leq H_{ji} L_j V_{ij} \leq 1.1 D_i L_i$$
$$H_{ij} \text{ and } H_{ji} \text{ integers}$$

The first constraint in (2) ensures that the total number of sheets produced for each order meets the required demand of order i within a 10% tolerance range. The second constraint ensures that the total vertical length of the paper roll required is the same for the two orders combined within the 10% acceptable margin. The above model is solved using LINGO. The data is imported from excel and the solution was exported back to the same excel sheet.

3.3. Step 3: Solve Roll Width Optimization Problem

The ***Roll Width Optimization problem*** consists of finding, for a set roll width, the optimal combination of orders from S_k that minimize total trim waste W_k and where all primary orders are processed either separately or paired with other primary or secondary orders. This process is repeated for all roll widths available and the total trim waste W_k for each roll width k is computed.

The roll width with minimum W_k is selected and used for production during the day/shift under consideration. Note that S_k is the set of paired jobs that was computed in step 2 where

$$S_k = \{(i, j, W_{ijk}); i \in POrders, j \in AOrders, j \geq i, V_{ij} \neq 0, V_{ji} \neq 0\}$$

The roll width optimization problem is modelled and solved as an assignment problem. Let O be the set of all orders i and j such that $(i, j) \in S_k$. The cost matrix would consist of the waste values W_{ijk} and W_{iik} found in step 2. Note that the cost matrix is symmetrical because $W_{ijk} = W_{jik}$ for a pair of jobs i and j. The optimal combination of jobs to run on the corrugator for a set roll width k is found by solving (3).

$$W_k = \sum_{i \in O} \sum_{j \in O} 0.5 Z_{ij} W_{ijk}$$

Subject to

$$\sum_{i \in O} Z_{ij} = 1; \forall j$$
$$\sum_{j \in O} Z_{ij} = 1; \forall i$$
$$Z_{ij} = \{0,1\}$$

(3)

(3) is solved for every roll width k and the roll width that result in the minimum W_k is set for use during the shift/day under consideration.

4. THE UNIPACK CASE

Motivated by our analysis of the day-to-day scheduling of the corrugator at the UniPaK facility, one of the largest packaging industries in Lebanon, we automated the three-step suboptimal procedure outlined in section 4 to provide UnipaK with a decision support tool that enables them to find a lower cost corrugator schedule. The tool is built in Excel and calls LINGO for solving programs (1), (2) and (3). Excel is used as a platform for preprocessing data files that come from the company and for post-processing results obtained from LINGO.

Customer order data comes in the form of an excel sheet including data such as

- Order name
- Factory Ticket Number
- Order sheet width
- Order sheet length

- Order sheet demand
- Order sheet GSM, fluting, and paper type

A small sample of actual orders that come with due dates in February 2006 are shown in Table 1 for illustration purposes. Customer names and other proprietary information is not shown. Orders are referred to using anonymous customer names.

Table 1: Small Sample of customer Orders for the Month of February

Orders with February 2006 Due Date						
Ticket #	Qty	Delivery Date	Customer Name	Sheet Size	Paper Specifications	Flute Type
26191	1500	2/2/2006	AB	96.4 X 122.9	127 IKL/127 ISC/127 IKL/	CF
28071	2000	2/24/2006	AC	53.8 X 144.3	137 IWT/127 ISC/151 IKL/	CF
29471	50000	2/6/2006	AD	31.1 X 99.5	127 IWT/137 ISC/127 IKL/	BF
33491	10000	2/14/2006	BB	31.1 X 99.5	127 IWT/137 SC/127 IKL/	BF
33991	50000	2/6/2006	CC	31 X 99.5	127 IWT/137 ISC/137 IKL/	BF
33991	50000	2/14/2006	DD	31 X 99.5	127 IWT/137 ISC/137 IKL/	BF
49971	5000	2/5/2006	DE	55.8 X 120.2	137 IWT/127 ISC/127 IKL/	CF
49971	5000	2/23/2006	DF	55.8 X 120.2	137 IWT/127 ISC/127 IKL/	CF
128601	30000	2/27/2006	EE	31.0 X 99.5	127 IWT/137 SC/137 IKL/	BF
128601	30000	2/10/2006	FF	31.0 X 99.5	127 IWT/137 SC/137 IKL/	BF
228091	10000	2/14/2006	GG	79.7 X 114.5	127 IWT/137 ISC/127 IKL/	BF
230091	10000	2/14/2006	HH	79.7 X 114.5	127 IWT/137 SC/127 IKL/	BF
254001	25000	2/28/2006	II	53.0 X 106.4	127 IWT/137 SC/127 IKL/	BF
268281	10000	2/6/2006	JJ	38.5 X 112.5	127 IWT/137 ISC/127 IKL/	BF
268281	10000	2/13/2006	KK	38.5 X 112.5	127 IWT/137 ISC/127 IKL/	BF
268281	10000	2/17/2006	LL	38.5 X 112.5	127 IWT/137 ISC/127 IKL/	BF
269771	5000	2/5/2006	LM	92.9 X 66.5	137 IWT/127 SC/137 IKL/	EF
270881	10000	2/15/2006	MM	38.3 X 89.1	127 IWT/137 SC/127 IKL/	BF
270961	25000	2/28/2006	MO	76.5 X 93.2	137 IWT/127 ISC/127 IKL/	BF

Table 2 shows those jobs from Table 1 that have the same fluting type and paper type and color, namely, it shows the list of jobs requiring B fluting and White 1st grade type paper top.

Table 2: Orders sorted by fluting and paper type

Orders with February 2006 Due Date						
Ticket #	Qty	Customer Name	Sheet Size	Paper Specifications	Flute Type	GSM
29471	50000	AD	31.1 X 99.5	127 IWT/137 ISC/127 IKL/	BF	500
33491	10000	BB	31.1 X 99.5	127 IWT/137 ISC/127 IKL/	BF	400
268281	10000	CC	38.5 X 112.5	127 IWT/137 ISC/127 IKL/	BF	250
33991	50000	DD	31 X 99.5	127 IWT/137 ISC/137 IKL/	BF	330
33991	50000	EE	31 X 99.5	127 IWT/137 ISC/137 IKL/	BF	490
128601	30000	FF	31.0 X 99.5	127 IWT/137 SC/137 IKL/	BF	612
128601	30000	GG	31.0 X 99.5	127 IWT/137 SC/137 IKL/	BF	429
228091	10000	HH	79.7 X 114.5	127 IWT/137 ISC/127 IKL/	BF	432
230091	10000	II	31.1 X 99.5	127 IWT/137 SC/127 IKL/	BF	260
254001	25000	JJ	53.0 X 106.4	127 IWT/137 SC/127 IKL/	BF	195
268281	10000	KK	38.5 X 112.5	127 IWT/137 ISC/127 IKL/	BF	325
268281	10000	LL	38.5 X 112.5	127 IWT/137 ISC/127 IKL/	BF	615
270881	10000	MM	38.3 X 89.1	127 IWT/137 SC/127 IKL/	BF	312

For the purpose of illustration the Primary orders (*POrders*) are identified to be AD through MM and the *AOrders* are the same list of orders. Solving the *Job Selection problem* (step 1) on the data in Table 1 gave the paired jobs shown in Table 3 and the set $S = \{(AD, EE) \ (CC, II) \ (DD, KK) \ (FF, LL) \ (GG, HH) \ (KK, MM)\}$. Jobs in S can be paired because they have GSM requirements within acceptable tolerance.

265

Table 3: Solution of Job Selection Problem

	AD	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM
AD		0	0	0	1	0	0	0	0	0	0	0	0
BB			0	0	0	0	0	0	0	0	0	0	0
CC				0	0	0	0	0	1	0	0	0	0
DD					0	0	0	0	0	0	1	0	0
EE						0	0	0	0	0	0	0	0
FF							0	0	0	0	0	1	0
GG								1	0	0	0	0	0
HH									0	0	0	0	0
II										0	0	0	0
JJ											0	0	0
KK												0	1
LL													0
MM													

The following shows the computations for a selected roll width of 1.8 m. The same steps are carried for all available roll widths but are not presented in this paper. Tables 4 and 5 show the results of solving the *Job Matching problem* (step 2) for the selected roll width of 1.8 m. The same computations are carried on all available roll widths and the results saved for the next step (step 3 of the proposed procedure). In particular, Table 4 shows the optimal number of vertical cuts per horizontal cut found by solving *nb_verticalcuts* for each pair of jobs in S in addition to the optimal number of vertical cuts found if each job is to be processed individually on the corrugator for the same roll width of 1.8 m (shown on the diagonal). For example, for the pair (*AD*, *EE*) in S, the number of vertical cuts of order *AD* is 3 cuts or 3 sheets per horizontal cut and the number of sheets for order *EE* is 2. The number of cuts or sheets per horizontal cut can be read similarly for the remaining pairs of orders in S. Note that the pair *(KK, MM)* of S was eliminated from further consideration as a result of applying *nb_verticalcuts* for a roll width of 1.8m and this is due to the fact that the run length of both jobs does exceed the 10% accepted tolerance. Hence, the set $S_{1.8}$ formed at the end of this step include the remaining jobs in S in addition to individual jobs.

Table 4: Solution of *nb_verticalcuts* for a Roll width of 1.8 m

	AD	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM
AD	4				3								
BB		4											
CC			4						2				
DD				4							4		
EE	2				4								
FF						4						3	
GG							2	1					
HH							1	4					
II			1						3				
JJ										4			
KK				1							4		
LL						1						4	
MM													4

Table 5 shows the results of solving nb_horizontal cuts for each pair of jobs in $S_{1.8}$. In particular it shows the number of horizontal cuts or sheets that will be generated for each job in a pair to satisfy the demand per job shown earlier in Table 1. Table 6 shows the geometric waste in square meters generated from running jobs in the combinations shown in Table 5 for a roll width of 1.8 m. The geometric waste that would be generated from running jobs individually on a roll width of 1.8 m is shown on the diagonal of Table 6.

Table 5: Solution of nb_horizontalcuts for a Roll width of 1.8 m

	AD	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM
AD	12500	0	0	0	15636	0	0	0	0	0	0	0	0
BB	0	2500	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	12500	0	0	0	0	0	24061	0	0	0	0
DD	0	0	0	12500	0	0	0	0	0	0	11250	0	0
EE	13500	0	0	0	7500	0	0	0	0	0	0	0	0
FF	0	0	0	0	0	7500	0	0	0	0	0	10176	0
GG	0	0	0	0	0	0	5000	9000	0	0	0	0	0
HH	0	0	0	0	0	0	10357	2500	0	0	0	0	0
II	0	0	22500	0	0	0	0	0	8333	0	0	0	0
JJ	0	0	0	0	0	0	0	0	0	2500	0	0	0
KK	0	0	0	9950	0	0	0	0	0	0	2500	0	0
LL	0	0	0	0	0	9000	0	0	0	0	0	2500	0
MM	0	0	0	0	0	0	0	0	0	0	0	0	2500

Table 6: Waste in square meters generated for a roll width of 1.8 m

	AD	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM
AD	6950				3837								
BB		1390											
CC			7000						15640				
DD				7000							1969		
EE	3837				4200								
FF						4200						4935	
GG							1030	6228					
HH							6228	1390					
II			15640						1749.9				
JJ										650			
KK				1969							650		
LL						4935						650	
MM													670

Table 7 shows the results obtained by solving the *Roll Width Optimization* problem (Step 3 of the procedure) for a roll width of 1.8 m. The results indicate the combination of jobs that would result in minimal total geometric waste. Note that the combinations *(CC, II)* and *(FF, LL)* were not optimal and thus only *(AD, EE),* and *(DD, KK)* were retained from S the remaining jobs are to be run individually if total geometric waste is to be minimized.

Table 7: Solution of *Roll Width Optimization* problem for a roll width of 1.8 m

	AD	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM
AD	0	0	0	0	1	0	0	0	0	0	0	0	0
BB	0	1	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	1	0	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	1	0	0
EE	1	0	0	0	0	0	0	0	0	0	0	0	0
FF	0	0	0	0	0	1	0	0	0	0	0	0	0
GG	0	0	0	0	0	0	1	0	0	0	0	0	0
HH	0	0	0	0	0	0	0	1	0	0	0	0	0
II	0	0	0	0	0	0	0	0	1	0	0	0	0
JJ	0	0	0	0	0	0	0	0	0	1	0	0	0
KK	0	0	0	1	0	0	0	0	0	0	0	0	0
LL	0	0	0	0	0	0	0	0	0	0	0	1	0
MM	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 8 summarizes the findings for a roll width of 1.8 m. The least cost corrugator schedule for this roll width calls for combining jobs AD and EE and jobs DD and KK together and running the remaining jobs separately. The resulting total waste for this job combination and roll width is found to be 24536.07 m^2 or the equivalent of $ 837.91.

The same steps were carried for all available roll widths namely 1.9, 2, 2.05, 2.1, 2.10, 2.15, and 2.2 meters and the total waste value was found for each roll width and resulting corrugator schedule. The minimum total waste among all roll widths did correspond to the roll width of 1.8 m and thus the schedule shown in Table 8 is the final schedule returned by the proposed method. Note that no setup time is recorded for the proposed

schedule because the roll width is fixed over the entire production.

Table 8: Optimal Schedule for a roll width of 1.8 m

Combination of Orders	# of sheets per horizontal cut	Waste (m²)	Roll Length (m) (Order 1, Order 2)	Production Time (min)	Setup Time (min)
(AD, EE)	(3, 2)	3837.39	(15458.32, 15457.5)	154.5832	No Setup
B	4	1390	2487.5	24.875	No Setup
C	4	7000	12437.5	124.375	No Setup
(DD, KK)	(4, 1)	1968.75	(11193.75, 11193.75)	111.9375	No Setup
FF	4	4200	7462.5	74.625	No Setup
GG	2	1030	5725	57.25	No Setup
HH	4	1390	2487.5	24.875	No Setup
II	3	1749.93	8866.312	88.66312	No Setup
JJ	4	650	2812.5	28.125	No Setup
LL	4	650	2812.5	28.125	No Setup
MM	4	670	2227.5	22.275	No Setup
Total Geometric Waste (m²)	24536.07				
Total Cost of Waste ($)	$837.91				
Total Setup time (min)	0				
Total Setup Cost ($)	$0				
Total Cost ($)	$837.91				

Table 9 shows the actual schedule for the job selection shown in Table 1. The roll width was changed depending on the combination of jobs to minimize geometric waste or material scrap.

Table 9: Summary of waste and other operations information of actual corrugator schedule

Combination of Orders	# of sheets/ horizontal cut	Roll Width (cm)	Waste (m²)	Roll Length (m) (Order 1, Order 2)	Production Time (min)	Setup Time (min)
(AD, EE)	(4, 3)	220	325	(12437.5, 11450)	124.375	-
B	4	180	1390	2487.5	24.875	15
C	4	180	7000	12437.5	124.375	0
(DD, KK)	(3, 2)	180	1666.7	(16583.665, 5625)	165.84	0
(FF, LL)	(3, 1)	180	4935	(9950, 5625)	99.5	0
(GG, HH)	(2, 1)	200	950	(9950, 5725)	99.5	15
II	4	220	500	6650	66.5	15
JJ	4	180	650	2812.5	28.125	15
MM	4	180	670	2227.5	22.275	0
Total Geometric Waste (m²)	18086.75					
Total Cost of Waste ($)	$617.57					
Total Setup time (min)	60					
Total Setup Cost ($)	$420					
Total Cost ($)	$1037.57					

The actual schedule as shown has a lower total materials waste cost, however if we consider the additional setup time associated with changing roll widths and its equivalent dollar value, we find that the total cost of the actual schedule is higher than the cost of the proposed schedule as shown in Table 10 below. The savings for the small selection of 20 jobs presented in this paper is around $200. This figure is much more significant if all orders over the month of February were considered and even more significant if we were to consider the cost of waste over a year of production.

Table 10: Comparison of proposed vs. actual corrugator schedule

	Cost of Waste Generated	$ Value of Setup Time Needed
Actual Schedule	615.57	420
Proposed Schedule	837.91	0
Savings dollar value	(220.34)	420
Total Savings ($)		200

5. CONCLUSION

This paper presented a three-step suboptimal procedure that draws a daily corrugator schedule while minimizing both trim waste cost and setup time cost where setup cost is measured as the loss of production resulting from stock rolls size changes. The procedure yields the optimal roll width to use over an entire shift. The proposed method was motivated by our analysis of the day-to-day scheduling of the corrugator at the UniPaK facility one of the leading manufacturer of cardboard boxes in Lebanon. An example application of the proposed method to UniPak operations showed that the proposed procedure yields a lower-cost daily schedule of jobs on the corrugator when compared to the cost of the actual schedule for the same day. Indeed and although the actual schedule had a lower total materials waste cost, if we consider the additional setup time associated with changing roll widths and its equivalent dollar value, the proposed procedure gave significant improvements. By fixing the corrugator roll width for a whole shift of eight hours, we were able to achieve significant improvement and savings in terms of the dollar value of paper loss.

REFERENCES

Eismann, K., 1957. The trim problem, *Management Science*, 3, 279-284.

Gilmore, P.C., Gomory, R.E., 1961. A Linear Programming Approach to the Cutting Stock Problem, *Operations Research*, 9, 848-856

Haessler, R. W., Talbot, F.B., 1983. A 0-1 Model for Solving the Corrugator Trim Problem, *Management Science*, 29 (2), 200-209.

Darely, V., Sanders, D., von Tessin, P., 2004. An Agent-Based Model of a Corrugator Box Factory: The tradeoff between Finished Goods Stock and On-Time-In-Full Delivery, *Proceedings 5th workshop on Agent-Based Simulation*, H. Coelho, B. Espinasse, eds.

Krichagina, E.V., Rubio, R., Taksar, M.I., Wein, L.M., 1998. A Dynamic Stochastic Stock-Cutting Problem, *Operations Research*, 46 (5), 690-701.

Viswanthan, K.V., Bagchi, A., 1993. Best-First Search Methods for Constrained Two-Dimensional Cutting Stock Problems, *Operations Research*, 41 (4), 768-776.

Bookbinder, J.H., Higginson, J.K., 1986. Customer Service vs Trim Waste in Corrugated Box Manufacture, *The Journal of the Operational Research Society*, 37 (11), 1061-1071.

WORKSTATION PRODUCTIVITY ENHACEMENT WITHIN HYDRAULIC HOSES MANUFACTURING PROCESS

Antonio Cimino[(a)], Duilio Curcio[(b)], Francesco Longo[(c)], Enrico Papoff[(d)]

[(a) (b) (c) (d)] Modeling & Simulation Center - Laboratory of Enterprise Solutions (MSC-LES)
Mechanical Department, University of Calabria, 87036 Rende (CS), Italy

[(a)]acimino@unical.it, [(b)]dcurcio@unical.it, [(c)]f.longo@unical.it, [(d)]epapoff@tin. it

ABSTRACT
The paper proposes the effective redesign of a workstation belonging to a manufacturing process devoted to produce high pressure hydraulic hoses. The objective of the study is to increase the workstation productivity in terms of number of Shop Orders being completed on monthly basis. To this end, the authors adopt an approach based on Modelling & Simulation as support tool for implementing a three dimensional environment capable of recreating, with satisfactory accuracy, the real workstation. A final configuration of the workstation is presented and higher productivity levels are achieved.

Keywords: Modelling & Simulation, productivity, workstation design

1. INTRODUCTION
An overview of the state of the art, starting from the second half of the 1990s, reveals that manufacturing systems continuously provide challenging problems for researchers and scientists working in this area. Actually the workstation effective design became more and more important due to their effects on system efficiency and productivity. A number of different research works and scientific approaches have been proposed, trying to achieve an effective design of the workstations belonging to a manufacturing system.

Das and Sengupta (1996) propose a workstation design procedure based on the optimization of the worker and total system productivity as well as worker physical and mental well-being, job satisfaction and safety.

Resnick and Zanotti (1997) develop a set of guidelines for the workstation design integrating ergonomics and productivity objectives. An application example is proposed for remarking that a workstation can be designed to maximize performance and reduce costs by considering both ergonomics and productivity together.

Engstrom and Medbo (1997) develop a video based observation method for time data collection and analysis of work time consumption. The method allows to measure the efficiency of the production system by separating between value and not value adding works activities. In this regards, the method can be used for increasing manufacturing systems productivity.

Vedder and Hellweg (1998) recorder twenty day and night shifts in a fibbre spinning area of a chemical plant by means of a stationary camera. A very long analysis of the videotapes allows them to povide the guidelines for redesigning the system under consideration in order to achieve higher productivity levels.

Laring et al. (2002) develop an ergonomic complement to a modern MTM system called SAM. In particular the authors propose a tool that gives the possibility to estimate simultaneously the consumption of time in the envisaged production, and the biomechanical load inherent in the planned tasks.

Udosen (2006) proposes a tools for construction, evaluation and improvement of a workplace for the assembly of a domestic fan. The tool can be used for minimizing the cycle time of the assembly operations.

In the last decade of the 20th century, researchers and practitioners started to develop research works for the workstation effective design by using Modelling & Simulation (M&S) as support tool for choosing correctly, for understanding why, for diagnose problems and explore possibilities (Banks, 1998).

Ben-Gal and Bukchin (2002) propose a structural methodology based on M&S for the workstation design. Factorial experiments and the response surface methodology are integrated in order to reduce the number of examined design solutions and obtain an estimate for the best design configuration with respect to multi-objective requirements (acceptable ergonomic risk levels and high productivity values).

Longo et al. (2006) use M&S for the effective design of an assembly line still not in existence. The authors propose a multi-measures approach with the aim of obtaining a different work assignment to each workstation, better line-balancing and better ergonomic solutions.

Chang and Wang (2007) propose a method for conducting workplace ergonomic evaluations and re-design in a digital environment with the aim of both preventing work-related musculoskeletal disorders during assembly tasks in the automotive sector and increasing the system productivity.

Santo et al. (2007) use M&S for solving ergonomic problems and achieving higher productivity levels. A case study based on a non repetitive manufacturing process is presented.

The objective of the paper is to achieve the effective design of a workstation (Pressure test workstation) belonging to a real manufacturing system. To this end, the authors adopt a M&S approach supported by a three dimensional virtual environment, in which the workstation has been recreated with high accuracy. The simulation model is used for developing an improved workstation configuration in terms of higher productivity levels. The choice of the final configuration is made according to the number of Shop Orders being completed on the monthly basis.

The paper is organized as follows. Section 2 describes the Pressure test workstation manufacturing process. Section 3 gives specific details on the simulation model implementation and validation. Section 4 presents a detailed analysis of the simulation results (number of Shop Order being completed on monthly basis) and shows how the final workstations configurations can be obtained. The last section reports the conclusions that summarize the scientific value of the work.

2. THE PRESSURE TEST WORKSTATION MANUFACTURING PROCESS

The research work focalizes on the effective redesign of the Pressure test workstation of a production process for manufacturing and assembling high-pressure hydraulic hoses. All the workstations of the production process are located inside the Assembly area and each workstation performs specific operations. A preliminary analysis carried out by the company top management shows that the productivity of the Pressure test workstation (evaluated on monthly basis) always falls below the target levels causing delays in Shop Orders (S.Os) completion.

Figure 1 shows the final products (the high-pressure hydraulic hoses).

Figure 1: Hydraulic hoses

Each hydraulic hose is made up of a rubber hose, two fittings and two ring nuts. In the workstation the test of the hydraulic hoses by using a pressure machine (setting a pressure value higher than the nominal value) is accomplished.

The operator of the Pressure test workstation performs the following operations:

1. he walks three steps to pick up the Shop Orders sheet, reads the information he/she needs and puts it back;
2. he takes the hydraulic hoses from a manual dolly and brings it on the work table;
3. he walks four steps to pick up the joints for connecting the hydraulic hoses to the pressure test machine;
4. he connects joints and hydraulic hoses;
5. he moves the hydraulic hoses from the work table to the testing machine;
6. he connects the hydraulic hoses to the testing machines, performs the security procedures and starts the testing phase;
7. he puts away from the machine the hydraulic hoses, performs the visual checks and moves the hoses on the work table;
8. he disconnects the joints from the hydraulic hoses;
9. he walks fours steps to put the joints back in the proper bins and comes back to the work table;
10. he brings the tested hydraulic hoses to a manual dolly;
11. he completes the Shop Order by setting the status "end of the operation" on the company informative system; in this regards he walks 8 steps to use the workstation PC;
12. he moves the materials to the successive workstation by using a manually operated dolly.

Note that, the operations 1, 11 and 12 are performed just once for the Shop Order completion. On the other hand all the other operations are cyclically performed for each hydraulic hoses. Obviously the frequency of such operations depends on the number of hydraulic hoses required for the Shop Order completion. In this regards, the authors take into consideration a typical Shop Order that requires the production of 12 medium section hydraulic hoses. Moreover, as concerns the work method, three hydraulic hoses must be simultaneously tested by means of the pressure machine.

3. THE SIMULATION MODEL DEVELOPMENT

The first step of the research work was the development of a simulation model representing the actual Pressure test workstation. The simulation modelling tools, used for developing the Pressure test workstation model are

the CAD software Rhinoceros and the simulation software eM-Workplace.

All the tools used during the production process have been modelled by means of Rhinoceros. Obviously, the implementation of the geometric models has required an accurate data collection of objects types, dimensions and weights. The data collection includes the following elements of the real workstation: machines, equipment and tools, worktables, manual operated dollies, raw materials and bins.

Table 1 reports the objects description, dimensions and weights.

Table 1: Data collection for geometric model implementation

Object Description	Object Type	Weight (Kg)	Dimensions (cm) L x W x H
Ring	Component	0.168	Depending on S.O.
Fitting	Component	0.336	Depending on S.O.
Marking die	Component	1.800	Depending on S.O.
Workstation stamp	Component	0.100	Depending on S.O.
Scanner	Component	0.400	12 x 7 x 18
Empty bin	Component	0.300	30 x 20 x 15
Rubber hose	Component	1.020	Depending on S.O.
Manual operated Dolly	Equipment	35.300	100 x 120 x 76
Rings bin	Equipment	0.300	30 x 20 x 15
Work table	Equipment	100.800	240 x 220 x 95
Pressure test machine	Machine	1020.040	368 x 90 x 150

After the modelling phase, the geometric models have been imported into the virtual environment provided by eMWorkplace. Note that the geometric models have been located in the virtual environment considering the same position the real objects take place in the real system in order to recreate exactly the Pressure test workstation. Figure 2 shows a panoramic overview of the virtual Pressure test workstation.
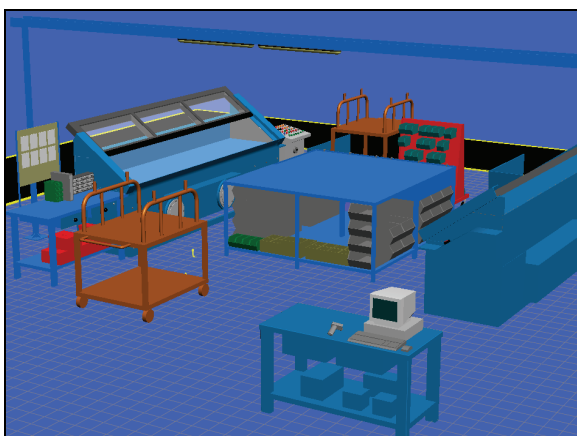
Figure 2: The virtual Pressure test workstation

The completion of the simulation model requires the insertion and training of human models capable of performing the workstation operations in the virtual environment. The selection of the human models type is based upon an accurate analysis of operators' characteristics (age, gender, height, weight and health). The objective is to select and import, from eM-Workplace libraries, human models representing as much as possible the real workers. After the insertion into the virtual environment, the human model is only able to stand in the waiting position; the model has to be trained to perform the manufacturing operations. In this regards, eMWorkplace provides the user with a programming language for teaching different types of activities and reproducing correctly each type of operation.

3.1. The simulation model validation

In order to increase the probability of success of a simulation study, the validation of the simulation model is a mandatory step. The main goal of the validation is to verify if the simulation model is capable of recreating the real system evolution over the time with satisfactory accuracy.

The simulation model validation has been carried out by the face validation technique. The face validation technique has been used for comparing the real process time and the simulated process time required for completing a typical Shop Order (production of 12 medium section hydraulic hoses). The real process time values are available on the company informative system (as mentioned in the section 2, at the end of each operation the worker sets the status "end operation" on the company information system). The real process time required for completing the Shop Order is 628.6 seconds. The simulated process time for completing the same Shop Order is equal to 602.2 seconds. An error of 4.2% reveals that the simulation model is capable of recreating the real Pressure test workstation with satisfactory accuracy.

4. THE SIMULATION RESULTS AND THE FINAL WORKSTATION CONFIGURATION

Let us consider the actual workstation configuration. The simulated process time for completing the Shop Order under consideration is equal to 602.2 seconds (about 10 minutes and 2 seconds). Taking into consideration the time allowance for personal, fatigue and delay (the time allowance is about the 16% of the total process time) and the 8 hours working shift, the number of Shop Orders per day being completed is 40. The workstation production on monthly basis is reported in figure 3.
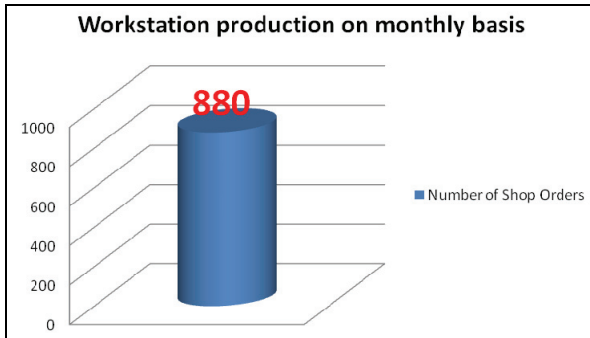
Figure 3: Monthly production of the Pressure test workstation

Once the initial solution was studied, a new workstation configuration was developed for achieving higher productivity levels. Let us list the workstation changes the authors propose:

- The manual dolly, the hydraulic hoses to be tested are placed on, has been located close the work table in order to eliminate the worker walking; figure 3 and figure 4 show respectively the actual and the final workstation configurations;


Figure 3: Manual dolly carrying the hydraulic hoses to be tested initial position

Figure 4: Manual dolly carrying the hydraulic hoses to be tested final position

- The manual dolly, the tested hydraulic hoses are placed on, has been moved near the work table; such change allows to reduce the time required by the operator for moving the tested hydraulic hoses. Figure 5 and figure 6 show respectively the actual and the final workstation configurations;
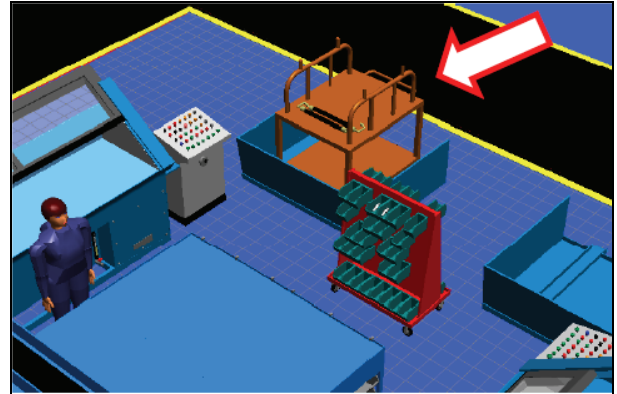
Figure 5: Manual dolly carrying the tested hydraulic hoses initial position
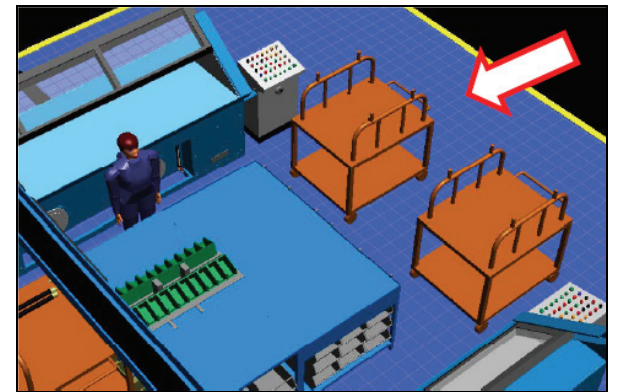
Figure 6: Manual dolly carrying the tested hydraulic hoses final position

- The PC location, used by the operator for setting the status "end of operation" on the company informative system, has been moved near the pressure test machine. Note that such change allows to reduce the number of steps required for performing the operation; figure 7 and figure 8 show respectively the actual and the final workstation configurations;
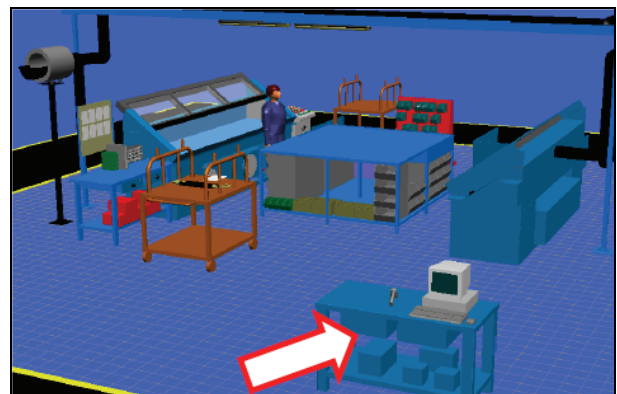
Figure 7: PC initial position.

271
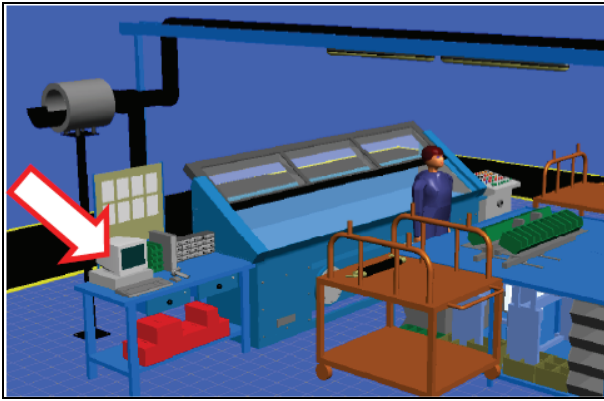
Figure 8: PC final position

- The joints bins have moved from the initial location (four steps far the work table) to the work table. Figure 9 and figure 10 show respectively the actual and the final workstation configurations;
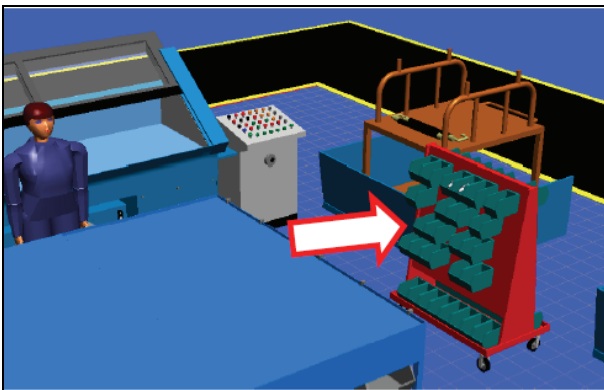
Figure 9: Joints bins initial position.
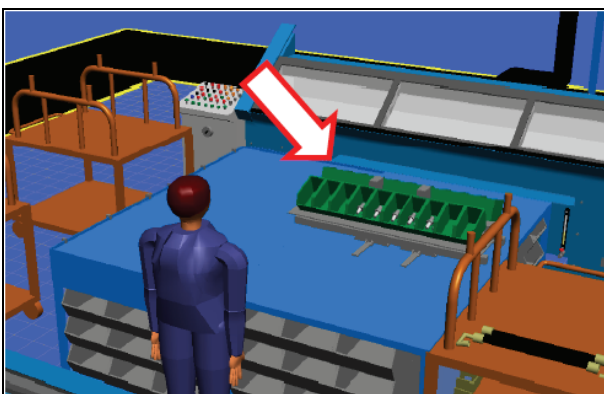
Figure 10: Joints bins final position.

The workstation changes allow the operator to have all the equipment and tools closer and in turn to reduce the time for performing the operations.

Let analyze the final workstation configuration in terms of number of Shop Orders being completed on monthly basis.

The simulation model evaluates in about 564 seconds (9 minutes and 24 seconds) the process time for completing the Shop Order under consideration. The number of Shop Orders per day being completed is 44.

The final workstation production on monthly basis is reported in figure 12.
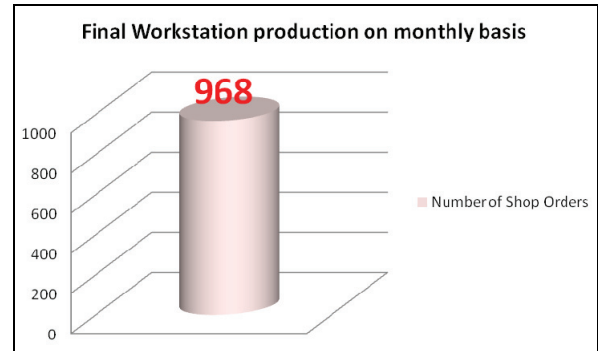
Figure 12: Monthly production for the final configuration of the Pressure test workstation

Compare now the productivity values for the actual and the final workstation configurations. Figure 13 reports the productivity values on monthly basis for both the configurations.
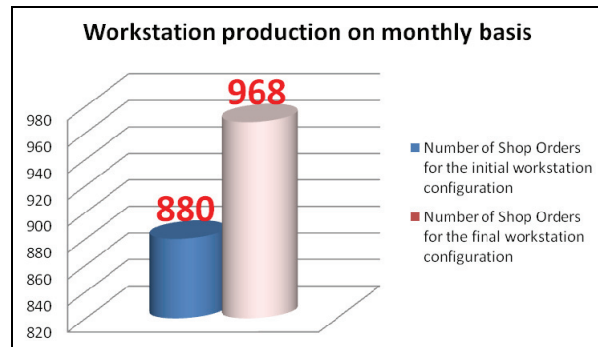
Figure 13: Productivity values for the actual and the final workstation configurations.

Note that the final workstation configuration allows to complete 88 Shop orders more the initial workstation configuration. In this regards the productivity improvement is about 10 % respect to the initial level.

The productivity growth has been appreciated by the company top management so that the workstation changes have been applied to the real system.

5. CONCLUSION

The research work proposes the effective redesign of a workstation belonging to a manufacturing process devoted to produce high pressure hydraulic hoses.

The objective of the study is to increase the workstation productivity in terms of number of Shop Orders being completed on monthly basis.

The authors started the research work by modelling the actual configuration of the workstation. The simulation model has been developed by using the CAD software Rhinoceros and the simulation software eM-Workplace.

The actual productivity level has been evaluated by means of the simulation model. Several workstation

changes in terms of objects position have been accomplished and tested by the simulation model.

A final workstation configuration has been proposed achieving productivity levels higher than the initial ones.

REFERENCES

Banks. J., 1998. *Principles of Simulation, Handbook of Simulation*. New York: Wiley Interscience.

Ben-Gal, I. and Bukchin, J., 2002. The ergonomic design of workstations using virtual manufacturing and response surface methodology. *IIE Transaction*, 34, 375-391.

Bocca, E., Longo, F., 2008. Simulation Tools, Ergonomics Principles and Work Measurement Techniques for Workstations Design. *Proceedings of Summer Computer Simulation Conference, 15-18 June, Edinburgh (UK)*.

Chang, S.-W., Wang, M.-J. J., 2007. Digital Human Modeling and Workplace Evaluation: Using an Automobile Assembly Task as an Example. *Human Factors and Ergonomics in Manufacturing,* 17, 445-445.

Das, B., Sengupta, A. K., 1996. Industrial workstation design: a systematic ergonomics approach. *Applied Ergonomics*, 27, 157-163.

De Sensi, G., Longo, F., Mirabelli, G., 2007-a. *Modeling & Simulation Based Approach for Optimizing Seal Press Workstation in a Manufacturing System*, Proceedings of Business and Industry Symposium, March 25-29, USA.

De Sensi, G., Longo. F., Mirabelli, G., 2007-b. *Ergonomic work methods optimization in a three dimensional environment*, Proceedings of Summer Computer Simulation Conference, July 15-18, San Diego, California, USA

Engström, T., Medbo, P., 1997. Data collection and analysis of manual work using video recording and personal computer techniques. *International Journal of Industrial Ergonomics,* 19, 291-298.

Feyen, R., Liu, Y., Cha, D., Jimmerson, G., Joseph, B., 2000. Computer-aided ergonomics: a case study of incorporating ergonomics analyses into workplace design. *Applied Ergonomics,* 31, 291-300.

Kadefors, R., Forsman, M., 2000. Ergonomic evaluation of complex work: a participative approach employing video computer interaction, exemplified in a study of order picking. *International Journal of Industrial Ergonomics,* 25, 435-445.

Laring J., Forsman M., Kadefors R., Örtengren R., 2002. MTM-based ergonomic workload analysis. *International Journal of Industrial Ergonomics*, 30, 135-148.

Longo, F., Mirabelli, G., Papoff, E., 2005. *Techiche di analisi avanzate per la progettazione efficiente delle postazioni di assemblaggio manuale,* SdA – Soluzioni di Assemblaggio, VNU Business Publications Italia.

Longo, F., Mirabelli, G., Papoff, E., 2006-a. *Material Flow Analysis and Plant Lay-Out Optimization of a Manufacturing System*, International Journal of Computing, 5(1), 107-116.

Longo, F., Mirabelli, G., Papoff, E., 2006. Effective Design of an Assembly Line Using Modeling & Simulation. *Proceedings of the Winter Simulation Conference*, Monterey, California, USA.

Resnick, M.L., Zanotti, A., 1997. Using ergonomics to target productivity improvements. *Computer & Industrial Engineering,* 33, 185-188.

Santos, J., Sarriegi, J. M., Serrano, N., Torres, J. M., 2007. Using ergonomic software in non-repetitive manufacturing processes: A case study. *International Journal of Industrial Ergonomics,* 37, 267-275.

Udosen U.J., 2006. Ergonomic workplace construction, evaluation and improvement by CADWORK. *International Journal of Industrial Ergonomics,* 36, 219-228.

Vedder, J., Hellweg, 1998. Identifying postural hazards with a video-based occurrence sampling method. *International Journal of Industrial Ergonomics*, 22, 4-5.

Wilson J.R., 1997. Virtual environments and ergonomics: needs and opportunities. *Ergonomics,* 40, 1057-1077.

Zandin K.B., 2001. *Maynard's Industrial Engineering Handbook*, 5th ed. New York: McGraw-Hill.

AUTHORS BIOGRAPHY

ANTONIO CIMINO was born in Catanzaro (Italy) in October the 1th, 1983. He took his degree in Management Engineering, summa cum Laude, in September 2007 from the University of Calabria. He is currently PhD student at the Mechanical Department of University of Calabria. His research activities concern the integration of ergonomic standards, work measurement techniques, artificial intelligence techniques and Modeling & Simulation tools for the effective workplace design. He collaborates with the Industrial Engineering Section of the University of Calabria to research projects for supporting innovation technology in SMEs.

DUILIO CURCIO was born in Vibo Valentia (Italy), on December the 15th, 1981. He took the degree in Mechanical Engineering from University of Calabria (2006). He is currently PhD student at the Mechanical Department of University of Calabria. His research activities include Modeling & Simulation and Inventory Management theory for production systems and Supply Chain design and management. He collaborates with the Industrial Engineering Section of the University of Calabria to research projects for supporting Research and Development in SMEs.

FRANCESCO LONGO took the degree in Mechanical Engineering from University of Calabria (2002) and the PhD in Industrial Engineering (2005). He is currently researcher at the Mechanical Department (Industrial Engineering Section) of University of Calabria. His research interests regard modeling & simulation of manufacturing systems and supply chain management, vulnerability and resilience, DOE, ANOVA. He is Responsible of the Modeling & Simulation Center – Laboratory of Enterprise Solutions (MSC-LES), member organization of the MS&Net (McLeod Modeling & Simulation Network) He is also member of the Society for Computer Simulation International and Liophant Simulation.

ENRICO PAPOFF was born in Naples (Italy) on February the 03rd, 1948. He took the degree in Mechanical Engineering from University of Napoli Federico II, in 1973. He is currently Associate Professor at the Mechanical Department (Industrial Engineering Section) of the University of Calabria. His research interests regard project management and business plans.

INTEGRATED PLANNING AND CONTROL ON RO-RO TERMINALS

Bernd Scholz-Reiter [a], Felix Böse [b], Michael Teucke [c] , Jakub Piotrowski[d]

[a] [b] [c] [d]BIBA – Bremer Institut für Produktion und Logistik GmbH, University of Bremen

[a]bsr@biba.uni-bremen.de, [b]boe@biba.uni-bremen.de, [c]tck@biba.uni-bremen.de [d]pio@biba.uni-bremen.de

ABSTRACT

In vehicle logistics several logistics service providers are involved across the process chain ranging from transport of new vehicles to the automobile terminal via storage and technical modifications on the terminal to delivery of the vehicles to the customer. Today, coordination of the process executors is marked by a short planning horizon, as none of the involved parties have access to all necessary planning information of the entire process. Currently no integrated process design of the logistic chain is executed.

This contribution deals with developing an integrated planning and control method for collaborative vehicle logistics processes. The method considers both vehicle terminal processes as well as transport processes executed by automobile logistics providers.

The logistic effects of applying this method will be determined in a simulation study using discrete event simulation. The method will be implemented in real automobile terminals by means of a prototypical software based control centre.

Keywords: automobile logistics, planning and control, Ro-Ro terminals

1. INTRODUCTION

This contribution deals with integrated planning and control of order processing in vehicle logistics performed at automobile terminals. Such terminals are situated at strategically important traffic nodes, like sea harbours, and provide a network of service points that is essential for vehicle supply chains marked by steadily increasing internationalisation of production and sales.

The German automotive industry e.g. in 2006 exported 70% of the production volume, or 3.9 million cars and 290,000 commercial vehicles, including 650,000 cars and 8,300 commercial vehicles to North America and 350,000 cars and 12,600 commercial vehicles to Asia. On the other hand, ca. 30% of all cars newly licensed in Germany in 2006 were produced by foreign manufacturers, while an additional 23% were cars of German brands but not produced in Germany (N.N. 2007).

Consequently vehicle outbound transhipments via European ports of the North Range (Zeebrugge – Ham-burg) during the 1990s more than tripled whereas inbound transhipments remained stable (Mattfeld 2005).

A large part of the vehicle exports as well as many inland sales are conducted via automobile terminals. At the automobile terminal of the German sea port Bremerhaven, which will serve as an implementation partner for the method proposed in this contribution, nearly 2 million new vehicles (1.85 million vehicles in 2006) are handled each year on their way from the manufacturer to the dealer or retailer and are exposed to a variety of value added services.

These services include shipping or arrival both from sea or land side, storage, finishing and technical modifications to meet the multitude of individual customer requirements for the ordered vehicles as well as distribution to the retailers. Thus the service portfolio of automobile logistics service providers includes not only transport logistics, but also services falling into the field of production logistics.

The typical process chain of an automobile terminal can be structured into a set of process steps as illustrated in Figure 1. These process steps can be described as follows (Böse and Windt 2007, Mattfeld 2006):

1. Intermodal transport of the new vehicles to the terminal is normally performed by ship (often with Ro-Ro ferries), train or flat-bed trailers. Terminals are normally situated with very good access to the traffic infrastructure, so that imported vehicles arriving by ship can be unloaded directly onto the terminal. For ship transports to the terminal the expected time of arrival as well as a freight list are sent to the terminal several days before arrival.

2. Vehicle registration on the terminal includes registration and collection of all relevant vehicle data and transfer of these data into the operational software systems. Each vehicle is identified by its vehicle identification number (VIN) from the terminal staff, often by using mobile data entry de-vices, which can read barcodes. The vehicle identification number than allows identification of the vehicle during all later processes.

3. Storage at the terminal area: Based on pre-defined priorities the IT control system

allocates a storage location of a storage area as a storage position to each vehicle. A handling employee moves the vehicle to the assigned storage location.

4. Technical treatment and the subsequent allocation and delivery processes are normally triggered by arrival of the delivery request for that vehicle. After receipt of the delivery request the vehicle is removed from stock. The sequence of the technical treatment stations a vehicle has to run through is specified in the technical treatment order of the vehicle. The technical treatment steps include e.g. re-coating of vehicles, or washing and removal of separate coating layers (e.g. from wax) temporarily added for vehicle protection during transport. For individual customer orders, additional technical services may include installation of special equipment like radios or satellite navigation de-vices, and removal and replacement of certain parts or components.

5. Allocation and delivery: Upon completion of all technical treatment tasks, the vehicle is provided to the shipment area of the terminal, where the transport service provider can take it over.

6. Transport from the terminal to the automobile dealer or customer is normally executed by specialised transport service providers using special transport vehicles like e.g. flat bed trailers or low-loading trucks.

Figure 1: Process Chain of an Automobile Terminal

As not every sea port terminal offers all technical ser-vices, such as removal of transport protection systems, re-pair and coating, or mounting of special devices, the terminal services are frequently offered jointly by several specialised and cooperating companies, which fosters increased participation and cooperation of the automobile logistics service providers (automobile producers, transport carriers, automobile terminals, automobile retailers) in cross-company networks.

Typically storage services are provided by the storage service provider, while technical treatment is performed by separate technical service providers situated on the automobile terminal and transport to the terminals and from the terminals to the customers is performed by specialist transport service providers.

In the mentioned Bremerhaven sea port automobile terminal e.g. arrival and storage services are executed by BLG LOGISTICS AUTOMOBILE acting as the storage service provider, while technical services are executed by EHH AUTOTEC acting as technical service provider. The transportation of the vehicles is achieved by EHH Auto-mobile-Logistics operating as transport service provider.

2. CURRENT PROCESS EXECUTION AND RESULTING WEAKNESSES

Until today no integrated planning and control systems have been established for the order execution of an automobile terminal that integrate processes and information of related logistic services like e.g. transport logistics. Additionally planning and control of order execution processes on automobile terminals like e.g. management of parking positions or scheduling and sequencing for technical stations is mostly performed manually and thus unsystematically.

Thus the generally existing logistic target conflicts between interests of the customers (high logistic service quality) and those of the service providers (low logistic service costs) are intensified by lacking coordination and partially competing interests of the separate operating process executors.

For example, vehicles arriving by ship are temporarily stored on intermediary buffer positions to optimize lay days and waiting times of ships in the harbour. Storage position management, however, tries to minimize vehicle transport times on the automobile terminal and thus, if possible, reduce or prevent storage relocations. Vehicle transfer to transporters has to cope with similar weaknesses stemming from sub-optimal coordination with terminal providers.

Coordination of the process executing parties is generally characterised by a short term planning horizon. Vehicle call-off orders by technical service providers can only be executed, when a call-off order for the respective vehicle has been called by the vehicle producer or retailer and the technical processing is due the next day. This situation results in bottleneck effects and high occupancy of commissioning areas,
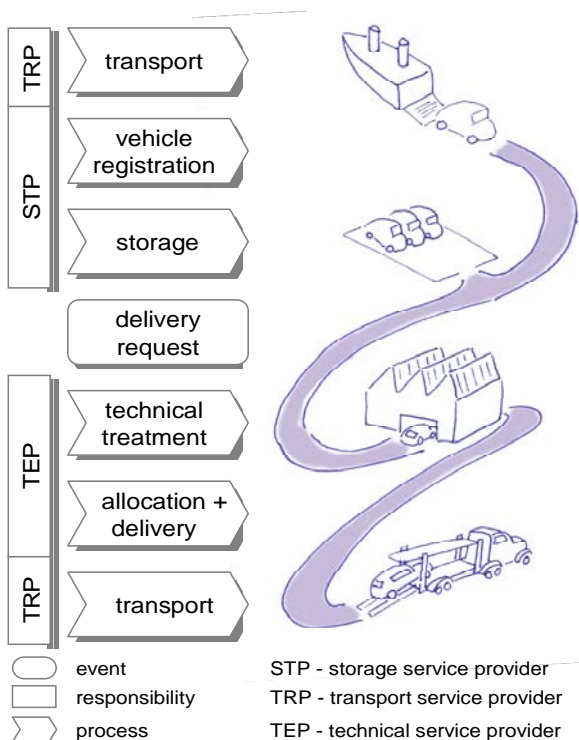
preventable restoring activities and reduced adherence to delivery dates.

Vehicle commissioning and transfer from terminal provider to transport carrier is mostly executed at short notice, too. Transport times to the automobile retailer are considered for scheduling during the disposition of the vehicles by the terminal provider; however they are not considered for optimization efforts of costs and times. Avail-able transport capacities and relations of the transport carrier are not considered during planning and control of the terminal processes.

Based on the requirements of capacity planning and sequencing of orders at the technical treatment stations the terminal provider determines, which vehicles are allocated for transport and when transport dates are scheduled. Vehicles, which are prepared for transport, are messaged to the transport carrier's disposition system only after they have been allocated to the consignment areas and can only then be disposed of in the route planning procedures of the transport carrier. Duly timed transport capacity planning and route plan optimization (in the sense of a high payload factor and a low number of empty drives) are thus made more difficult.

Negative effects of the lacking coordination between terminal providers and transport carriers include high occupancy of the already large consignment areas, non-optimal utilization and efficiency of the carrier lorries, high uploading times for the vehicles, which are insufficiently concentrated and thus spread over a large area and non-optimal adherence to delivery dates. Insufficient utilization of transport carrier lorries due to short planning horizons with limited timeframes for search of amendment loads results in a larger traffic load on harbour to hinterland communications.

3. METHODICAL PROCESS IMPROVEMENT

Optimization of single processes of the closely interrelated vehicle logistics process chain in isolation is only possible to a limited extent and in numerous cases adversely affects previous or successive processes.

Logistic and manufacturing processes on automobile harbour terminals characterised by several involved logistic service providers, like e.g. the automobile terminal of the German sea port Bremer-haven, have to be regarded as a unified system, which also can be optimized only as a whole system. Thus an integrated planning and control method, such as proposed in this contribution, has to consider both, terminal as well as transport processes.

A graphical software control centre (Kurbel 1999) will provide all relevant information to all process executing parties, thus enabling a more efficient and timely planning and control of order execution and related processes on sea port automobile terminals.

The data contributed by the three parties, which are considered by the integrated control centre proposed in this contribution, are shown in Figure 2.
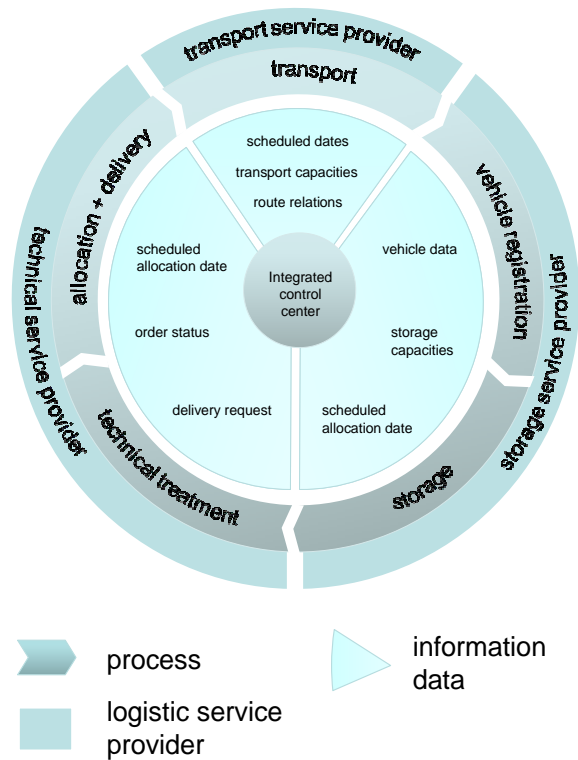
Figure 2: Integrated information management by means of an integrated control centre

The contributions can be described as follows:

- The transport service provider is responsible for transport of the cars to the customer. Relevant process planning and control data of the transport provider includes route relations (starting points and end points), transport capacities of lorries, and scheduled vehicle transport and delivery dates.
- The technical service provider is responsible for technical treatment and modifications of the cars such as e.g. coating or mounting special equipment. Relevant process planning and control data of the technical service provider includes delivery requests, order status as well as scheduled allocation dates of vehicles.
- The storage service provider is responsible for registration and storage of vehicles at the automobile terminals area. Relevant process planning and control data of the storage service provider contains vehicle data (such as vehicle identification number, vehicle orders), storage capacities as well as dates of vehicle allocation.

This way transparency of the planning relevant information will be increased for all process executors, who will be able to communicate their respective process requirements to the other parties involved and thus influence planning of the operational process

execution. This will increase their benefit as well as contribute to improved over-all process execution effectiveness.

The collaborative process planning and control method will be described in the next section.

4. DESCRIPTION OF THE METHOD

Figure 3 outlines the activity chain of collaborative process planning and control of automobile terminal processes.

As already explained, the method involves the described roles of an automobile terminal storage provider, an independent technical service provider situated at the terminal and an automobile transport provider delivering the automobiles to the customers.

As shown in Figure 3, the collaborative planning and control method for the integrated terminal process consists of number sequential and parallel planning and information transfer steps, which are periodically performed by the respective process executors in a timely and logical sequence during both the process planning and execution phase.
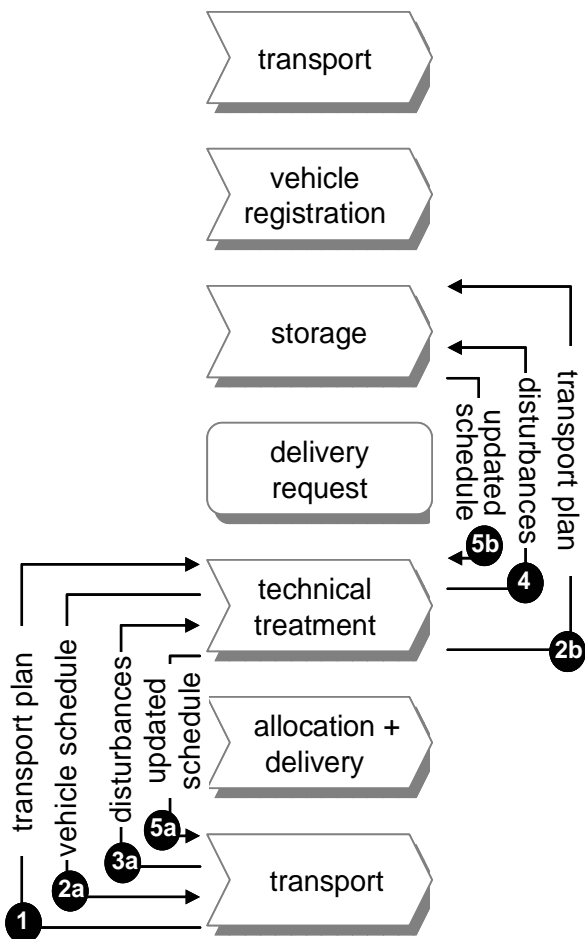
Figure 3: Activity chain of collaborative terminal process planning and control

These steps effect in the creation of mutually adjusted transport plans, technical treatment plans and storage allocation plans. The actions performed during the steps can be de-scribed as follows:

- Step 1: The car transport provider transfers information on his preliminary transport plan data to the technical service provider situated at the automobile terminal. This information includes mainly data on transport route relations (starting and end points of transport movements), data on the capacities and a rough schedule of transport carrier movements. This data determines the planned magnitude of the vehicles to be trans-ported during the planning period. The technical service provider receives this data and uses it for his own process planning, which will take into ac-count the data from the transport provider.

- Step 2a): After the technical service provider has created his own draft technical service plan for the planning period, he sends the sequence of technical service orders and scheduled treatment dates and times as of this plan back to the transport ser-vice provider, who uses the data to modify and re-fine his transport plan.

- Step 2b): Parallel to step 2a the technical service provider sends the data on the scheduled technical treatment dates of the vehicles to the storage ser-vice provider, who takes this data into account for creation of his own vehicle allocation schedules. Steps 2a and 2b and the respective actions taken by the transport and storage provider finish the planning and scheduling process.

- Step 3a: During the execution phase the transport service provider immediately sends any information on disturbances of his own transports like e.g. delays or breakdowns of transport vehicles to the technical service provider, who adapts his own schedules accordingly. If e.g. transport of several vehicles is announced to happen later than originally scheduled, technical treatment of these vehicles may be delayed and more urgent technical treatment orders may be brought forward in the sequence.

- Step: 4: The technical service provider in turn in-forms the storage provider on all disturbances of his own process execution. These may include e.g. changes in his own order sequence and schedules due to disturbances of the transport service provider or delays in the technical treatment processes, e.g. by machine failure.

- Step 5a: Based on the disturbance reports of the transport service (step 3a) provider the technical service provider adapts his own schedules and re-ports the modified schedules of the treatment orders back to the transport service provider who in turn can also adapt his own transport schedules.

278

- Step 5b: Based on the disturbance reports of the technical service provider the storage service provider adapts his own vehicle allocation schedules and sends the modified schedules back to the technical service provider, who can in turn modify his schedules accordingly.

5. EFFECTS

Employment of the method will increase transparency of the planning relevant information for all process executors, who will be able to communicate their respective process requirements to the other parties involved and thus influence planning of the operational process execution. This will increase the benefit of their operations as well as contribute to improved over-all process execution effectiveness.

Bundling of the transport orders already during order processing planning of the automobile terminal would make execution of the transport more economical. Better utilisation and efficiency of the transport carrier space capacities and a reduction of the transport cycle rates and acquisition effort for load amendments could be achieved.

Automobile terminal providers in their function as technical modifiers of the vehicles would profit from lower capital commitment costs due to reduced storage inventories and from lower vehicle processing times by preventing sequence errors or modifications at the technical stations. Across the process chain as a whole, adherence to delivery dates can be increased.

Application of the method is expected to result in the following potential improvements:

- Capacity increase of the automobile terminal (number of orders processed) due to more efficient use of resources (number of transport means, technical stations, personnel etc.).
- Reduction of total process costs of the whole terminal process chain.
- Reduction of the total vehicle order execution time from storage to delivery.
- Increase of adherence rates to delivery dates and prescribed maximal order execution times.
- Reduction of required disposition areas and carrier lorry loading times,
- Increase of carrier lorry utilisation rate and reduction of rate of empty drives.
- Reduction of traffic load at the port terminal and between port and hinterland.

Increased process security due to improved planning and control of order execution along the whole vehicle logistic process chain is regarded as an important additional criterion for success.

6. CONCLUSION AND OUTLOOK

The concept of an integrated planning and control strategy and its effects on logistic process quality will be validated by execution of a computer based simulation of a process scenario along the whole logistic process chain, thus going beyond the limits of one service provider's process execution.

The scenario will be based on real terminal processes, as performed at the automobile terminal at the sea port Bremerhaven, including initial transport, storage, handling, technical stations and distribution, performed by a terminal service provider (BLG LOGISTICS AUTOMOBILE), a specialised car storage service provider (EHH AUTOTEC) and a transport carrier (EHH Automobile-Logistics).

The simulation study will simulate and compare three different process scenarios. Vehicle order management will be modelled (a) as practised today, and be compared with (b) locally optimized processes and (c) an optimized, integrated management across the complete process chain, with all significant interests and requirements of the different parties considered.

Vehicle processing from the time of arrival on the terminal via storage and technical treatment at the terminal to delivery of the modified vehicles on the consignment areas will thus be scheduled with full consideration of estimated processing times, available buffer times until the order delivery date, and available technical and transport capacities and constraints.

As a tool for the simulation study, discrete event simulation will be used, which allows modelling of vehicles, storage areas and storage positions, delivery areas, technical treatment stations as separate entities. Such discrete event simulation models of automobile terminals have al-ready been used in simulation studies to compare the effects of different storage allocation strategies on vehicle transfer times on the terminals (Böse and Windt07).

The procedure for creation of the simulation model and calculation of the simulation results includes definition of the simulation scenarios, collection and preparation of the simulation input data, which will be taken from the real-world process data of the service providers, model creation, model validation, definition and execution of the simulation runs, and documentation and analysis of the simulation results.

Based on the validation by the simulation results, the planning and control method will be prototypically implemented as a software control centre. This control centre will provide a real-time process and status overview with capabilities for regulatory operations provided for all process executing parties mentioned in the reference scenario.

The proposed support method for planning and control of terminal and transport logistics services in vehicle logistics will allow to increase logistic performance.

REFERENCES

Böse, F., Windt, K., 2007: Autonomously Controlled Storage Allocation on an Automobile Terminal. In: Hülsmann, M.; Windt, K., eds. *Understanding Autonomous Co-operation and Control in Logistics – The Impact on Management, Information and Communication and Material Flow.* Berlin:Springer, 351-363.

Kurbel, K., 1999: *Produktionsplanung und –steuerung. Methodische Grundlagen von PPS-Systemen und Erweiterungen.* München, Wien:Oldenbourg.

Mattfeld, D. C., 2005: Development of Vehicle Transshipment at European Ports. Proceedings of the 16th Mini - EURO Conference and 10th Meeting of EWGT, pp. 379-385. September 13-16, Poznan (Poland).

Mattfeld, D. C., 2006: *The management of transshipment terminals: decision support for terminal operations in finished vehicle supply chains.* New York (New York, USA):Springer.

N.N., 2007: *Jahresbericht 2007.* Verband der Automobilindustrie e.V. Available from: http://www.vda.de/de/ ser-vice/jahresbericht/files/VDA_2007.pdf [accessed 2 June 2008].

AUTHORS BIOGRAPHY

Prof. Dr.-Ing. **Bernd Scholz-Reiter** is Managing Director of BIBA - Bremer Institut für Produktion und Logistik GmbH and professor at the University of Bremen, Germany. His research includes production planning and control and logistics.

Dipl.-Wirtsch.-Inf. **Felix Böse** works as a research scientist at BIBA – Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Germany. His research emphasis are modelling of production and logistics systems, production planning and control as well as the application of innovative information and communication technologies to logistic processes.

Dipl.-Wirtsch.-Ing. **Michael Teucke** works as a research scientist at BIBA – Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Germany. His research emphasis are planning and control of production, transport and logistics systems as well as the application of innovative information and communication technologies to logistic processes.

Dipl.-Inf. **Jakub Piotrowski** works as a research scientist at BIBA – Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Germany. His research emphasis are modelling of production and logistics systems as well as the application of innovative information and communication technologies to logistic processes.

PROJECT OF AN AGV TRANSPORT SYSTEM THROUGH SIMULATION TECHNIQUES

V.Duraccio, D. Falcone, A Silvestri, G. Di Bona
Department of Industrial Engineering University of Cassino (Italy)
duraccio@unicas.it

ABSTRACT

The aim of the present work is the application of simulation techniques to the project of the handling system of a semiautomatic line for the assembly of electrical switches. In particular, we have studied an Automatic Guided Vehicle (AGV) transport system. The maximum assembly capacity is 48 units per hour:
- Switches single-pole: 25%, 12 units;
- Switches 2 poles: 25%, 12 units;
- Switches 3 poles: 50%, 24 units.

A main line has been designed in order to connect the following stations:
• An input station for the loading of components for the assembly of each kind of product;
• Two machines for the preparation of subgroups;
• A machine characterized by two stations: assembling and testing of all products (physical, electrical and magnetic);
• Two final warehouses; one of finished products and another of products sent back to the testing station by an operator (inspector), through an handling system. We have a worker in each working centre.

1.1 The internal transport system

We have considered a particular type of AGV: ROBOMATIC 500p. This vehicle is a trolley with 3 wheels, wire-guided, suitable for handling loads up to 300 Kp and equipped with a battery to feed the onboard computer and the transfer along the direction of an inductive circuit (FL circuits). It is possible to recharge the battery through the contacts placed under the chassis at the charging stations. The plant is controlled by a Layout Controller: LC2-40/2/1. The data transfer between controller and vehicles, is realized through fixed installations on the floor, the signal frequency modulation FE (6.3 KHz) and the vehicle transmitter AW. There are two stop-points where data about path and destination are transferred to the vehicle, able to communicate its presence, its identification number and the battery recharging conditions. Each trolley, equipped with rollers, permits to load and unload the trays containing the assembly kits, from the warehouse to the assembly area.In order to reduce the problem of balancing the production line and to maximize the worker saturation, a job-shop layout has been chosen. The signals received by each vehicle, are converted by the onboard computer into driving instructions. Starting from the warehouse, the AGVs are able to recognize the workstation for unloading the assembly kit, thanks to a particular code located on the floor. Arrived at its destination, the trolley stops with a precision of ± 25 mm.

1.2 Detailed Design

Some authors propose very approximate methods to value the needed number of AGVs; often, these deterministic evaluations don't work correctly. In the following, a particular formulation is used:

$$C = \sum i \frac{Ni \cdot ti}{ni}$$

Where:
C = number of trolleys needed;
Ni = number of load units per hour for each item i;
ni = number of load units carried by a trolley;
ti = trolley transport time, considering its speed and the particular path followed (curves, crossings, arrests, loading and unloading of load units). In the case studied, the number of trolleys calculated is 5.

1.3 Simulation Model

The above deterministic methods are very uncertain, so it is useful their validation through computer simulation. It is necessary to realize a plant model, including all data concerning the geometry of the system and the functional properties of each elementary layout part. The model of the analysed system is shown in the below figure:
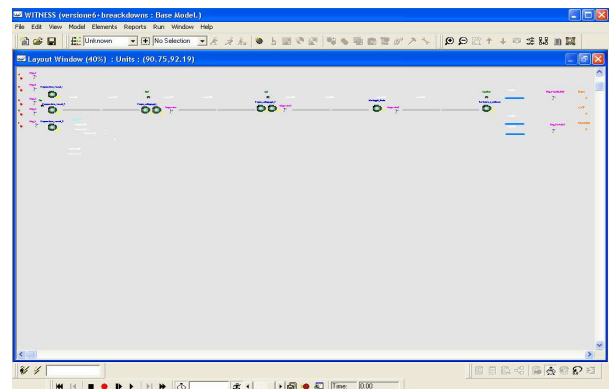
We have used the simulation software "Witness", Release 2.

In order to understand the plant model, we describe the main characteristics of each workstation:

• At the beginning of the production line, there are five buffers to storage five different components. Subsequently, in the first workstation, there are three machines to assembly the different trays of switches, characterized by the same box, called component C. To simplifying the model comprehension, the trays characterized by a different switch production cycle, are marked with different colours: green for single-pole, blue for two-poles, grey for three-poles.
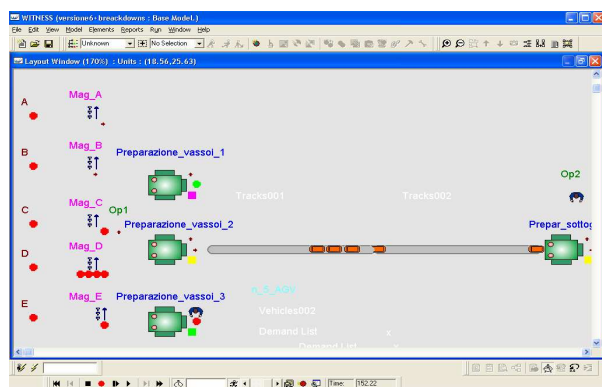
The AGV path starts from the exit of the above machines, where they load and unload the trays.

• In the second e third position, we have two identical workstations, both with two machines to assembly subgroups. Only for the first workstation, we have set a busy_time breakdown, therefore failures happen according to total working time; therefore we have defined the time between breaks (breakdown interval), through a lognormal distribution with an average value of 60 minutes, and the time of reparation (repair time of 3 minutes), made by a specialized worker.
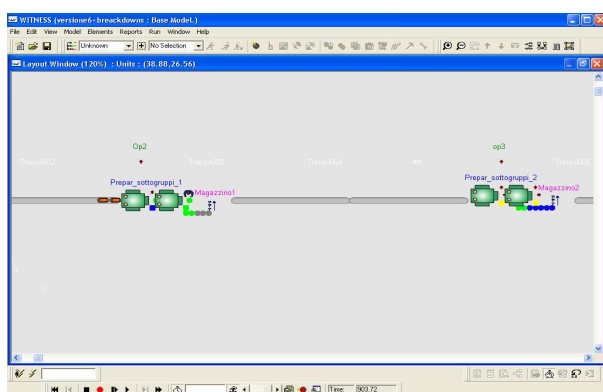
At the exit of each workstation, there is an interoperation buffer (Warehouse 1 and Warehouse 2), useful to separate the different switch production speeds. In the first buffer there are only single and three-poles switches; instead in the second one only single and two-poles switches.

• In the ending part, there is a machine in witch all products are assembled and tested by a worker. At the exit of this workstation there are two different warehouses, one for finished products and another for products sent back by the inspector to the testing station and finally to the warehouse of finished products. To realize the above diversification, in the machine form we have set a particular output rule, called "percent".
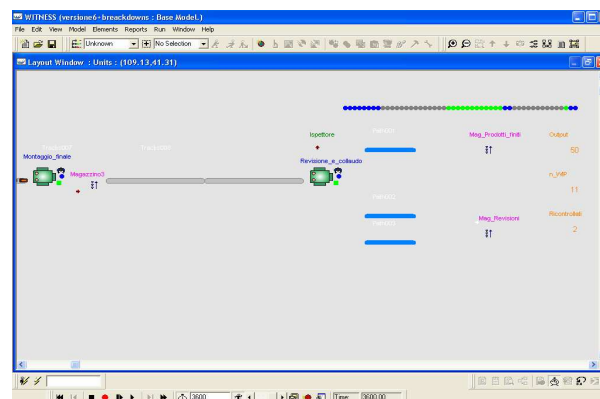
Statistical Reports

Running the simulation, after 60 minutes, we observe 50 finished products realized: 13 single-pole switches, 12 two-poles, 25 three-poles; according to the real production capacity, in spite of the introduction of failure events. So, thanks to simulation, we can verify the correctness of the number of AGVs, previously valued with the analytical method. The below diagram shows the data about production output and work in progress (WIP):

	0	5	10	15	20	25	30	35	40	45	50	55	60
— Output	0	0	0	3	9	14	20	26	29	34	38	44	50
— n_WIP	0	14	27	32	26	34	27	21	25	20	16	17	11
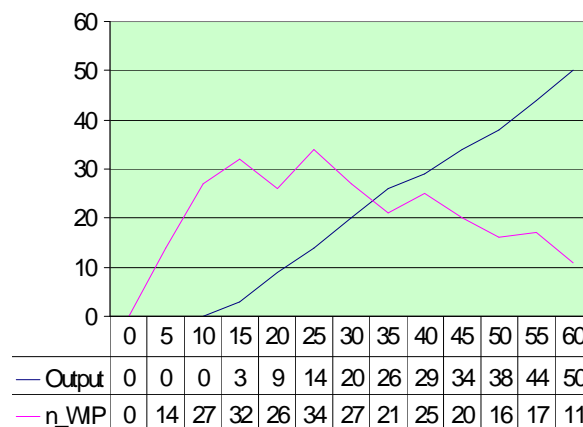
Then, we report statistical data relating to each machine and trolley, including delays and failures, in order to suggest possible improvements of the handling system.

- First workstation:

Machine	%free	%busy	%stop	n. operations
1	19.37	5.33	71.25	20
2	24.41	3.33	68.63	14
3	20.34	6.00	69.82	27

- Second workstation:

Machine	%free	%busy	%stop	%breakdown	%reparation	n. operations
1	31.39	36.11	0.00	1.75	0.25	26
2	32.94	31.94	0.00	1.58	1.33	23

- Third workstation:

Machine	%free	%busy	%stop	n. operations
1	46.58	28.06	0.00	29
2	46.58	27.50	0.00	27

- Fourth workstation:

Machine	%free	%busy	%stop	n. operations
1	21.20	71.30	0.00	51

- Fifth workstation:

Machine	%free	%busy	%stop	n. operations
1	70.64	29.36	0.00	50

- AGVs:

Trolley	%free	%called	%transport	%load	%stop	distance	n. loads
1	0.22	44.03	46.69	9.06	42.48	390	5
2	4.23	69.85	19.15	7.36	68.22	388	4
3	2.17	41.26	42.27	14.30	46.90	387	5
4	7.98	10.47	45.91	35.63	44.56	340	5
5	0.73	58.15	38.74	12.97	51.75	337	5

Analysis of Results

Analyzing the statistical reports, we can notice very high stop percentages of the trolleys, causing high values for the first work station too and consequently a reduction of production speed.This problem is mainly due to two factors:
• The single transport path for all trolleys, causing frequent jams of the AGVs;
• Long machine cycle times, causing frequent queues of the AGVs, waiting for loading and unloading products in the workstations.
Both the above causes are non simply to remove, requiring significant economic investments:
- In the first case, it is possible to increase the number of paths for the AGVs, through the creation of new inductive circuits followed by the trolleys;
- In the second case, it is possible to buy new machines of the first workstation, able to reduce cycle times.

Conclusions

The main reason for using computer simulation is the evaluation in advance of changes of the production process, reducing realization time and cost. In the case-study proposed, we have analysed the material handling system of a company assembling electrical switches, using wire-guided vehicles.
Initially, we have calculated the needed number of trolleys through an analytical formulation. Subsequently, we have decided to verify the deterministic result through a simulation model, using a specific software. The simulation software lets us represent graphically the material handling system and confirms a correct choice of the number of AGVs. Subsequently, thanks to the analysis of statistical reports, it was possible to identify the critical points along the production line, characterized by many stops, and to suggest possible solutions to apply.
After the economic evaluation of the technical solutions proposed, consisting of a reduction of machine cycle times and/or an increase of transport paths, it will be possible to create the new model of the production line and to value the productivity increase.

References

[1]. JAMES M. MOORE
"*Progettazione e layout degli* impianti" 1995 3[a] edizione
Mc Graw Hill New York
[2]. TURCO A.
"*Principi generali di progettazione degli impianti industriali*" 1990
Città degli Studi Milano
[3]. WITNESS MANUAL USER, Release 9.10, AT&T ISTEL
[4]. PARESCHI A.
"*Impianti industriali*" 1998
Progetto Leonardo
Pitagora Editrice Bologna

[5]. DE' ROSSI FERDINANDO
"*Corso di impianti meccanici*"Parte prima 1985
Liguori editore Napoli

[6]. T.H. NAYLOR, J.L. BALITY, D.S. BURDICK, K. CHU, 1968,
Computers Simuation Techniques, John W & Sons, New York

[7]. B. MARTINOLI,
Guida alla simulazione: metodi, linguaggi e modelli di ricerca operativa per simulare processi ed eventi con esempi relativi ad applicazioni aziendali 1988, Franco Angeli Editore

[8]. F. CARON,
Introduzione ali linguaggi e modelli di simulazione, in Simulare per decidere meglio, Milano Convegno IRI 6-7 febbraio 1992

[9]. SAIFALLAH BENIAAFAR,
Intelligent Simulation for Flexible Manufacturing System: an Integrated Approach, Computer & Industrial Engineering, vol.22, n°3. 1992,

[10]. P. BRATLEY, B.L. FOX, L.E. SCHRAGE,
A Guide to Simulation, 1983 Springer New York

[11]. LAW M. AVERIL,
*Simulation Modeling Analalysis,*1991*, II edit,* Mc Graw-Hill

[12]. K.D. TOCKER,
The Art of Simulation, 1963 English University Press, London

[13]. MORGAN, B.J.T.,
Elementary Simulation, 1984 Chapman 6 Hall, London

[14]. D. FALCONE, F. DE FELICE
Simulation For The Sizing Of The Buffer In A Manufacturing Company - SCSC 30th Anniversary The Proceedings of the Summer Computer Simulation Conference - Rheno, Nevada - (USA) 19-22 Luglio1998

[15]. D.FALCONE, F. DE FELICE
Operating planning of production in a mechanical industry through simulation techniques - ESM98-12th European Simulation Multiconference – Manchester, United Kingdom – 16-19 Giugno 1998

MODELING AND SIMULATION OF BEHAVIORAL SCENARIOS BY USING COUPLED DEVS MODELS

Sqali Mamoun[(a)], Lucile Torres[(b)]

[(a)] [(b)]LSIS (Laboratoire des Sciences de l'Information et des Systèmes)
Université Paul Cezanne Aix-Marseille III
avenue Escadrille Normandie-Niemen13397 MARSEILLE CEDEX 20

[(a)] mamoun.sqali@lsis.org, [(b)]lucile.torres@lsis.org

ABSTRACT
Sequence diagrams, such as those included in UML, are widely used to express behavioral requirements of a system. They are employed to refine use cases and to induce an abstract model of the system behavior. In the approach we propose, the global behavioral model is obtained from a set of sequence diagrams and it is expressed in the DEVS (Discrete EVent System Specification) formalism. Finally, the DEVS model can be validated and verified by simulation.

Keywords: Sequence Diagrams, DEVS modeling, Simulation.

1. INTRODUCTION

In general, the scenarios use Sequence Diagrams (SD) to represent graphically the interactions between objects. They can be composed by using flow control operators (alternative, sequence, parallelism and repetition) in order to form more complex scenarios. A SD show, as parallel vertical lines, different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur.

This allows the specification of simple runtime scenarios in a graphical manner. It seems illusive, to define the system by conceiving all its scenarios. Moreover, designing the system behavior directly with statecharts is not an intuitive process, since the concept of state is not obvious in the first stages of the development process. Many approaches address the scenario synthesis problem (Liang, Dingel and Diskin 2006) and makes possible to induce a total behavior model expressed in a state machine format starting from a set of scenarios. All of them agree with the need to have a sufficient number of scenarios. They especially differ in the source notation and semantics, in the target notation, in their ability to support scenario composition mechanisms, in their use of merging techniques of identical states, and in the implemented synthesis algorithm. We present in the following sections, the SD notation, the Discret EVent system Specification (DEVS) formalism, and the synthesis problem.

When the synthesis has been realized, the induced coupled DEVS model represents the system behavior that includes all the behaviors expressed by the scenarios and it is used to validate the specification by simulation.

2. RECALL
2.1. Sequence Diagrams

A SD is one of the five diagrams used in UML for modeling the system dynamic aspects of systems. A SD shows an interaction, consisting of a set of objects and their relationships. The diagram includes messages that may be dispatched among them with an emphasis on their time orderings. Graphically, a SD is a table that shows objects arranged along the X axis and messages, ordered in increasing time, along the Y axis.

The vertical lines in the SD represent the lifelines of the objects taking part in the scenario. Time is usually assumed to flow downwards along each lifeline. The directed arrows going across the lifelines represent the causal links from a sent event (the source of the arrow) to the corresponding received event (the target of the arrow), with the label on the arrow denoting the event.

A great number of notations are commonly used for the description of SD: Message Sequence Charts (MSC) defined within an international standard (ITU 2000), that can be distinguished between basic MSC (bMSC) and compound MSC (hMSC for High-Level MSC), Live Sequence Charts (LSC), which are an extension of bMSC proposed by (Damm and Harel, 2001), the UML SD (OMG 2005), which are a simplified version of bMSC (Brian and Hans 2004), …

In this article, we propose a method for translating a set of SD into state machines represented in the formal DEVS specification. We use a simple example to illustrate our approach. This example (see Figure 1) shows how to obtain a coffee from a coffee machine. The SD is composed by three objects: Customer, CoffeeMachine and Stock. Customer is considering as an environment object.

A SD has a structure (E, ≤, α, Φ, A, I, T) where:

- E is a set of events

- ≤ is a partial ordering between events imposed by lifelines and messages
- α is a set of action names
- Φ denotes the location to an event (i.e the object affected by the event)
- A is a set of actions (message sending and receiving);
- I is a set of objects
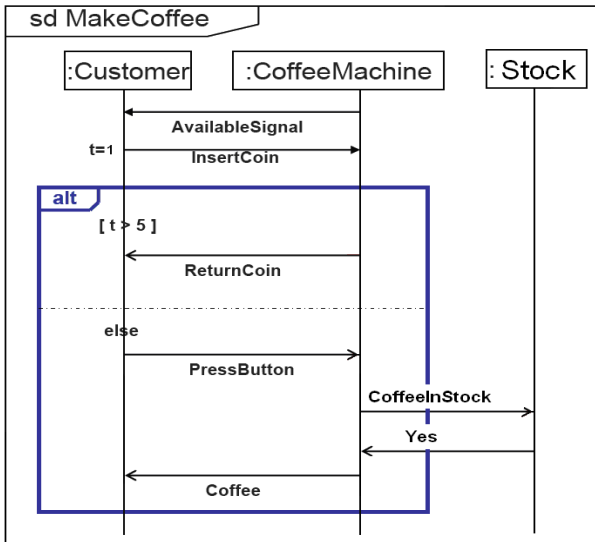- T is a set of interaction time constraints;

Figure 1: A SD for making coffee

The behavior represented by a SD is a set of event sequences determined by the causal priority. The events are totally ordered for one considered object. The causal relationship determines a partial order, noted ≤, on the events between all the objects. This partial order can be derived from the SD in respect with two principal rules:

- an event e drawn higher than another event e' on the same lifeline of an object precedes necessarily e';
- the event associated with a message sending precedes necessarily the event associated with the reception of this message (in the case of an asynchronous communication). For a synchronous communication, messages are used to be considered instantaneous in the order to do not differentiate the message sending from the message reception.

2.2. The DEVS Formalism
The DEVS formalism introduced by Zeigler (Zeigler 2000) provides a means for modeling discrete event system in a hierarchical and modular manner. The use of this formalism facilitates the modeling activity and guarantees a best accuracy of the model by decomposing the system into component models and by specifying the coupling between them. There are two kinds of DEVS models, atomic models and coupled ones.

2.2.1. Atomic DEVS model
An atomic model describes the behavior of a component, which is indivisible, in a timed state transition level. Formally, an atomic model is defined by a 7-tuple <X, Y, S, δint, δext, λ, ta> where:

- X is the set of the external input events;
- Y is the set of the external output events;
- S is the set of sequential states;
- δint : S -> S is the internal transition function that defines the state changes caused by internal events;
- δext : Q × X -> S is the external transition function, where Q = {(s,e)|s∈ S, 0≤e≤ta(s)} is the set of total state; this function specifies the state changes due to external events, with the ability to define a future state according to the elapsed time in the current state;
- λ is the output function that generates output events;
- ta : R+ ∪ {0, ∞} gives the lifetime of the states, where R+ ∪ {0, ∞} is the set of positive reals with 0 and ∞; ta(s) represents the interval during which the model remains in the state s if external event occurs.

While the internal transition function expresses the autonomous evolution of the model, the external transition function defines its evolution when occurring external events.

2.2.2. Coupled DEVS model
A coupled model is a compound component consisting of atomic models and/or coupled models. The coupled model can itself be employed as a component in a larger coupled model, there by giving rise to the construction of complex models with hierarchical structures. A coupled model is formally defined by a 7-tuple < X, Y, M, EIC, EOC, IC, SELECT > where:

- X is the set of input events;
- Y is the set of output events;
- M is the set of all the DEVS component models;
- EIC ⊆ X × UiXi is the external input coupling relation;
- EOC ⊆ UiYi × Y is the external output coupling relation;
- IC ⊆ UiXi × UiYi is the internal coupling relation;
- SELECT: 2M - ϕ, M is a function which chooses one model when more than 2 models are scheduled simultaneously.

EIC, EOC and IC specify the connections between the input and output ports of the various DEVS models.

3. STATE OF THE ART
Many approaches address the scenario synthesis problem, like (Harel, Kugler and Pnueli 2005) who

proposed a synthesis approach using the scenario-based language of Live Sequence Charts (LSC) as requirements, and synthesizing a state-based object system composed of a collection of finite state machines. (Letier, Kramer and Uchitel 2005) who presents a technique to generate Labelled Transition System (LTS) from Hight Level Message Sequence Chart (hMSC), in this approach, complex system behaviour can be modelled by parallel composition of the component LTS models. Parallel composition models components that execute asynchronously but synchronize on shared events; also they presented a technique to detect implicit scenarios.

Ziadi, Hélouët and Jézéquel (2004) have proposed an idea to synthesize statecharts starting from scenarios expressed by UML2.0 Sequence Diagrams, and give an algorithm for synthesizing a composition of statecharts between them. Also (Damas and Lambeau 2006) has presented an approach to generate Lablled Transition System from a collection of basics MSC's, and use a technique to merge the identical states. That will help us for our approach, and will return our atomic DEVS models minimal and determinists.

4. FROM SEQUENCE DIAGRAMS TO DEVS
4.1. Transformation steps of SD without operators
Let us show transformation from SD sets to DEVS in the five following steps.

Step I) Constructing an atomic DEVS model for one object belonging to one scenario:

1. For one object in one SD, identifying an atomic DEVS model (We consider for this example, the CoffeeMachine object).

2. Defining input/output events of the atomic DEVS model (X, Y)

3. Defining a set of states. A state must include information of input/output events in a SD. As shown in the figure 2, each state should be associated with both an external and/or internal transitions (δint, δext) in an atomic DEVS model.
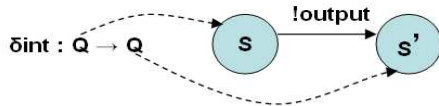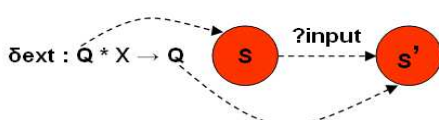
Figure 2: Definition rules of the set of states

4. Defining state transitions between states obtained in 3 from the SD:

A. Showing message transmission among the defined states (figure 3).
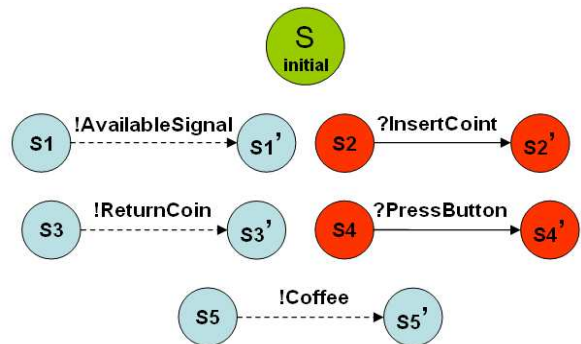
Figure 3: Translating message transmission

B. Merging the start point of the SD and its end point into the initial state of the DEVS model (figure 4).
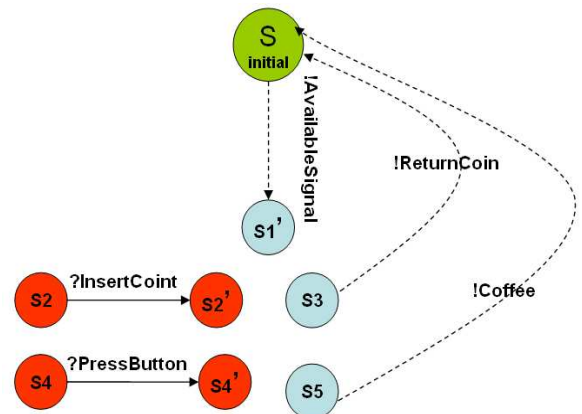
Figure 4: Defining the initial state

C. Translating all the message sequences of the SD to state transitions of the atomic model, and minimizing the number of states if equivalent states exist.
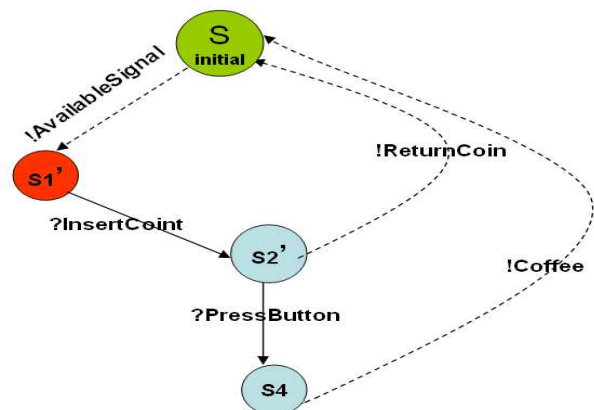
Figure 5: Minimizing the number of states

D. After the previous steps A~C, the specification of time advance functions in atomic DEVS models is remained. If a special time constraint is defined in the SD, the

condition must be translated in the confirmed state. If only an input causes the state transition from a state s, it can be describe by a passive state with ta(s)= ∞. If a state s' describes only output events, then ta(s')=1.

Step II) Repeating the first step (I) for all objects of the considered SD. This step generates the set of atomic models associated with the SD.

Step III) Repeating the step II for all the SD. This step generates the set of atomic models associated with the set of SD.

Step IV) Building one atomic model from the atomic models on a given object. This step generates as many atomic models as there are objects in the set of SD.
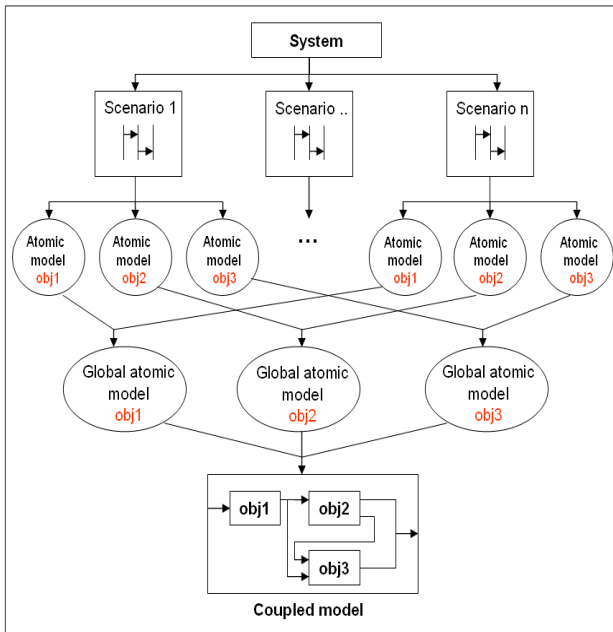
Figure 6: Example of the steps IV and V

Step V) Building a coupled final model which is composed of the atomic models obtained in the previous step, so that the output of an atomic model is the input of another, then all objects in the system can communicate between them. The final coupled model describes the overall behavior of the system.

An algorithm which translates a basic SD into an atomic DEVS model by object was presented in (Sqali and Torres 2008a). Also, the principle of the construction of an atomic DEVS model by object starting from several composed SD, and the construction of coupled DEVS model is described in (Sqali and Torres 2008b).

4.2. Transformation of SD with seq, loop and alt operators

▪ Seq specifies a sequence between two SD (strong sequential composition).

Da1 seq Da2 = <X, Y, S, δint, δext, λ, ta>:
- S =- S1 ∪ S2- {s02} if (Da2 ≠ ∅)
  - S2 if (Da1= ∅)
- X = X1 ∪ X2
- Y = Y1 ∪ Y2
- δint = δint1∪ δint1
- δext = δext1∪ δext2

▪ Loop specifies an iteration of a SD

Loop (Da1) = <X, Y, S, δint, δext, λ, ta>:
- S = - (S1-sn1) ∪ {s01}
  - s0=s01
- X=X1
- Y=Y1
- δint = δint1∪ {sn -> s0} ∨ δext = δext1∪ {Q ×X -> s0}
- Loop (Da∅) = Da∅

▪ Alt defines a choice between a set of

Da1 alt Da2 = <X, Y, S, δint, δext, λ, ta>:
- S =- S1 if (Da1 ≠ ∅ ∧ Da2 = ∅)
  - S2 if (Da1 = ∅ ∧ Da2 ≠ ∅)
  - {s0} if ((Da1 = ∅ ∧ Da2 = ∅)
  - S1 ∪ S2 ∪ {s} if (Da1 and DA2 are loops) ∧ (Da1 ≠ ∅ and Da2 ≠ ∅)
  - S1 ∪ S2 otherwise
- s0 =- A new state if (Da1 and DA2 are loops) ∧ (Da1 ≠ ∅ and Da2 ≠ ∅)
  - s01 if Da2 are loops
  - s02 if Da1 are loops
- X = X1 ∪ X2
- Y = Y1 ∪ Y2
- δint = δint1∪ δint1
- δext = δext1∪ δext2

5. CASE STUDY AND SIMULATION

Let us consider SD of the figure 1 that contains three objects; Customer, CoffeeMachine and Stock. By using the previous steps of transformation into DEVS models, we first obtain the DEVS atomic models represented in the figures 6 and 7.
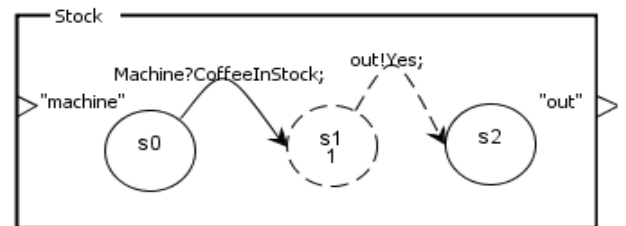

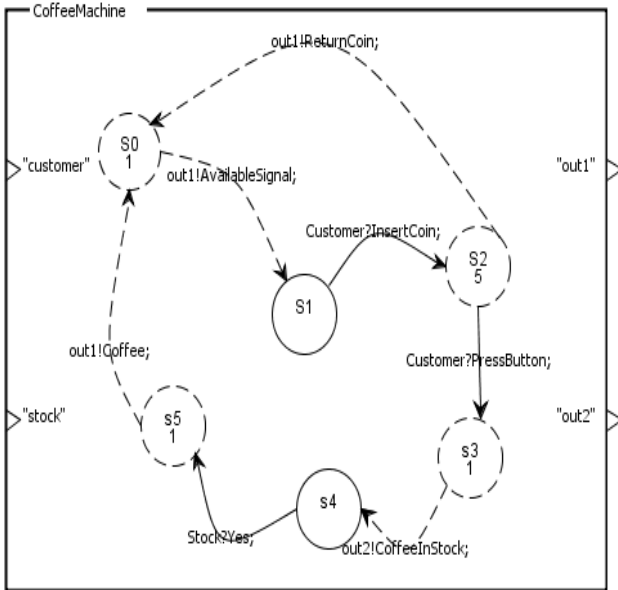Figure 6: An atomic DEVS model for the object Stock

288

Figure 7: An atomic DEVS model for the object CoffeeMachine

After the global atomic models for all objects of the system have been built, we construct the coupled model given in the figure 8 who describes the overall behavior of the system.
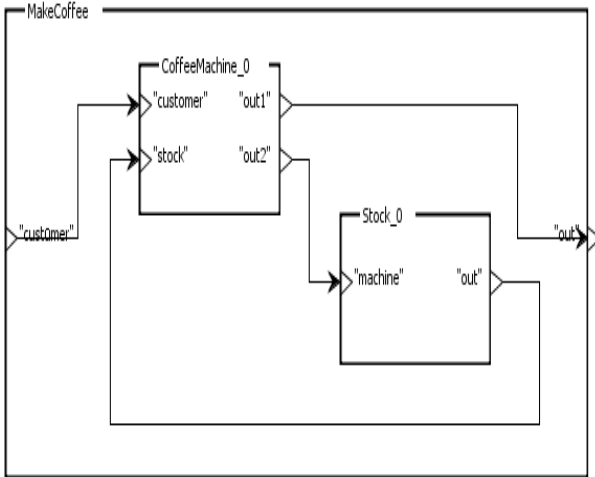

Figure 8: A final coupled model

To validate the specification of the system behavior from the final coupled DEVS model, we use the LSIS-DME tool developed within our laboratory. From a model and a set of data, the simulator provides the simulation results. The dataset is defined by the user who enters all external events supposed to occur during the simulation. For that, he entered for each event the port of the model on which the event will occur and the value of the event. For the final model, with the inputs specified in the figures 9 and 11, the results obtained during the simulation are given in the figures 10 and 12 respectively. The simulation of the global DEVS model produces a sequence of events for the considered sequence of input events belonging to the set of event

sequences defined by the SD. We note that the behaviors obtained are exactly the same as those specified in the SD.
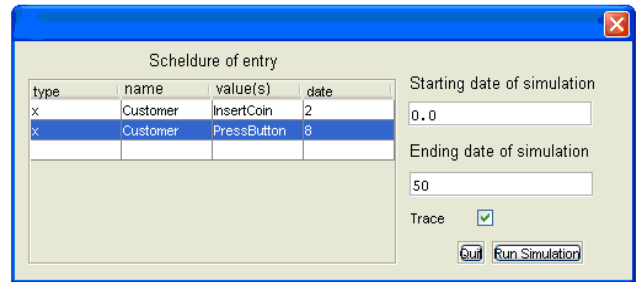
Figure 9: Filling input schedules
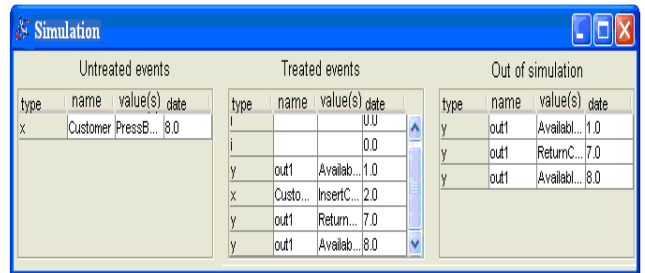
Figure 10: Simulation results for Figure 9
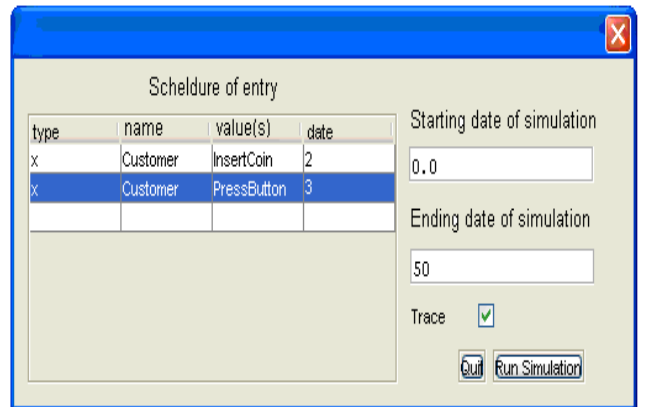
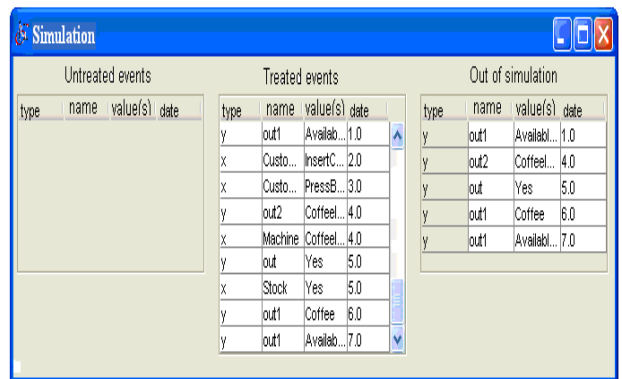Figure 11: Another filling input schedules

Figure 12: Simulation results for Figure 11

The LSIS_DME tool also provides an interface XML to describe models DEVS textually. This makes possible to transform models DEVS obtained under

format XML. We can thereafter modify them and simulate them. For this, we propose to use XML and XSL to realize the synthesis in models DEVS expressed in XML
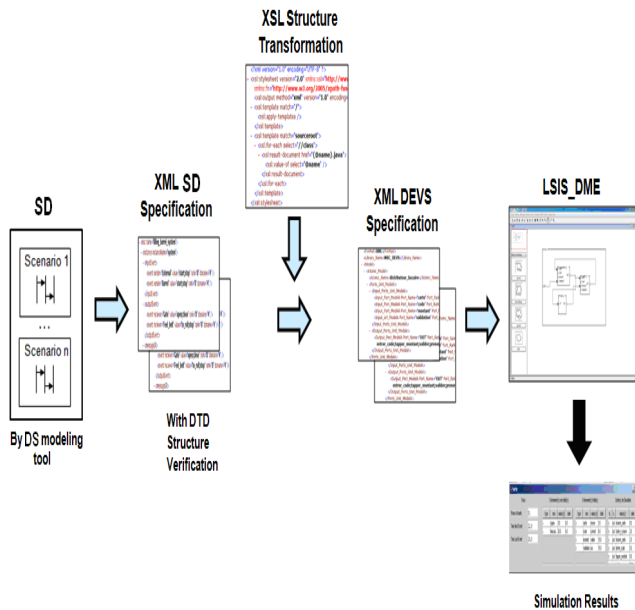

Figure 13: Transformation using XML

The transformation procedure consists to:

1) Generate a XML representation of an atomic DEVS model for each object in all the scenarios of the system, by defining the XML format of the scenarios class with a DTD grammar, to apply XSL language on XML representation of the scenarios models, and generate a XML document for each object of the system.
2) Construct a final coupled DEVS model, composed of the various atomic models obtained, describing the total behavior of the system.

By using the library saxon8.jar, we import the javax.xml.transform.*; javax.xml.transform.stream._ Stream_Result; and javax.xml.transform.stream._ StreamSource;. The following code creates an XSL transformer which has as parameters the path of scenario XML document which we want to transform, and generates a DEVS XML file.

```
// Create a transform factory instance.
TransformerFactory tfactory = TransformerFactory_
.newInstance();
// Create a transformer for the stylesheet.
Transformer trans = null;
    try {
        trans     =     tfactory.newTransformer(new
        StreamSource("C:/Path/convert.xsl"));
    }
    catch (TransformerConfigurationException ex) {
// Transform the source XML to System.out.
    try {
        trans.transform(
        new StreamSource("C:/Path/scenario.xml), new
        StreamResult(                              new
        File("exampleSimple.out")));
    }
    catch (TransformerException ex)
```

6. CONCLUSION

Many approaches address the scenario synthesis problem. All of them agree with the need to have a sufficient number of scenarios. They especially differ in the source notation and semantics, in the target notation, and the identification of common states for merging scenarios (Liang, Dingel and Diskin 2006).

We have presented the Sequence Diagrams synthesis into a DEVS model in order to verify and validate by simulation the behavioral specification given by the scenarios. We have chosen a scenario semantics restricted to event sequences with the notion of (repetition, alternativity and sequence). Thereafter, we proposed to use XML and XSL to realize the synthesis in DEVS models expressed in XML (José, Risco-Martin and Mittal 2007).

REFERENCES

Brian, L., Hans, E., 2004. UML 2 toolkit. *Whiley Publishing OMG press*
.
Zeigler, B., Herbert, P., Tag Gon, K., 2000. Theory of Modeling and Simulation. *ACADEMIC PRESS.*

Damas, C., Lambeau, B., Lamsweerde A., 2006. Scenarios, Goals, and State Machines: a Win-Win Partnershi for Model Synthesis. *14th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pp. 197-207. Portland, Oregon, USA.

Damm, W., Harel, D., 2001. LSCs: Breathing life into message sequence charts. *Formal Methods in System Design*, 19(1):45--80.

David, H., Kugler, H., Pnueli, A., 2005. Synthesis Revisited: Generating Statechart Models from Scenario-Based Requirements. *Formal Methods in Software and Systems Modeling.*

Letier, E., Kramer, J., Magee, J., Uchitel, S., 2005. Monitoring and Control in Scenario-Based Requirements Analysis**.** *International Conference on Software Engineering, Proceedings of the 27th international conference on Software engineering.*

José, L., Risco-Martin, Mittal, S., 2007, *A W3C XML Schema for DEVS Scenarios*, submitted to DEVS Integrative M&S Symposium DEVS' 07, Spring Simulation Multi-Conference, March.

ITU, 2000. Message Sequence Charts. Recommendation Z.120. *International Telecommunications Union*. Telecommunication Standardisation Sector.

Liang, H., Dingel, J., Diskin, Z., A, 2006. Comparative Survey of Scenario-based to State-based Model Synthesis Approaches. *SCESM'06 : International Workshop on Scenarios and State Machines: Models, Algorithms, and Tool*, pp.5-12. May, Shangaï, China.

OMG, 2005. *UML 2.0 Specification*. Object Management Group. Avaibale from: http://www.omg.org [August 2005].

Sqali, M., Torres, L., Frydman, C., 2008. Synthèse de scénarios en DEVS (a), *7ème conférence internationale de Modélisation et SIMulation (MOSIM08)*. Paris, 31 mars-2 avril, .

Sqali, M., Torres, L., Frydman, C., 2008. Synthetizing scenarios to DEVS models (b). *Poster Session of SpringSim08*. Ottawa,Canada.

Ziadi, T., Hélouët, L., Jézéquel, J., 2004. Revisiting Statechart Synthesis with an Algebraic Approach. *Proc. of 26th International Conference on Software Engineering (ICSE), IEEE Computer Society*.p. 242-251.Edinburgh, May.

WSDL-BASED DEVS AGENT FOR NET-CENTRIC SYSTEMS ENGINEERING

Saurabh Mittal[a], Bernard P. Zeigler[b], Jose L. Risco Martin[c], Jesús M. de la Cruz[c]

[a]DUNIP Technologies, Inc. New Delhi, India
[b]Arizona Center for Integrative Modeling and Simulation
ECE Department, University of Arizona, Tucson, AZ USA
[c] Departamento de Arquitectura de Computadores y Automática
Facultad de Informática, Universidad Complutense de Madrid, Madrid, Spain

[a]saurabh.mittal@duniptechnologies.com, [b]zeigler@ece.arizona.edu, [c] jlrisco@dacya.ucm.es, jmcruz@fis.ucm.es

ABSTRACT

This research work provides a methodology to use Discrete Event Systems Specification (DEVS) to design and evaluate the performance of web services within a Service Oriented Architecture (SOA). We will show how a Web Service Description Language (WSDL) document can be mapped to a DEVS model in an automated manner through a DEVS abstract service wrapper. This work will describe the underlying architecture of abstract DEVS service wrapper and a workflow example made executable using the DEVS/SOA framework. This work will establish DEVS as a production environment for net-centric systems as well as a solid M&S analysis tool for SOA design.

Keywords: SOA, WSDL, DEVS/SOA, DEVSML

1. INTRODUCTION

Industry and government are spending extensively to transition their business processes and governance to Service Oriented Architecture implementations for efficient information reuse, integration, collaboration and cost-sharing. Service Oriented Architecture (SOA) enables orchestrating web services to execute such processes using Business Process Execution Language (BPEL). Business Process Modeling Notation (BPMN) is another method that outputs BPEL for deployment. As an example, the Department of Defense's (DoD grand vision is the Global Information Grid that is founded on SOA infrastructure. As illustrated in Figure 1, the SOA infrastructure is to be based on a small set of capabilities known as Core Enterprise Services (CES) whose use is mandated to enable interoperability and increased information sharing within and across Mission Areas, such as the Warfighter domain, Business processes, Defense Intelligence, and so on) (DoD GIG Architecture, 2007). Net-Centric Enterprise Services (NCES) is DoD's implementation of its Data Strategy over the GIG. NCES provide SOA infrastructure capabilities such as service and metadata registries, service discovery, user authentication, machine-to-machine messaging, service management, orchestration, and service governance.

However, composing/orchestrating web services in a process workflow (a.k.a Mission thread in the DoD domain) is currently bounded by the BPMN/BPEL technologies. Moreover, there are few methodologies to support such composition/orchestration. Further, BPMN and BPEL are not integrated in a robust manner and different proprietary BPMN diagrams from commercial tools fail to deliver the same BPEL translations. Today, these two   technologies is by far the only viable means whereby executives and managers can devise process flows without touching the technological aspects. With so much resting on SOA, their reliability and analysis must be rigorously considered. The BPMN/BPEL combination neither has any grounding in system theoretical principles nor can it be used in designing net-centric systems based on SOA in its current state.
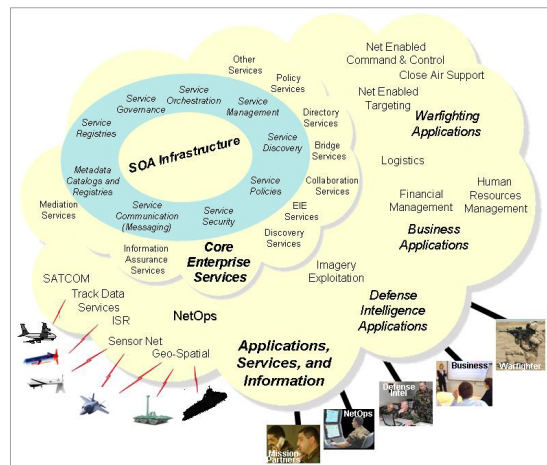

Figure 1: Core Enterprise Services in GIG

In this research work we provide a proof of concept of how Discrete Event System Specification (DEVS) Formalism can deliver another process work flow mechanism to compose web services in a SOA. We will show how the resulting agent based DEVS system can be executed on the recently developed DEVS/SOA (DUNIP, 2007; Mittal, Martin and Zeigler, 2008) distributed modeling and simulation framework. In addition to supporting SOA application development the framework enables verification and validation testing of application. The developed DEVS models from WSDL lie in the subset of DEVS specifications known as Finite Deterministic DEVS or FDDEVS (Hwang and Zeigler, 2006; Mittal, Zeigler and Hwang, 2007) that can be used for verification. However, V&V is not the focus of this paper. We will demonstrate the execution of the DEVS agent in a complete case-study in which a workflow is composed and executed using

DEVS/SOA framework. The paper is organized as follows. Section 2 presents the related work in the area of BPEL, BPMN and Agent based studies focused towards SOA. Section 3 describes the underlying technologies that include DEVS, Web Services framework. Section 4 deals with Abstract DEVS Service wrapper in detail and also discusses how statistics gathering is integrated with the wrapper design. Section 5 presents the implementation of DEVS WSDL agent and how it can be used in a process workflow using the proposed Web Services Workflow Formalism, WSWF. Section 6 presents layered architecture of Agent-based Test Instrumentation System on/using Global Information Grid using SOA (GIG/SOA) that provides a larger perspective on the application of DEVS-WSDL agent architecture. Finally, Section 7 presents conclusions and future work.

2. RELATED WORK

In 2003 there were more than 10 recognized groups defining standards for BPM related activities, 7 of these bodies were working on modeling definitions so it's no wonder that the whole picture got very confused (Pyke, 2007) Fortunately there has been a lot of consolidation, and currently only 3 key standards to really take notice:

1. BPMN
2. XPDL
3. BPEL

The Business Process Modeling Notation (BPMN) is a standardized graphical notation for graphically representing business processes workflows. BPMN's primary goal is to provide a standard notation that is readily understandable by all business stakeholders. Stakeholders in this definition include business analysts, technical developers and business managers. BPEL is an "execution language" the goal of which is to enable definition of web service orchestrations. Ultimately, BPEL is all about bits and bytes being moved from place to place and manipulated. XPDL is described not an executable programming language like BPEL, but specifically a process design format that literally represents the "drawing" of the process definition. XPDL is effectively the file format or "serialization" of BPMN. More generally, it can also support any design method or process model that uses the XPDL meta-model. XPDL is a proven format for process design interchange, and it is the most practical standard for establishing a Process Design Ecosystem.

Summarizing, currently there is no popular means other than BPMN/BPEL to design a web service workflow orchestration.

3. UNDERLYING TECHNOLOGIES

This section will give an overview of the technologies used in the development of DEVS Web service M&S framework.

3.1. DEVS

Discrete Event System Specification (DEVS) (Zeigler, Kim and Praehofer, 2000) is a formalism, which provides a means of specifying the components of a system in a discrete event simulation. In DEVS formalism, one must specify *Basic Models* and how these models are connected together. These basic models are called *Atomic Models* (Figure 2a) and larger models which are obtained by connecting these atomic blocks in meaningful fashion are called *Coupled Models* (Figure 2b). Each of these atomic models has *inports* (to receive external events), *outports* (to send events), set of *state variables*, *internal transition*, *external transition*, and *time advance functions*. Mathematically it is represented as 8-tuple system as eq. (1) below:

$$M = <X, S, Y, \delta_{int}, \delta_{ext}, \delta_{con}, \lambda, ta> \text{ eq. (1)}$$

where,

X is the set of input values
S is the set of state
Y is the set of output values
δ_{int}: $S \rightarrow S$ is the internal transition function
δ_{ext}: $Q \times X^b \rightarrow S$ is the external transition function, where
X^b is a set of bags over elements in X,
Q is the total state set.
δ_{con}: $S \times X^b \rightarrow S$ is the confluent transition function, subject to $\delta_{con}(s,\Phi) = \delta_{int}(s)$
λ: $S \rightarrow Y^b$ is the output function
ta: $S \rightarrow R_0^+{}_{,inf}$ is the time advance function

The model's description (implementation) uses (or discards) the message in the event to do the computation and delivers an output message on the *outport* and makes a state transition.

A DEVS-coupled model designates how atomic models can be coupled together and how they interact with each other to form a complex model. The coupled model can be employed as a component in a larger coupled model and can construct complex models in a hierarchical way. The specification provides component and coupling information. The coupled DEVS model is defined as eq. (2) below:

$$M = <X, Y, D, \{M_{ij}\}, \{I_j\}, \{Z_{ij}\}> \text{ eq. (2)}$$

where,

X is a set of inputs
Y is a set of outputs
D is a set of DEVS component names
For each $i \in D$,
M_i is a DEVS component model
I_i is the set of influences for I
For each $j \in I_i$,

A Java-based implementation of DEVS formalism, DEVSJAVA (Zeigler and Sarjoughian, 2000) can be used to implement these atomic or coupled models.

DEVS formalism consists of models, the simulator and the Experimental Frame as shown in Figure 3. It categorically separates the three of them and they can be perceived of components of a DEVS system architecture. We will focus our attention to the two types of models i.e. atomic and coupled models.
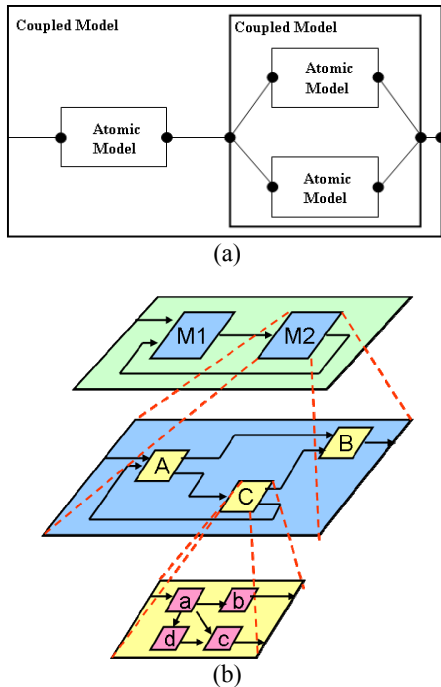

(a)

(b)

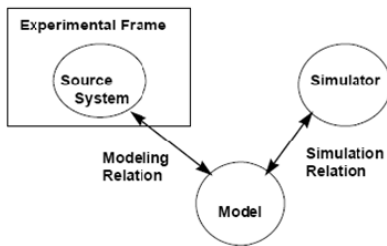Figure 2: Hierarchical composition of Atomic and Coupled DEVS models.

Figure 3: DEVS Separation of Model, Simulator and the Experimental Frame

3.2. Web Services and Interoperability using XML

The Service oriented Architecture (SOA) framework is the orchestration of multiple web services engaged towards a business goal. A Web Service is a component consisting of various W3C standards, in which various computational components are made available as 'services' interacting in an automated manner that achieve machine-to-machine interoperable interaction over the network. The interface is specified using Web Service Description language (WSDL) that contains information about ports, message types, port types, and other relating information for binding two interactions. It is essentially a client server framework, wherein client requests a 'service' using a SOAP message that is transmitted via HTTP in the XML format. A Web service is published by any commercial vendor at a specific URL   to be consumed/requested by another commercial application on the Internet. It is designed specifically for machine-to-machine interaction. Both the client and the server encapsulate their messages in SOAP wrappers.

The fundamental concept of web services is to integrate software application as services. Web services allow the applications to communicate with other applications using open standards. To offer DEVS-based simulators as web services, they must have the following standard technologies: communication protocol (Simple Object Access Protocol, SOAP), service description (Web Service Description Language, WSDL), and service discovery (Universal Description Discovery and Integration, UDDI).

4.   AN ABSTRACT DEVS SERVICE AGENT

As a crucial part of our workflow, we have designed an Abstract DEVS Service Agent to link DEVS models with Web Services and to generate statistics regarding remote method calls and response times.
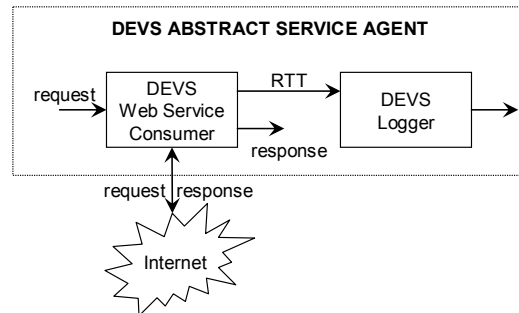

Figure 4: Schematic showing the architecture of our DEVS Agent Service model.

Figure 4 depicts an illustrative example. Our proposed model consists of two DEVS atomic models. The DEVS Web Service Consumer invokes the remote operation provided by means of an external transition. When the operation is processed, this atomic model informs about the round-trip-time (RTT) taken by such operation to the DEVS Logger atomic model as well as the response provided by the Web Service. At the end of the simulation, the DEVS Logger provides statistics such as operations executed successfully, the RTT consumed per operation, etc.

The DEVS Web Service Consumer needs to be configured by means of: (a) the URL of the Web Service, (b) name of the operations offered, and (c) the parameters needed by such operations. This information is specified in the WSDL document.

In order to avoid to the user to extract this information by hand, we have implemented a wrapper which automatically generates the DEVS Web Service Consumer for a Web Service. Thus, given a WSDL address, our framework is able to generate the corresponding DEVS Service Agent.

Web services are utilized using web service clients that are created by various open source and commercially available tools such as Eclipse Web Service Toolkit (WST), Netbeans IDE, Websphere etc.. All of them use the Web Service Description Language (WSDL) as the input to generate the web service client. In our

implementation we utilize the Axis2 framework to generate clients. Our choice of Axis2 plugin is driven by the implementation platform of DEVS/framework which is Axis/Java. However, it doesn't matter which method is used to generate the client.
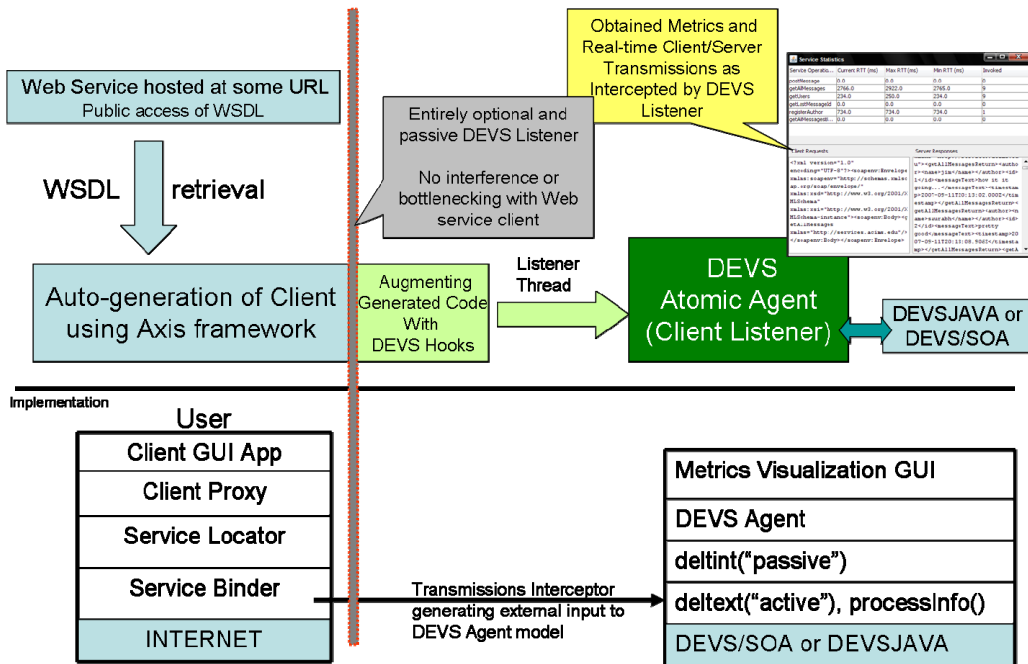
Figure 5: DEVS wrapper implementation over an Axis web service client

A DEVS model has two modes of operation: an internal behavior representation and an external behavior representation. In developing a DEVS wrapper, which would be effectively a DEVS web service client, we will implement the external behavior. The concept is shown in the top half of Figure 5. The detail is shown in the lower half of the same Figure 5. It shows the mapping between the Axis layers, specifically the Axis binding layer and the DEVS elements. It describes the external event that is triggered whenever there is message exchange through the Axis client. This triggered event informs the DEVS atomic model that wraps this Axis client. Such an arrangement does not create any bottleneck or any pipe between the actual Axis client and the network. The DEVS wrapper is informed of the round-trip-time (RTT) when the actual service has been executed its completion. Consequently, it is a passive observer and offers no interference to the true communication between the client and the live web service. By inserting a specific set of code in any Axis generated client, we can create a DEVS wrapper that is ready to become a part of a test-agent federation coupled system, as described in Section 6. Further, having such automated design, it allows augmentation of a comprehensive log mechanism that can provide many other instrumentation data than just RTT.

Having described the basic DEVS Web service wrapper, the next task in line is the creation of a coupled model, a web service workflow to be more specific to actually utilize the DEVS modeling and simulation capabilities. The coupled model where this DEVS WSDL model is a component of a bigger networked model is not the focus of this article and more details will be available in the extended article. It is not hard to understand that once you have an atomic model, it can be easily used as a component in a DEVS coupled model.

5. IMPLEMENTATION OF DEVS AGENT
This case study demonstrates the execution of a web service encapsulated in a DEVS wrapper Agent and the associated obtained statistics.

5.1. Web Service Work Flow Formalism
We compose a process workflow based on certain goals, objectives or requirements. We can deduce the information we need to compose a workflow and develop an automated procedure towards DEVS based design and analysis. The information set for a Web Service workflow can be described in a four element tuple as:

WSWF: **< W,M,F,X>**
where,

W: Set of Web service definitions (WSDLs) or Agents each with a valid URL
M: Set of web service methods
F: defined as <C,L,D>
    C is a set of W-M pairs with each pair as a source or destination
    L is a set of partner links with each link containing a src and dest pair defined in C
    D is a type of workflow mode which can either be a sequence, while, *holdSend* or *concurrent type*, which are corresponding to the BPEL specifications
X. Set of messages, where
    Each Message contains Data and is defined by time of entry in system, rate, whether it is periodic or stochastic and can be either an Input message or an Output message

5.2. DEVS Wrapper Agent

In this most basic demonstration, we use only one web service. This web service executes a chat session between two users. The schematic is shown in Figure 6. In our example, we execute the session with a live person and a DEVS agent. The live person here is 'Jim Client' that connects to the CHAT service via an Internet browser at (CHAT, 2007). The chat session is executed using the GUI as shown in Figure 7.
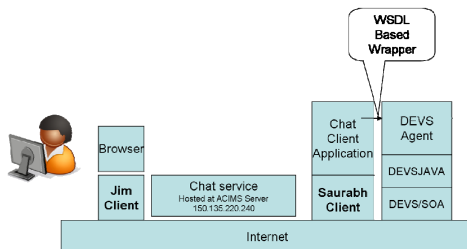

Figure 6: Schematic showing basic execution of DEVS Wrapper agent

The DEVS agent is defined according to the WSWF formalism as follows:

```
<W>: "CHAT":
    <W1:CHAT>:http://150.135.220.240:8080/C
hatServiceCollaboration/services/ChatServic
e?wsdl
    <A1:Jim>: "Jim:localhost:8080"
<M>: "Methods":
    <M1> postMessage()
    <M2> getAllMessages()
    <M3> getLastMessageId()
    <M4> registerAuthor()
    <M5> getUsers()
    <M6> getAllMessagesForAuthor()
<F>:"Flow specifications"
    <C>
<C1:Src>A1-M1
<C2:Src>A1-M2
<C3:Src>A1-M4
<C4:Src>A1-M5
<C5:Dest>W1-M1
<C6:Dest>W1-M2
<C7:Dest>W1-M4
<C8:Dest>W1-M5
    <L>
```
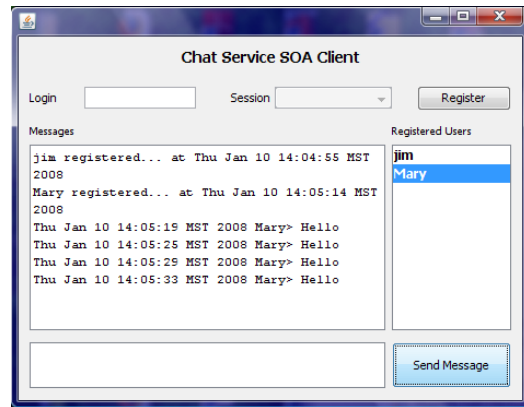
```
    <L1>C1-C5
    <L2>C2-C6
    <L3>C3-C7
    <L4>C4-C8
<D>
    <D1>M1-HoldSend
    <D2>M2-While-infinity
    <D3>M4-HoldSend
    <D4>M5-While-infinity
<X>: Set of Messages
    <InputMsg>
        <I1>W1-M1{string:T1:0:false:false}
        <I2>W1-M4{string:T0:0.1:true:false}
    <OutputMsg>
        <O1>M2{string:T2:1:true:false}
        <O2>M5{string:T2:1:true:false}
```


Figure 7: Chat Service Client connected to CHAT Service

<W> tag contains description of the Chat Web Service as W1 and the agent description as A1 along with their URL. **<M>** contains the list of service methods that may be used in the process flow. **<F>** contains the flow description categorized into **<C,L,D>** as per the WSWF. **<C>** provides the source and destination specification for a W/A defined in *<W>* with *<M>*. **<L>** specifies the coupling between the sources and destinations as defined in *<C>*. **<D>** contains the execution of methods in *<M>* in logic implementation. For example, `<D1>M1-HoldSend` implies that the method *M1* is to executed in *HoldSend* manner. Similarly, `<D2>M2-While-infinity` implies that *M2* will be executed indefinitely when invoked or bounded by any condition. **<X>** specifies the input and output message structures in **<InputMsg,OutputMsg>** tags. The structure of **<InputMsg>** as specified in WSWF SES is *<SystemComponent-Method{Data: time of Start: R+: isPeriodic: isRandom>*. For example, the specification `<I1>W1-M1{string:T1:0:false:false}` implies that the message *I1* is an input to *W1*, method *M1* with data as *string*. It starts at *T1* with period 0. Any non-zero value means that the message will be incoming at a periodic rate. The next boolean variable '*false*' implies that it is not periodic. The last variable '*false*' implies that it is not random either. Similary, `<I2>W1-M4{string:T0:0.1:true:false}` implies that *M4* at *W1* is to be invoked by *string* data message with a periodic rate of 0.1. The <OutputMsg> has a similar structure except the fact that it does not contain any

information about the system component. It only contains information about the method in <M> as it is just an output message. Whenever method <Mx> is invoked, it returns with the parametric details as in `<O1>M2{string:T2:1:true:false}`.

It is worth stressing here that the messages flow through the linkages as specified in **<L>**. This acts as a coupling for the DEVS models. There are two DEVS models in the WSWF instance described above, viz. *W1* and *A1*. Based on the coupling information for ex. `<L4>C4-C8` implies that the source is Agent `<C4:Src>A1-M5` and the destination is Web service `<C8:Dest>W1-M5`. The source sends a message invoking method *M5* at the destination. If there is a specification on how *M5* should be invoked in <*InputMsg*> listing, then the source has to ensure that it conforms to that specification. In this example there is no specification for M5. This implies that there are no parameters to be passed, but just the invocation. At the destination side, M5 has a specification `<O2>M5{string:T2:1:true:false}`, which implies that whenever *M5* returns a value, it will according to this <*OutputMsg*> specification.

The statistics for each of the methods in <*M*> is gathered according to the autogenerated agent GUI monitor at the agent's end. The statistics are largely the round trip time (RTT) for each of <*M*>. The GUI in Figure 8 also shows the SOAP messages that are exchanged between the pairs as specified in <*W*>.
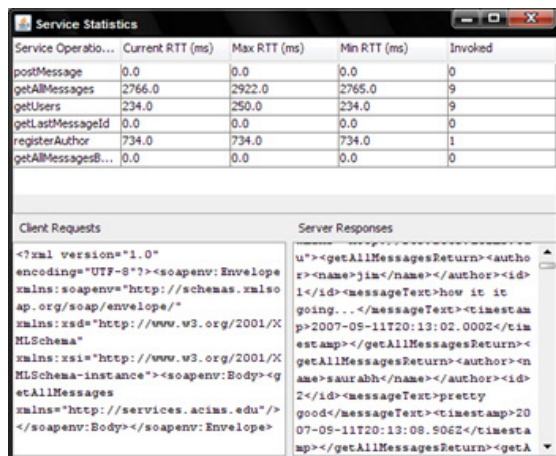

Figure 8: Associated Statistics GUI for an encapsulated Web Service in DEVS atomic model

6. MULTI-LAYERED AGENT BASED TEST INSTRUMENTATION SYSTEM USING GIG/SOA

A DEVS distributed federation is a DEVS coupled model whose components reside on different network nodes and whose coupling is implemented through middleware connectivity characteristic of the environment, e.g., SOAP for GIG/SOA, The federation models are executed by DEVS simulator nodes that provide the time and data exchange coordination as specified in the DEVS abstract simulator protocol. The DEVS Agent Monitoring System (TIS) is a DEVS

coupled system that observes and evaluates the operation of the DEVS coupled system model. The DEVS models used to observe other participants are the DEVS test-agents. The TIS should provide a minimally intrusive test capability to support rigorous, on-going, repeatable and consistent testing and evaluation (T&E). Requirements for such a test implementation system include ability to

1. deploy agents to interface with SoS component systems in specified assignments
2. enable agents to exchange information and coordinate their behaviors to achieve specified experimental frame data processing
3. respond in real-time to queries for test results while testing is still in progress
4. provide real-time alerts when conditions are detected that would invalidate results or otherwise indicate that intervention is required
5. centrally collect and process test results on demand, periodically, and/or at termination of testing.
6. support consistent transfer and reuse of test cases/configurations from past test events to future test events, enabling life-cycle tracking of SoS performance.
7. enable rapid development of new test cases and configurations to keep up with the reduced SoS development times expected to characterize the reusable web service-based development supported on the GIG/SOA.

Many of these requirements are not achievable with current manually-based data collection and testing. Instrumentation and automation are needed to meet these requirements.

Net-centric Service Oriented Architecture (SOA) provides a currently relevant technologically feasible realization of the concept. As discussed earlier, the DEVS/SOA infrastructure enables DEVS models, and test agents in particular, to be deployed to the network nodes of interest. Details on how such observers can be auto-generated and be executed using DEVS/SOA are provided in (Mittal, Zeigler, Martin, Sahin and Jamshidi, 2008; Zeigler and Hammonds, 2007).

6.1. Deploying Test Agents over the GIG/SOA

Figure 9 depicts a logical formulation test federation that can observe a SUT to verify the message flow among components as derived from information exchange requirements. In this context, a mission thread is a series of activities executed by operational nodes. In playing out this thread, DEVS test models are informed of the current activities (or see to detect their onset) as well as the operational nodes that execute these messages. These test models watch messages sent and received by the components that host the participating operational nodes. The test models check whether such messages are the ones that should be sent or received under the current function.

The test-agents are contained in DEVS Experimental Frames (EF) are implemented as DEVS models, and distributed EFs are implemented as DEVS models, or agents as we have called them, reside on network nodes. Such a federation, illustrated in Figure 10, consists of DEVS simulators executing on web servers on the nodes exchanging messages and obeying time relationships under the rules contained within their hosted DEVS models. This DEVS Agent Monitoring System that contains DEVS models interacts with real world web services through the DEVS agents that were described earlier.
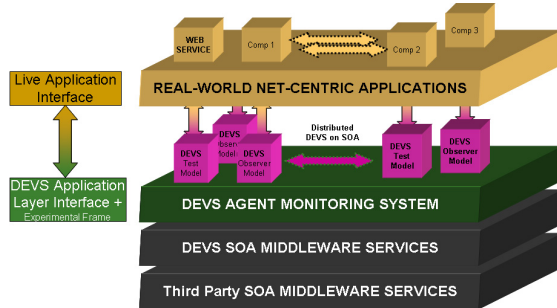
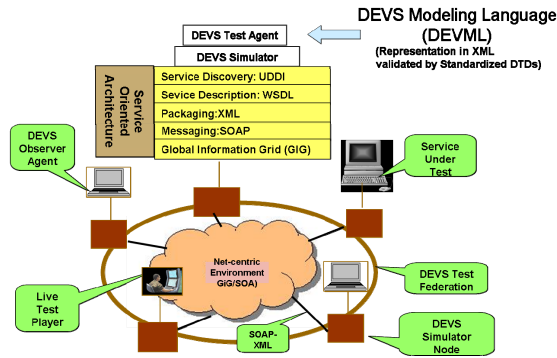Figure 9: Multi-layered Agent-based test instrumentation framework

Figure 10: Prototypical DEVS Test Federation

6.2. Implementation of Test Federations

A test federation observes an orchestration of web-services to verify the message flow among participants adheres to information exchange requirements. As derived from requirement, a process workflow is a series of activities executed by operational nodes and employing the information processing functions of web-services. As discussed in (Mittal, Zeigler, Martin, Sahin and Jamshidi, 2008; Zeigler and Hammonds, 2007), test agents watch messages sent and received by the services that host the participating operational nodes. Depending on the mode of testing, the test architecture may, or may not, have knowledge of the driving process workflow under test. If it is available, DEVS test agents can be aware of the current activity of the operational nodes it is observing. This enables it to focus more efficiently on a smaller set of messages that are likely to provide test opportunities.

To help automate set-up of the test we use a capability to inter-covert between DEVS and XML.

DEVSML (Mittal, Martin and Zeigler, 2007) allows distributing DEVS models in the form of XML documents to remote nodes where they can be coupled with local service components to compose a federation (Mittal, Martin and Zeigler 2007, Mittal, 2007) The layered middleware architecture capability is shown in 11 and (Mittal, Martin, and Zeigler 2007)
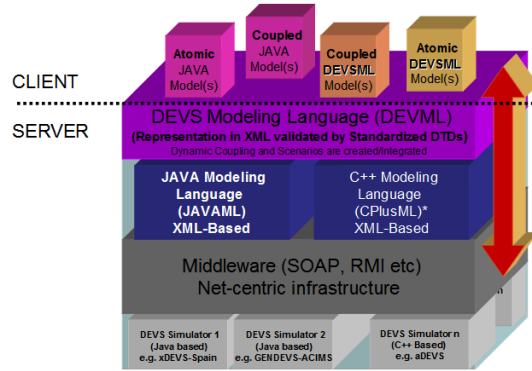

Figure 11: Layered Architecture of DEVSML towards transparent simulators in Net-centric domain

At the top is the application layer that contains model in DEVS/JAVA or DEVSML. The second layer is the DEVSML layer itself that provides seamless integration, composition and dynamic scenario construction resulting in portable models in DEVSML that are complete in every respect. These DEVSML models can be ported to any remote location using the web-service infrastructure and be executed at any remote location.

The simulation engine is totally transparent to model execution over the net-centric infrastructure. The DEVSML model description files in XML contains meta-data information about its compliance with various simulation 'builds' or versions to provide true interoperability between various simulator engine implementations. Such run-time interoperability provides great advantage when models from different repositories are used to compose models using DEVSML seamless integration capabilities. Recent articles provide an evidence in the direction to achieve interoperability for DEVS and non-DEVS models (Zeigler, Mittal and Hu, 2008; Mittal, Zeigler and Martin, 2008). Finally, the test federation is illustrated in Figure 10 where different models (federates) in DEVSML collaborate for a simulation exercise over GIG/SOA.

This section has laid out the framework on the creation and execution of a DEVS-based test instrumentation system.

7. CONCLUSIONS

Service Oriented Architecture (SOA) is still under development and many of the businesses are seriously considering migration of their IT systems towards SOAs. DoD's initiative towards migration of GIG/SOA and NCES requires reliability and robustness, not only

in the execution but in the design and analysis phase as well. Web service orchestration is not just a research issue but a more practical issue for which there is dire need. Further, Service Oriented Architecture must be taken as another instance of system engineering for which there must be laid out engineering process. Modeling and Simulation provides the needed edge. Lack of methodologies to support design and analysis of such orchestration except BPEL related efforts cost millions in failure. This research has proposed that Discrete Event Formalism can be used to compose and analyze Web service workflows. The DEVS theory, which is based on system theoretic concepts, gives solid grounding in the modeling and simulation domain.

We have shown how a web service can be encapsulated into a DEVS atomic model and can be put towards a coupled DEVS system with other live web services as well as other DEVS models. We also have demonstrated the proposed use of Web Service Work Flow (WSWF) formalism in composing SOA, much like of the same functionalities of BPEL. We have also described creation of DEVS net-centric coupled systems based on SOA. We have also shown how the developed DEVS coupled system can be simulated using the basic DEVSJAVA framework as well as distributed DEVS/SOA framework. Further, on the basis of our earlier work on DEVS/SOA we have basis for:

- Agent-Implemented Test Instrumentation
- Net-centric Execution using Simulation Services
- Distributed Multi-level Test Federations
- Analysis that can help optimally tune the instrumentation to provide confident scalability predictions.
- Mission Thread testing and data gathering:
- Definition and implementation of military-relevant mission threads to enable constructing and/or validating models of user activity.
- Comparison with current commercial testing tools shows that by replicating such models in large numbers it will be possible to produce more reliable load models than offered by conventional use of scripts.

We have taken the challenge of constructing net-centric systems as one of designing an infrastructure to integrate existing Web services as components, each with its own structure and behavior with DEVS components and agents. The net-centric system is analogous to a System of System (SoS) where in hierarchical coupled models could be created. Various workflows can be integrated together using component based design. The net-centric system can be specified in many available frameworks such as BPMN/BPEL, UML, or by using an integrative systems engineering-based framework such as DEVS.

In this research, we illustrated how M&S can be used strategically to provide early feasibility studies and aid the design process. We have established the capability to develop a live workflow example with

complete DEVS interface. In this role, DEVS acts as a full net-centric production environment. Being DEVS enabled, it is also executable as a system under test (SUT) model towards various verification and validation analysis that can be performed by coupling this SUT with other DEVS test models. Last but not the least, the developed DEVS system can be executed by both real and virtual users to the advantage of various performance and evaluation studies.

As components comprising SoS are designed and analyzed, their integration and communication is the most critical part that must be addressed by the employed SoS M&S framework. We discussed DoD's Global Information Grid (GIG) as providing an integration infrastructure for SoS in the context of constructing collaborations of web services using the Service Oriented Architecture (SOA). The present research is being considered and refined for testing GIG/SOA at Joint Interoperability Test Command [JITC], which is *the* agency to test future DoD systems. Clearly, the theory and methodology for such net-centric SoS development and testing are at their early stages.

REFERENCES

ACIMS software site, Available from http://www.acims.arizona.edu/SOFTWARE/softw are.shtml [Accessed July 2008]

IBM, BEA Systems, Microsoft, SAP AG, Siebel, 2007, Business Process Execution Language for Web Services version 1.1, Available from http://www.ibm.com/developerworks/library/speci fication/ws-bpel/ [Accessed July 2008]

Business Process Modeling Notation http://www.bpmn.org [Accessed July 2008]

Mittal, S., 2007, CHAT SOA web service, Available from http://www.saurabh-mittal.com/demos/ChatClient, [Accessed July 2008]

Mittal, S., 2007, DUNIP: A Prototype Demonstration, Available from http://www.acims.arizona.edu/dunip/dunip.avi [Accessed July 2008]

Joint Interoperability Test Command, a Defense Information Systems Agency at http://jitc.fhu.disa.mil/ [Accessed July 2008]

Department of Defense GIG Architectural Vision, Ver. 1.0, 2007, prepared by DoD CIO, available from http://www.defenselink.mil/cio-nii/docs/GIGArchVision.pdf [Accessed July 2008]

Hwang, M.H., Zeigler, B.P., 2006, A Modular Verification Framework using Finite and Deterministic DEVS, Proceedings of 2006 DEVS Symposium, pp. 57-65, April, Huntsville, Alabama, USA

Mittal, S., Martin, J.L., Zeigler, B.P., 2007, DEVS/SOA: A Cross Platform framework for Net-Centric Modeling and Simulation in DEVS Unified Process, SIMULATION: Transactions of SCS, under review

Mittal, S., Martin, J.L., Zeigler, B.P., 2007, DEVSML: Automating DEVS Simulation over SOA using Transparent Simulators", DEVS Symposium, March, Norfolk Virginia

Mittal, S., Martin, J.L., Zeigler, B.P., 2007, DEVS-Based Web Services for Net-centric T&E, Summer Computer Simulation Conference, San Diego, CA, USA

Mittal, S, 2007, DEVS Unified Process for Integrated Development and Testing of Service Oriented Architectures, Ph. D. Dissertation, Dept. of Electrical and Computer Engineering, University of Arizona

Mittal, S., Hwang, M.H., Zeigler, B.P, 2007, XFD-DEVS: An Implementation of W3C Schema for Finite Deterministic DEVS, Demo available at: http://www.saurabh-mittal.com/fddevs

Mittal, S., Zeigler, B.P., Martin, J.L.R., Sahin, F., Jamshidi, M., 2008, Modeling and Simulation for System of Systems Engineering, in Jamshidi, M., ed. Systems of Systems engineering for 21st Century, in press

Mittal, S., Martin, J.L.R., Zeigler, B.P., Nutaro, J.J., 2008, Design and Analysis of Service Oriented Architectures using DEVS/SOA-Based Modeling and Simulation, submitted to SIMULATION: Transactions of SCS

Mittal, S., Zeigler, B.P., Martin, J.L.R., 2008, Implementation of Formal Standard for Interoperability in M&S/System of Systems Integration with DEVS/SOA, Abstract accepted, manuscript in review at International Command and Control, C2 Journal

Net-Centric Enterprise Service http://www.disa.mil/nces/ [Accessed July 2008]

Pyke, Jon, 2007, XPDL: The Silent Workhorse of BPM, Available from http://www.bpm.com/FeatureRO.asp?FeatureId=2 32 [Accessed July 2008]

Mittal, S., 2007, DEVS/SOA sample demonstration in .avi format, Available from http://www.saurabh-mittal.com/demos/demoSOADEVS.avi [Accessed July 2008]

W3C Recommendation, 2007, Simple Object Access Protocol, Available from http://www.w3.org/TR/soap12-part1/ [Accessed July 2008]

W3C Recommendation, 2001, Web Services Description Language, Available from http://www.w3.org/TR/wsdl [Accessed July 2008]

Zeigler, B.P., Kim, T.G. and Praehofer, H., 2000, Theory of Modeling and Simulation. New York, NY, Academic Press.

Zeigler, B.P., Sarjoughian, H., DEVSJAVA, 2000: Available from http://www.acims.arizona.edu/SOFTWARE/devsj ava_licensed/CBMSManuscript.zip [Accessed July 2008]

Zeigler, B.P., and Hammonds, P., 2007, Modeling & Simulation-Based Data Engineering: Introducing Pragmatics into Ontologies for Net-Centric Information Exchange, New York, NY: Academic Press.

Zeigler, B.P., Mittal, S., Hu, X., 2008, Towards a Formal Standard for Interoperability in M&S/Systems of Systems Engineering, Critical Issues in C4I, AFCEA-George Mason University Symposium, May, Washington, D.C, USA

AUTHORS BIOGRAPHY

Saurabh Mittal is the CEO at DUNIP Technologies, India. Previously he worked as Research Assistant Professor at the Department of Electrical and Computer Engineering at the University of Arizona where he received his Ph. D in 2007. His areas of interest include Web-based M&S using SOA, executable architectures, Distributed Simulation, and System of Systems engineering using DoDAF. He can be reached at saurabh.mittal@duniptechnologies.com

José L. Risco-Martín is an Assistant Professor in Complutense University of Madrid, Spain. He received his PhD from Complutense University of Madrid in 2004. His research interests are computational theory of modeling and simulation, with emphasis on DEVS, Dynamic memory management of embedded systems, and net-centric computing. He can be reached at jlrisco@dacya.ucm.es

Bernard P. Zeigler is Professor of Electrical and Computer Engineering at the University of Arizona, Tucson and Director of the Arizona Center for Integrative Modeling and Simulation. He is developing DEVS-methodology approaches for testing mission thread end-to-end interoperability and combat effectiveness of Defense Department acquisitions and transitions to the Global Information Grid with its Service Oriented Architecture (GIG/SOA). He can be reached at zeigler@ece.arizona.edu

Jesús M. de la Cruz is Professor at the Department of Computer Architecture and Automation at the Complutense University of Madrid, Spain, where he is the head of the Automatic Control and Robotics Group. His interest covers broad aspects of automatic control and its pplications, real time control, simulation, optimization, statistical learning, and robotics. He can be reached at jmcruz@fis.ucm.es.

MODELLING AND SIMULATION USING STATECHART-BASED ACTORS

Franco Cicirelli[a], Angelo Furfaro[b], Libero Nigro[c]

[a][b][c]Laboratorio di Ingegneria del Software
(www.lis.deis.unical.it)
Dipartimento di Elettronica Informatica e Sistemistica
Università della Calabria
87036 Rende (CS) – Italy

[a]f.cicirelli@deis.unical.it, [b]a.furfaro@deis.unical.it, [c]l.nigro@unical.it

ABSTRACT
This paper describes an agent infrastructure centred on statechart-based actors for modelling and simulation of complex systems. Actors are lightweight reactive autonomous agents which communicate to one another by asynchronous message passing. Actor dynamism is specified through a "distilled statechart" which simplifies the expression of complex behaviour. The threadless character of actors conserves memory space and ensures efficient execution. The paper highlights current implementation status of statechart actors and demonstrates their practical use and programming style in Java through a manufacturing system model. Simulation experiments are reported. Finally, on-going and future work are summarized in the conclusions.

Keywords: Multi agent systems, statecharts, modelling and simulation, Java.

1. INTRODUCTION

Statecharts (Harel and Politi, 1998)(Booch *et al.*, 2000) are a well-known extension to classical state transition diagrams with constructs to control state explosion during development of complex systems. The basic mechanism consists in the possibility of nesting a sub automaton within a (macro) state thus encouraging step-wise refinement of complex behaviour. In addition a macro state can be *and*-decomposed for supporting a notion of concurrent sub automata. Statecharts have been successfully applied to the design of reactive event-driven real-time systems (Harel and Polity, 1998; Selic and Rumbaugh, 1998; Fortino *et al.*, 2001; Furfaro *et al.*, 2006), as well as to modelling and performance analysis, e.g. (Vijaykumar *et al.*, 2002-2006).

In the work described in this paper an original framework in Java is developed which enables modelling and simulation of agent-based systems (Wooldridge, 2002; Jang *et al.*, 2003) using statecharts. The proposal is an evolution of previous work of authors described in (Fortino *et al.*, 2001; Furfaro *et al.*, 2006) specifically to support M&S activities. The agent infrastructure is part of a more general distributed architecture named Theatre (Cicirelli *et al.*, 2007a). A particular notion of agents is adopted which rests on actors (Agha, 1986; Cicirelli *et al.*, 2007b). Actors are lightweight, threadless reactive components which communicate to one another by asynchronous message passing. An actor is characterized by its message interface, a set of hidden data variables and a behaviour for responding to messages (events). A multi-actor subsystem (theatre) is orchestrated by a control machine which provides basic timing, scheduling and dispatching message services to actors. In this paper the behaviour of an actor is specified through a "distilled" statechart, where only the *or*-decomposition of states is admitted. Concurrent sub states are avoided. All of this complies with the basic assumptions of the adopted actor computational model where concurrency exists at the actor level but not within actors. In other words, concurrency stems from reacting to messages and not from the use of heavyweight multi-threaded agents which are difficult to exploit, for space/time constraints, for M&S of large systems. A message reaction represents an atomic action which can modify the actor internal data, generate messages to known actors (*acquaintances*) including itself for pro-activity, create new actors, change the current state of the actor.

In this paper statechart-based actors are demonstrated by applying them to modelling and simulation of a manufacturing system. Results from simulation experiments are reported. Finally, conclusions are presented together with indications of future work.

2. A MODELLING EXAMPLE

In the following a system model is described concerning a manufacturing system. The example is adapted from (Vijaykumar *et al.*, 2002-2006) where it was handled by statecharts with *and*-decomposition and event broadcasting (Harel and Politi, 1998), and analytically studied by preliminarily transforming the model into a continuous time Markov chain (CTMC). In this paper the model is simulated. Obviously, simulation opens to the possibility of using probability distribution functions for event occurrences beyond the exponential one which is normally a prerequisite for building a CTMC. In addition, simulation can be exploited for more general exploration of system behaviour properties.

In the system (see Fig. 1), jobs are first processed by Machine A and then by Machine B. Job movements are mediated by a bounded capacity Inventory. A Robot is actually in charge of loading/unloading machines with jobs, possibly using the Inventory for temporary job buffering. Machines and the robot are subject to failure. Repairing from a failure is responsibility of an Operator.

Machines, Robot, Operator and the Inventory are modelled as actors. Event broadcasting is simulated in Fig. 1 by having a shared environment (passive) object Env which is updated by robot and machines and inspected by the robot and the operator to decide next action.
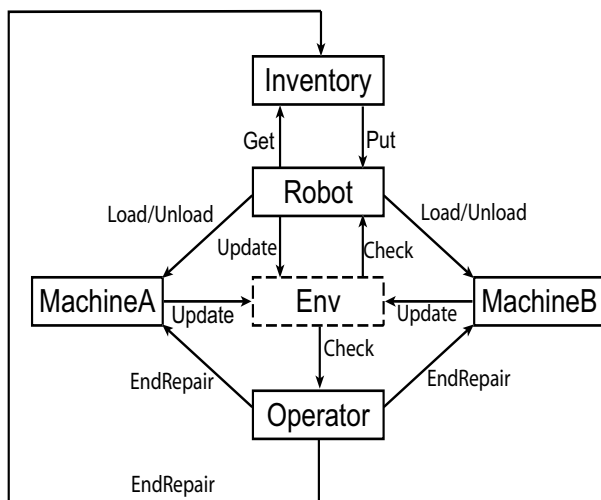
Figure 1: A manufacturing system model

Figg. from 2 to 5 depict respectively the behaviour of Machine, Operator, Inventory and Robot actors. The default state of each statechart is represented by a New leaf state from which the actors move into operation when receiving an Init message carrying the initialization parameters. Or-decomposition of states means that an actor occupying a macro state can find itself in one sub state only at each time. However, due to state hierarchy, for example, when the Robot is in the U2 sub state, it effectively is simultaneously in the U2, P, and Robot Top states (such a state chain is said a configuration). For Robot, Robot Top and P are examples of macro states. Remaining states are leaf states (they do not admit further decomposition). State transitions are represented by edges with arrows. Each transition is labelled by ev[guard]/action where ev is the trigger (event causing the transition), guard a logical condition which enables the transition when it evaluates to true, and action the action "à la Mealy" associated with the transition (not shown in the figures for the sake of simplicity). In Fig. 5 the transition with trigger Failure is an example of a group transition. It means that whatever is the internal sub state of P, the arrival of the Failure message causes the state P to be exited and state B to be entered, where the Robot requires to be repaired by the Operator. The transition with trigger EndRepair transmitted by the Operator, causes the

Robot to return into macro state P with history (see the shallow connector history H). This way the actor returns exactly into the internal sub state of P which was current when it was last left off at the time of Failure. Another form of history connector (not used in the example) is H* (deep history) which would case recursive state re-assumption within P and its current sub macro state down to a leaf state. In this case H and H* would be equivalent because all the internal sub states of P are leaf states.

The transition labelled by Tr[condP] and connecting state W of Robot to the boundary of state P asks for the default state of P (leaf state D) to be entered.

After being initialized, a machine (see Fig. 2) goes into the W state where it waits for a job to be processed. From W it moves into P state when loaded by the robot with a job. In the P state the loaded job is processed. At processing end, the machine moves into WU waiting for unloading. The message EndProcess is an internal message which the machine sends to itself for simulating the processing time (dwell time in P). During its stay in P the machine can have a failure in which case it passes to B state. Failure is another internal message which the machine sends to itself according to the next time to failure defined by a corresponding probability distribution function.
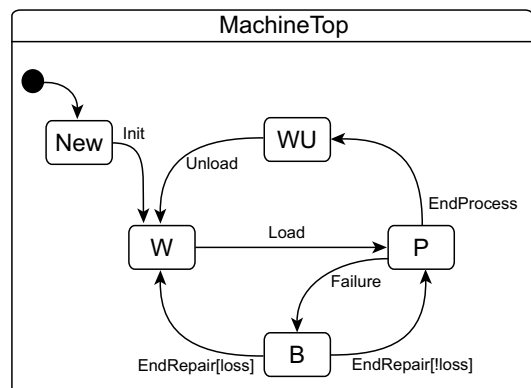

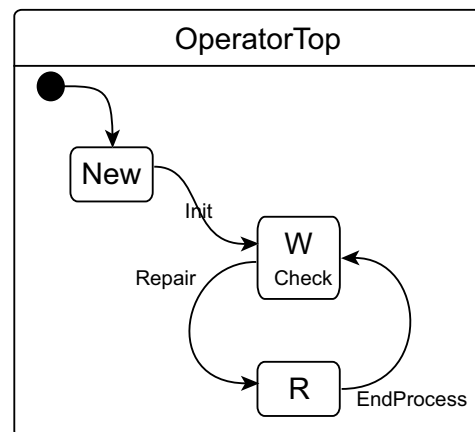Figure 2: Machine behaviour

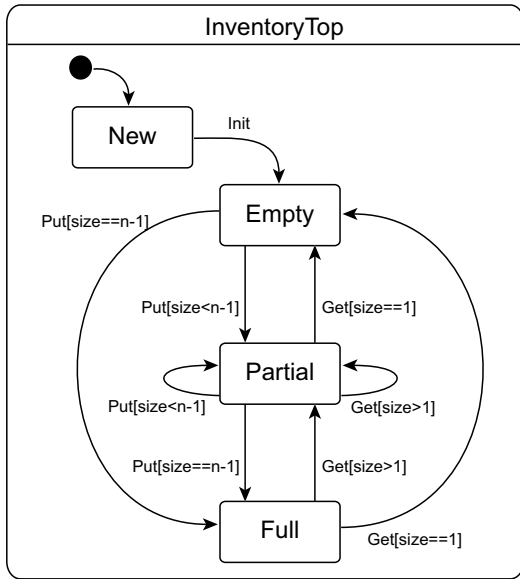Figure 3: Operator behaviour

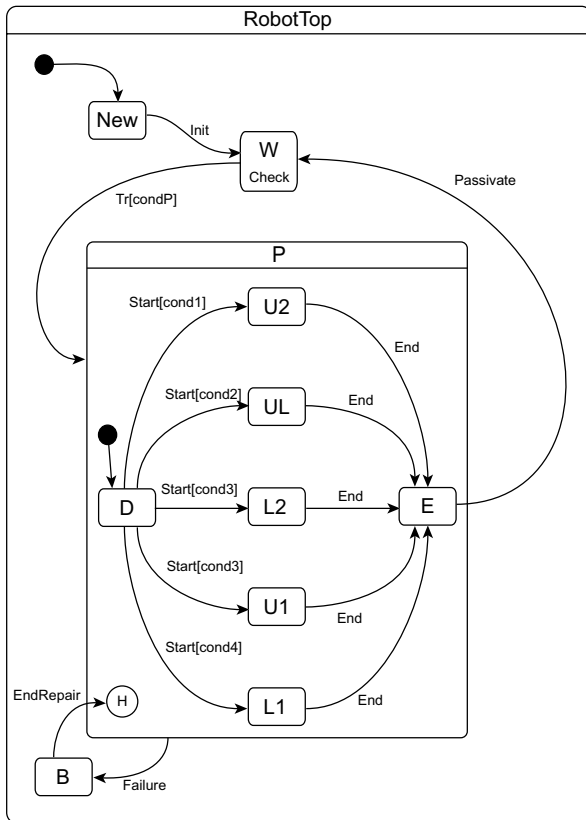Figure 4: Inventory behaviour

Figure 5: Robot behaviour

While staying in B the machine is repaired by the Operator. After repairing (arrival of the external originated message EndRepair from the Operator), the machine can return into P for continuing processing of interrupted job, or it can go back to W if the job is lost. The two possibilities are controlled by the loss guard which obeys to loss probability.

The Operator (Fig. 3) waits for a repairing request in the W state. The Check message is an internal event to W. It is received from the environment and alerts the operator to consult the environment for checking if a machine or the robot are in the B state requiring repair. If a repair request exists, the operator moves to the R state where the dwell time models the repairing time, which obviously is different from Machine A, Machine B and the Robot. The dwell time is tied to the internal message EndRepair whose action sends to a machine or robot a similar EndRepair message.

When faced with the decision of which entity to repair first, the operator chooses according to the following priority order: first machine Mb, then machine Ma, then the Robot.

Fig. 4 portrays the behaviour of the Inventory. It is a bounded buffer of capacity n, which can be 0 or a positive value. At each moment the initialized Inventory can be in one state among Empty, Partial or Full. Depending on current inventory size a Get/Put message can switch the inventory between Empty, Partial or Full as shown in Fig. 4.

When in the state W the robot (Fig. 5) waits for an operation to be exercised on the machines. Table 1 shows some logical conditions which influence the order in which the robot decides its next move.

Table 1: Environmental conditions affecting robot operation

cond1 = Mb is in WU;
cond2 = Ma is in WU && Mb is in W;
cond3 = Mb is W && ¬(Inv is empty);
cond4 = Ma is in WU && ¬(Inv is Full);
cond5 = Ma is in W;
condP = cond1 || cond2 || cond3 || cond4 || cond5;

As one case see from Fig. 5 the robot gives priority to unloading machine Mb (cond1 and sub state U2), then to simultaneously unloading machine Ma and loading machine Mb (cond2 and sub state UL), then to loading machine Mb (cond3 and sub state L2), then to unloading machine Ma (cond4 and sub state U1) and, finally, to loading machine Ma (cond5 and sub state L1). Robot effectively passes from state W to state P (and then to default state D) when condP which is the logical or of the various conditions is true. From D an internal immediate message Start is sent to itself for moving conditionally to a particular processing state among U2 to L1. The time spent by the robot in any operating state depends on the particular operation (see Table 2). Internal message End is self-sent for witnessing the end of operation, in which case the robot moves first into the E state and sends itself a Passivate message whose arrival takes the robot from state E to state W where the behaviour repeats again.

While being into an operating state, the Robot can fail (internal message Failure received). In this case the on-going operation is interrupted and the operator intervenes for repairing. Which event arrives first between End and Failure depends on the next time

303

respectively for completing the operation and for failing.

For the purpose of experimentation, all the timed events in the system are assumed to be exponentially distributed. Table 2 summarizes the rates (number of events per time unit) used for simulation (as in (Vijaykumar et al., 2002-2006)). Loading/unloading rate of machines are shown as dwell times in the corresponding operation state of the robot.

Table 2: Simulation parameter values

	Machine A	Machine B	Robot
Production rate	β=8,10 or 12	β=10	
Failure rate	λ=1	λ=0.5	λ=1
Repairing rate	μ=10	μ=15	μ=10
Loss probability	p=0.5	p=0.3	
Loading rate machine A			γ_{L1}=100
Unloading rate machine A			δ_{U1}=100
Loading rate machine B			γ_{L2}=100
Unloading rate machine B			δ_{U2}=100
Moving rate from m. A to m. B			α_{UL}=70

3. OUTLINE OF JAVA FRAMEWORK FOR STATECHART-BASED ACTORS

The following recapitulates the most important Java classes which were developed for supporting hierarchical actors.

Time, *AbsoluteTime*, *RelativeTime*. Are interfaces specifying a time notion. An absolute time is an instant in time when something can happen. A relative time is a duration, e.g. an amount of time measured from now. The interfaces have methods for adding/subtracting times as meaningful.

AbsoluteDiscreteTime, *AbsoluteDenseTime*, *RelativeDiscreteTime*, *RelativeDenseTime*. Are concrete classes implementing basic time interfaces. A discrete time is a long. A dense time is a double. The classes have a value() method which returns discrete/dense value of a time instance.

Clock. A basic interface for clocks. A clock has a method for checking current time, and methods for advancing the clock according to an absolute or relative time.

SimulationDiscreteTimeClock, *SimulationDense-TmeClock*. Are concrete classes implementing a clock for simulation, based respectively on discrete time or dense time.

Actor. Is the base abstract class for actors. An application actor derives directly or indirectly from Actor. Methods of Actor include send(Message) for sending a message to an acquaintance, handler(Message) which triggers an actor into operation for processing an arrived message (making a state transition), now() which returns an absolute time

indicating the current time. The ultimate meaning of now() depends on a control machine.

Message. Is the base class for messages. A message carries the receiver information.

Timer. Is an heir of Message. A timer is a triple: <Message timeout, Actor receiver, RelativeTime firetime>. At fire time the timeout message is consigned to its receiver actor. A created timer can be set and reset. Moreover its remaining time to firing and the elapsed time from its set time can be checked. The firetime is expected as a relative time which added to current time establishes the absolute fire time.

ControlMachine. Is the base class for simulation control engines. Its methods allow to schedule/unscheduled a message. A fundamental method is controller() which starts the control-loop of the engine.

Simulation. Is a concrete class deriving from ControlMachine. Its constructor receives the simulation time limit (an absolute time) and a clock to be used for managing the simulation time. The application passes a SimulationDenseTimeClock when it wants a dense time model to be used. Otherwise it has to pass a SimulationDiscreteTimeClock. Simulation uses a PriorityQueue for timers and a LinkedList for immediate concurrent messages, which are to be processed at current time.

State. Is the base abstract class for states. A state can be entered in a DEFAULT, THROUGH, HISTORY and DEEP_HISTORY mode. The THROUGH mode occurs when a transition reaches its destination state by crossing a state hierarchy. A state has a parent when it is nested into a macro state. To each state are associated the two basic methods entryAction() and exitAction() which have a default void implementation. They are executed respectively at each enter and exit from the state.

MacroState, *LeafState*. Are concrete classes extending State, respectively modelling a macro (o super) state which has inner nested states, and a LeafState which is a leaf in the state hierarchy tree.

Transition. Is a base abstract class modelling the concept of a transition in a statechart. A transition object carries its source state and the trigger message. Methods void action(Message) and boolean guard(Message) can be redefined in concrete transition objects for programming respectively the action and guard components of the transition.

InternalTransition. Is a concrete class extending Transition. It models internal events to a state, which do not cause entryAction()/exitAction() to be executed. Message Check in Figg. 3 and 5 is modelled by an internal transition.

Become. Is a concrete class extending Transition in the more general case of transferring from a source state to a destination state according to given enter mode. Following a become requires the exit-path (configuration) of source state and the enter-path (configuration) of destination state to be determined. Such paths extend from source or destination up to and

excluding their common ancestor state. Doing a become will imply exiting the states of the exit-path and entering the states of the enter-path.

DefaultPolicy. Is a class realizing the *Policy* interface. The class provides the default strategy for choosing among candidate transitions outgoing from a state (also considering hierarchy), corresponding to a given trigger. Selection always gives preference to a transition exiting from an inner state with respect to a transition exiting from an enclosing state and for a given state the choice is non-deterministic.

RandomGenerator. Is a concrete class providing common methods for random variate generators, e.g. uniform, exponential, normal etc.

4. PROGRAMMING STYLE

To figure out the resultant Java programming style, Listing 1 shows an excerpt from the Robot actor. The constructor of the Robot class is in charge of creating the state hierarchy of the corresponding statechart. Both states and transition objects must be defined. tR and tF are two timers for timing robot operation and next failure occurrence. eval() is an helper method checking environmental conditions as in Table 1.

Operating states U2, UL etc. redefine the entryAction() so as to set both tF and tR timers. This is important because the entry action gets executed every time the state is entered, e.g. by history.

As one can see the programming style is essentially declarative. In addition there is no need to redefine the handler() method coming from the Actor base class. All of this simplifies programming and makes it possible to automating translation from visual design to Java code.

5. SIMULATION EXPERIMENTS

The manufacturing model was simulated using the parameter values in Table 2, with the aim of validating the runtime infrastructure of the achieved implementation of statechart-based actors. Simulation uses dense time. Each experiment lasts after a time limit of tEnd=5×10^5. The environment object in Fig. 1 was extended in order to collect statistical information about system productivity, utilization of machines, robot and operator, losses in machines, average inventory size etc. The system model was studied in three cases: when machine A has respectively a lower/equal/greater production rate than B (see Table 2). System properties were analyzed vs. the inventory bounded capacity which was varied from 0 to 20 and then unbounded. Experimental results comply with those reported in (Vijaykumar *et al.*, 2002-2006), but furnish more detailed information about system behaviour.

Fig. 6 portrays measured system productivity (number of unloads from machine B per time unit) vs. the inventory capacity. As one can see, starting from 0, an increase in the inventory capacity increases the system productivity until the system reaches full-busy condition. In this condition the system exhibits maximum parallelism among components, with the

inventory which smooths out instantaneous differences in the production speed of the two machines. The smoothing effect is obviously greater when the production rate of machine A grows.

```
...
public Robot(){ //constructor
    super( new DefaultPolicy() );
    //make state hierarchy
    MacroState Top=new MacroState(null);
    State New=new LeafState(Top);
    State W=new LeafState(Top);
    State B=new LeafState(Top);
    MacroState P=new MacroState(Top);
    Top.setDefaultChild(New);
    State D=new LeafState(P);
    State E=new LeafState(P);
    State U2=new LeafState(P){
        public void entryAction(){
            tR.set( new End(), Robot.this,
                new RelativeDenseTime(rg.exponential(deltaU2)) );
            tF.set( new Failure(), Robot.this,
                new RelativeDenseTime(rg.exponential(lambdar)) );
        }
    };
    State UL=new LeafState(P){
        public void entryAction(){
            tR.set( new End(), Robot.this,
                new RelativeDenseTime(rg.exponential(alfaUL)) );
            tF.set( new Failure(), Robot.this,
                new RelativeDenseTime(rg.exponential(lambdar)) );
        }
    };
    …//similar definitions for L2, U1 and L1 sub states
    //make state transitions
    new Become( New, W, State.Mode.DEFAULT, Init.class ){
        public void action( Message m ){ … }
    };
    new InternalTransition( W, Check.class ){
        public void action( Message m ){ send( new Tr() ); }
    };
    new Become( W, P, State.Mode.DEFAULT, Tr.class ){
        public void action( Message m ){ send( new Start() ); }
        public boolean guard( Message m ){ eval(); return condP; }
    };
    new Become( B, P, State.Mode.HISTORY, EndRepair.class ){
        public void action( Message m ){ env.setRbInB(false); }
    };
    …//other transitions
    init( Top ); //initializes actor top state
}//end constructor
```

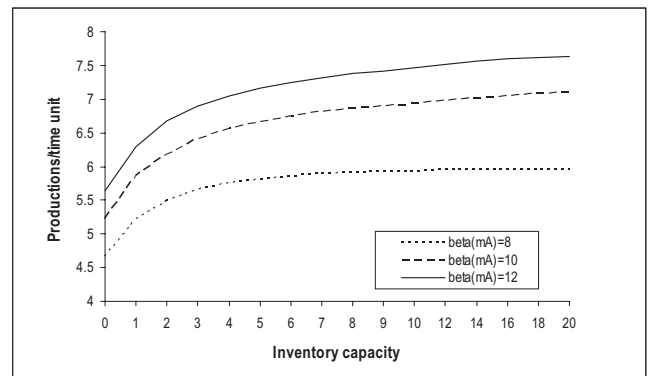Listing 1: A fragment from the Robot constructor

Figure 6: Observed system productivity vs. inventory capacity.

The positive effect of using a not zero inventory size can be checked in Fig. 7 which shows the waiting time for unloading machine A vs. the inventory capacity. This statistic was achieved by summing up the dwell time of machine A in state WU, waiting for the robot to unload the finished product, and then dividing the sum for the simulation time limit.
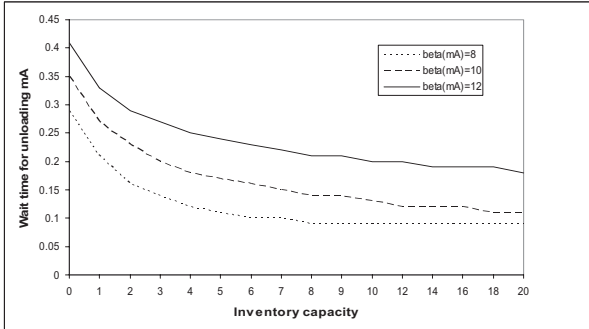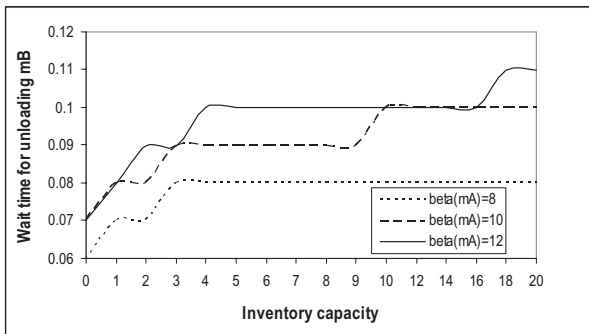

Figure 7: Average wait time for unloading machine A

Figure 8: Average wait time for unloading machine B

For completeness, Fig. 8 illustrates the wait time for unloading machine B, which has priority with respect to unloading machine A. As expected, machine B has a lower wait time. The two wait times become similar in full-busy operation of the system.

System behaviour can also be studied by watching the utilization factor (cumulative service time divided by the simulation time limit) of the various components (see Figg. 9 to 12).

Machine B utilization grows as the production rate of machine A augments and the inventory capacity is increased so as to buffer products delivered by machine A.
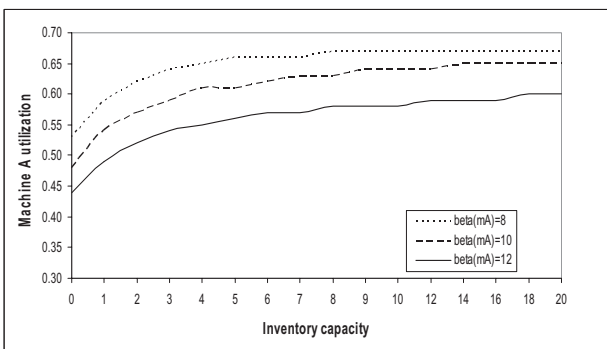

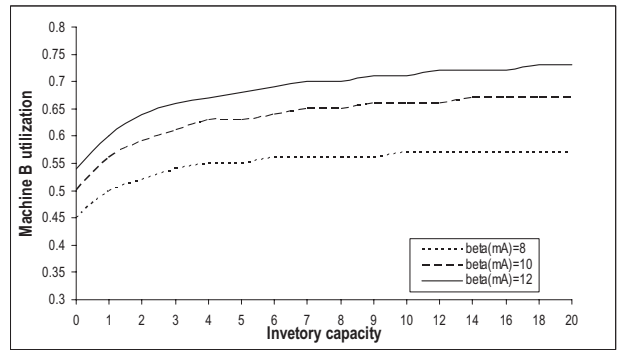Figure 9: Machine A utilization vs. inventory capacity

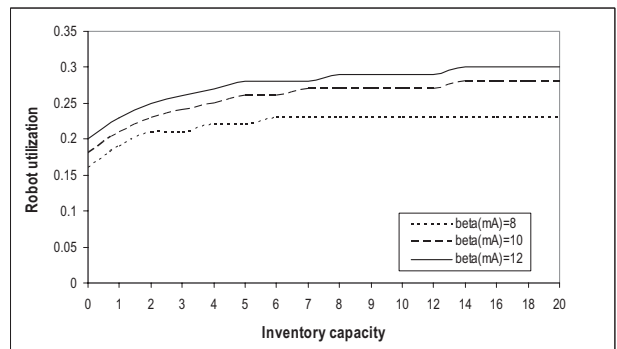Figure 10: Machine B utilization vs. inventory capacity

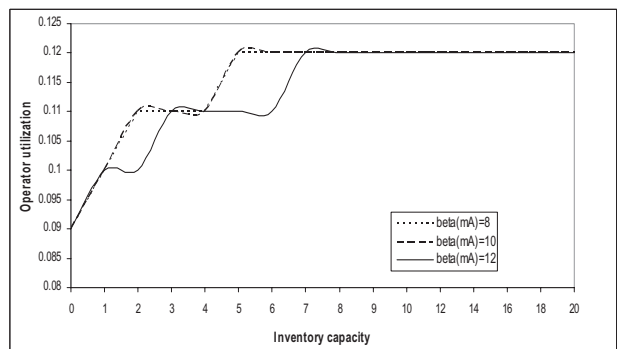Figure 11: Robot utilization vs. inventory capacity

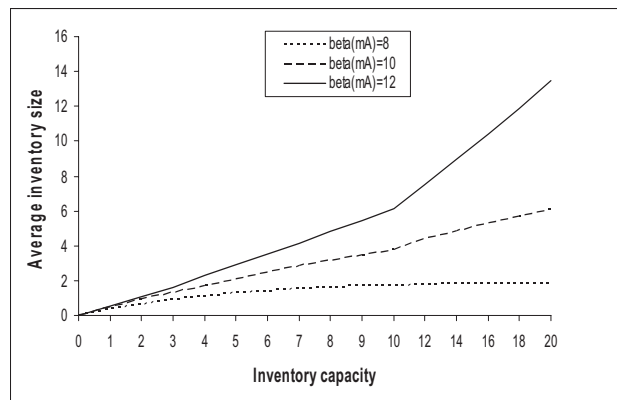Figure 12: Operator utilization vs. inventory capacity

Figure 13: Average inventory size vs. inventory capacity

At best operating conditions, the utilization of machine B is about 73%. From the perspective of machine A, robot availability and synchronization

concerns have the effect to diminish a little machine A utilization as its production rate augments from 8 to 12. However, system productivity and utilization of machine B are good in any case due to the inventory mediation. The incidence on the overall system behaviour of repairing failed components can be checked on Fig. 12 which portrays the operator utilization.

Fig. 13 depicts specifically the inventory usage, by showing the (temporal) average size of the inventory (collected through a path object) vs. the inventory capacity. When the production speed of machine A is 8, the average size of the inventory is definitely about 1.87 (even with unbounded capacity). With an unbounded inventory, though, it was found an instantaneous peek value of the inventory size of about 40.

By increasing machine A production speed, as expected, machine A tends to fill up the inventory as much as possible. In particular, for beta(mA)=10 (the two machines have identical speed), the average inventory size tends to be about 8 with a peek value, with an unbounded inventory, of about 90. In the case beta(mA)=12, the inventory is more intensely occupied. As witnessed by Fig. 13, the inventory average size continually increases, meaning that as long as there is a free slot in the inventory, machine A tends to fill it. The average inventory size, with capacity set to 20, was found to be about 13.5. Using unbounded capacity, it emerged that the average inventory size is about 139000 with a peek value of about 277000.

CONCLUSIONS

This paper argues that statechart-based actors (agents) have the potential to be really useful in the modelling and simulation of complex systems. A flexible and efficient (from both time and space) Java framework was developed which can directly be used through programming.

On-going and future work is geared at:

- porting the implementation in the Theatre architecture (Cicirelli *et al*., 2007b) for distributed simulation of very large systems, e.g. over HLA
- extending the realization toward design and implementation of embedded real-time systems
- experimenting with the use of statechart actors in the support of Parallel DEVS systems (Zeigler *et al*., 2000; Cicirelli *et al*., 2008)
- implementing a graphical tool for visual design of statechart actors and automatic generation of Java code.

REFERENCES

Agha, G., 1986. Actors: *A model for concurrent computation in distributed systems*. Cambridge:MIT Press.

Booch, G., Rumbaugh, J. and Jacobson, I., 2000. *The Unified Modeling Language User Guide*. Reading, MA:Addison-Wesley.

Cicirelli, F., Furfaro, A., Giordano, A., and Nigro, L., 2007a. An agent infrastructure for distributed simulations over HLA and a case study using unmanned aerial vehicles. *Proceedings of 40th Annual Simulation Symposium*, IEEE Computer Society Press, pp. 231-238. March 26-28, Norfolk (VA, USA).

Cicirelli, F., Furfaro, A., Nigro, L. and Pupo, F., 2007b. A component-based architecture for modelling and simulation of adaptive complex systems. *Proceedings of 21st European Conference on Modelling and Simulation (ECMS'07)*, pp. 156-163. June 4-6, Prague.

Cicirelli, F., Furfaro, A. and Nigro, L., 2008. Actor-based simulation of PDEVS Systems over HLA. *Proceedings of 41st Annual Simulation Symposium (ANSS'08)*, pp. 229-236. April 14-16, Ottawa, Canada.

Fortino, G., Nigro, L., Pupo, F. and Spezzano, D., 2001. Super actors for real-time. *Proceedings of Sixth Int. Workshop on Object-Oriented Real-Time Dependable Systems (WORDS01)*, IEEE Computer Society, pp. 142-149. Rome (Italy).

Furfaro, A., Nigro, L. and Pupo, F., 2006. Modular design of real-time systems using hierarchical communicating real-time state machines. *Real-Time Systems*, 32(1-2), 105-123.

Harel, D. and Politi, M., 1998. *Modeling reactive systems with statecharts*. Mc Graw-Hill.

Jang, M.-W., Reddy, S., Tosic, P., Chen, L. and Agha, G., 2003. An actor-based simulation for studying uav coordination. *Proceedings of the 15th European Simulation Symposium (ESS 2003)*, pp. 593–601. October, Delft (The Netherlands).

Selic, B. and Rumbaugh, J., 1998. Using UML for modeling complex real-time systems. Available on-line at: http://www.ibm.com/developerworks/rational/library/content/03July/1000/1155/1155_umlmodeling.pdf [accessed April, 2008]

Vijaykumar, N.L., de Carvalho, S.V. and Abdurahiman, V., 2002. On proposing statecharts to specify performance models. I*nt. Trans. in Operational Research*, 9, 321-336.

Vijaykumar, N.L., de Carvalho, S.V., Andrade, V.M.B. and Abdurahiman, V., 2006. Introducing probabilities in statecharts to specify reactive systems for performance analysis. *Computer and Operations Research*, 33(8), 2369-2386.

Wooldridge, M., 2002. *An introduction to multi-agent systems*. John Wiley & Sons, Ltd.

Zeigler, B. P., Praehofer, H. and Kim, T., 2000. *Theory of modeling and simulation*. New York: Academic Press. 2nd edition.

MULTI-AGENT BASED SIMULATION OF LARGE RANDOM BOOLEAN NETWORK

Dang Hai PHAM

Laboratoire de Cognition Humaine et Artificielle, Ecole Pratique des Hautes Etudes
41 Rue Gay Lussac, 75005 Paris, France

haipd-fit@mail.hut.edu.vn

ABSTRACT
Random Boolean networks, a generalization of cellular automata, were originally introduced as a simple model of genetic regulatory networks, but they are also used as mathematical models for studying complex dynamical systems with a large number of coupled variables. Simulating sequentially large networks with high connectivity meets often with difficulties on the time and the memory. We propose here a multi-agent based approach for simulating large random Boolean networks, which promises to give an improvement of its performance by using multiprocessors systems.

Keywords: Random Boolean network, Multi agent based simulation, Synchronization algorithm

1. INTRODUCTION
Random Boolean Networks (RBN), also known as Kauffman network or Nk-network, were originally introduced by Stuart Kauffman (Kaufman 1969) as a simple model to study the dynamics of complex genetic regulatory systems. In a random Boolean network, each gene is represented by a vertex in a directed graph. An edge from one vertex to another implies a causal link between two genes. The ON state of a vertex corresponds to the gene being expressed. Time is discrete, and at any time point, the new state of a vertex v is a Boolean function of the previous states of the vertices which are predecessors of v.

Up to now, the scope of existing analytical result is rather restricted, very few exact solutions have been determined. Instead computer simulations and statistical analysis are used to gain theoretical understanding. There are several tools available for exploration of different properties of RBNs, such as **DDLab** (*http://www.ddlab.com*), **RBN Toolbox** (*http://www.teuscher.ch/rbntoolbox),* **RBNLab** (*http://rbn.sourceforge.net*), but none of them can be used for the network with many vertex and high connectivity. That is because the sequential simulation of the large one meets often with difficulties on the time and the memory.

Large simulation consumes usually enormous amounts of time on sequential machines, while Multi-Agent Based Simulation can distribute its agents over different processors, thus reduces the computation time,

even though, it always requests synchronization for agents and depends robustly on the communication between agents. We believe that it will be an appropriate approach for simulating a large random Boolean network.

Having it in mind, we organize the rest of this paper as follows. The classical random Boolean network and the multi-agent based simulations are introduced in the section 2 and section 3 respectively. We propose in the section 4 a multi-agent based modeling random Boolean network. The section 5 gives some experiment results. This paper is closed with the conclusion and future work in section 6.

2. RANDOM BOOLEAN NETWORKS
A Classical RBN is a directed graph consisting of N nodes, where each node is connected to k other (or the same) nodes. The parameter k is known as the *connectivity* of the network. The set of k nodes connected to a given node is referred as its neighbourhood. Each node i is characterized by a Boolean (or binary) value $x_i \in \{0,1\}$ and a Boolean (or binary) function f_i, which assigns to each of the 2^k states of its k input nodes an output 1 with probability p or 0 with probability 1-p. The parameter $p \in [0, 1]$ is known as the *bias* of the logic function. The k connections and the Boolean function of each node are chosen randomly at the beginning but remain fixed during the dynamics of the network

In an RBN model, time is viewed as proceeding in discrete steps. At each step, the state of each node evolves according to its logic function and the previous states of its neighbourhoods. Let $v_i = (x^1_i, \ldots x^k_i)$ be set of k inputs of node x_i, then $x_i(t+1) = f_i(x^1_i(t), \ldots x^k_i(t))$ or $x_i(t+1) = f_i(v_i(t))$. Updating the state of all nodes in classical RBN is done synchronously. If $v(t) = \{v_1(t), \ldots v_N(t)\}$ is the state vector of the network at time t then evolution equation can be written $v(t+1) = f(v(t))$ with $f(v(t)) = \{f(v_1(t), \ldots f(v_N(t))\}$

Because the number of possible state of a network is finite, i.e. it equals to 2^N, so giving some initial state, evolution of the network in time traces a trajectory through the state space. If, after some period of time, the network comes back to one of the states it already visited L steps in the past, it will forever retrace this sequence of L states. We say that the system has fallen

into a *period L attractor*. Attractors of length L = 1 are called *point attractor* or *fixed points* of the network. The attractor states and the transient states leading to the attractor together constitute the basin of attraction. The number and length of attractors represent two important parameters of the cell modelled by a RBN. The number of attractors corresponds to the number of different cell types. The attractor's length corresponds to the cell cycle time which refers to the amount of time required for a cell to grow and divide into two daughter cells (Kauffman 1993). There are analytic solutions of RBN for k = 1 and k = N (Flyvberg and Kjaer 1988; Derrida and Flyvbjerg 1987), but finding general analytic solution is still an open issue. An example of a RBN with N=3 and k= 2 is shown in Figure 1.
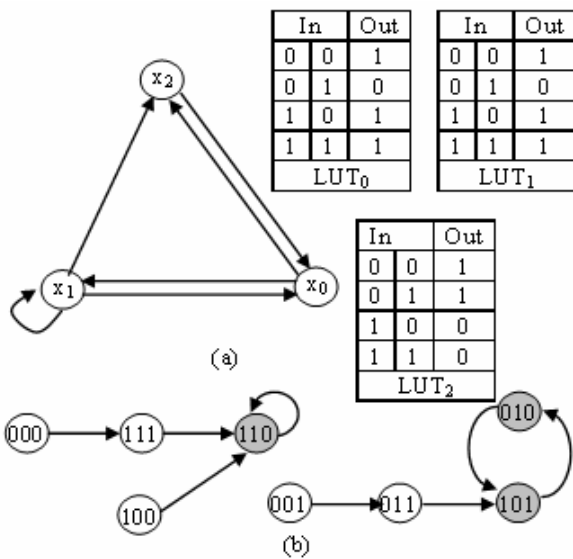

Figure 1: Example of classical RBN with N=3 and k=2 a) Directed graph and Boolean function using lookup table (LUT) for each node. b) State space diagram: there is one point attractor (110) and one cycle attractor of period two (101 010)

The parameters k and p determine the dynamics of the network. For a given bias p, there is a critical line $K_c = 1/2 \, p(1-p)$, below which the network is in the *frozen phase* and above which the network is in the *chaotic phase* (Derrida and Pomeau 1986). On the critical line between the frozen and the chaotic phases, the network exhibits *self-organized critical behaviour*, ensuring both stability and evolutionary improvements.

In the classical random Boolean networks, one important aspect is that the networks are assumed to be synchronous; however it is not very realistic as genes do not change their state all at the same moment, hence it is more appropriate to update these systems asynchronously using a random order. The exploration of such asynchronous RBNs can be found in (Harvey and Bossomaier 1997; Di Paolo 2001; Gershenson 2004).

3. MULTI-AGENT BASED SIMULATION

Agent-based simulation models are normally based on a discrete time model, where system states are changed at discrete time points only. The two most common types of such discrete time simulation are called *time-stepped* and *event-driven* (or *event-stepped*) models (Fujimoto 2000). They are distinguished by the mechanism used to advance simulation time. In the first approach, the simulation advances in equal-size time steps. In each step, the clock is adjusted and the participants are informed about the new time. They then can check if an activity has to be executed at this point in time. In the other one, a timestamp is assigned to each event which is an abstraction used in simulation to model some instantaneous action in the real system, to indicate the point in simulation time, when the event occurs. Each event usually results in some change in state variables defined by the model.

In Multi-agent based simulation (MABS), system entities are modelled using multiple agents which interact with the other agents by sending messages to them and to their environment. Agents can be viewed as threads of control, which execute concurrently (Wooldridge 2002). Therefore, it is important that the agents perceive a common view of the virtual environment. For example, any pair of agents should perceive a set of messages in the same sequence, or in other words, each agent has to process events/messages according to their timestamp order. This is called *local causality constraint* (Fujimoto 1990). Two main approaches have been proposed to ensure that the local causality constraint is not violated: *conservative* and *optimistic* synchronization.

The first one avoids strictly the causality violations by processing safe events only. The well-know algorithm **CMB** (Bryant 1977; Chandy and Misra 1979) ensures the local causality constraint is never violated but may lead to deadlock situation. The *null message* scheme (Misra 1986), on its turn, can avoid this deadlock but may generate an excessive number of null message. In addition, this technique relies heavily on the concept of *look-ahead* (Nicol 1996) and also typically requires a static connection between agents.

The optimistic approach, in contrast, allows causality error to occur but provides suitable techniques to recover from an incorrect system state. The best know is **Time Warp** algorithm (Jefferson 1985). It uses a mechanism called *rollback* to cancel the erroneous computation and reprocess messages in their timestamp order whenever a *straggler message* occurs. Several additional strategies of memory management and cancelling incorrect computation to improve efficiency of this algorithm can be found in (Fujimoto 1990; Jefferson 1990; Lin and Preiss 1991). The Time Warp exploits full parallelism of systems, nevertheless it is hard to implement because of requiring complex manipulations.

Another problem in multi-agent systems is the accessing to shared state variables which are maintained by the different agents. This usually leads to an all-to-all

communication between agents and results in the degradation of simulation performance. Many investigations in the context of Multi-Agent Based Simulation have addressed to the avoiding this broadcast communication. Most of them limit the number of agents receiving message based on their location or other application-specific attributes. Algorithms such as can be found in (Macedonia et al. 1995; Logan and Theodoropoulos 2001; Ewald et al. 2006).

4. MODELING RANDOM BOOLEAN NETWORK

We model the Random Boolean Network as a multi agent system. Each node of the network is modeled as an agent that consist of three main fields: agent's state which is a Boolean variable, list of its neighborhoods, which is a dynamic array (because the networks can be non-homogeneous, k is not the same for all agents, such as a scale-free network has an exponential distribution of connectivity, so although most nodes may have a small connectivity, some can have large values) and a lookup table which can be represented as an (array of) integer. The agents behave step by step. At each step in its evolution, an agent reads the state of all its neighborhoods refers to its lookup table and then updates its state.

In the distributed simulation, we can simply model each agent as a process which behaves independently. These agents are distributed over different processors and interacts each other by exchanging messages. At each step, after updating its state, each process send (k) messages to (k) its neighborhoods. Clearly, it is an effectless approach because there are too many of message sent in each step (kN messages). A more effective approach decomposes random Boolean network model by regrouping lightweight agents to a "*big-agent*" and assigns a logical process (LP) to simulate each one. The *big-agent* contains a Boolean (or bit) array variable (referred as *subNetState*) to present its internal state which is constructed from all state of its lightweight agents (Figure 2). At each step, instead of sending directly state to neighborhoods, each lightweight agent updates appropriate element in this *subNetState* variable. Hence, to read state of neighborhoods, agents only need refer to the *subNetState* variables. Obviously it would be possible for neighborhoods of an agent p simulated by LP_i to be an agent q residing in another LP_j. That means agent p has to refer to q^{th} element of the variable *subNetState* of the LP_j. Thus each LP, for computing its internal state at step t + 1, must know the ones which were computed by all others LPs at step t. In other words, LPs must be synchronized. Two approaches are suggested.

The first one is based on the distributed memory model, in which each LPs after finishing its computation at step t will send its internal state to all others LPs. It continues to compute next step t + 1 only when it received *subNetState* variables from all others LPs. Suppose that the RBN model is simulated by Q

different LPs, then the number of message must be sent per step is Q * (Q - 1). That reduces its performance. Detailed analyses about the performance of this approach are beyond the scope of this paper, but are discussed elsewhere.
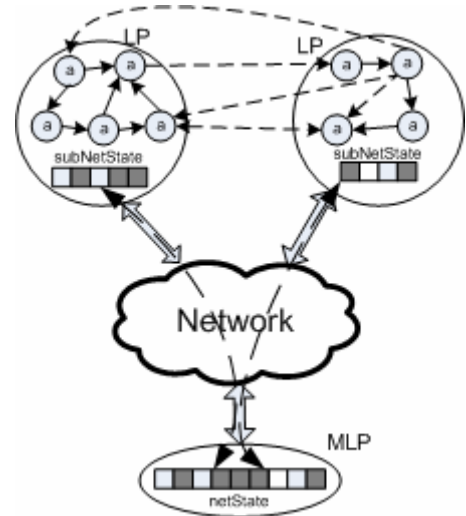

Figure 2: Decomposing a Random Boolean Network model.

The other approach is based on the share memory model, in which each LP when have finished its computation at step t, writes its *subNetState* to a *pseudo-share* memory. This *pseudo-share* memory object can be modelled by a *memory logic process* (MLP). The MLP receives all *subNetState* variables written from LPs, constructs a *netState* array that present the state of RBN and then broadcasts it again to all LPs (figure 2). In the case of Q logical processes, the number of message sent per step is 2 * Q.

By broadcasting *netState*, the MLP forces others LPs to advance together in lock step. This is a synchronous simulation and it guarantees that the causality errors does not happen but the potential for speedup is limited, especially when the RBN model is partitioned into too many sub-models, it can create a bottleneck at the memory logical process and this reduces its performance.

Let **t** be the average time value which one lightweight-agent needs for simulating a step, δ be the average time value for communicating the state of one agent (*it is usually a bit*) between MLP and LPs then the average simulation time T, which be computed at MLP in one step can be conjectured:

$$T_{MLP} = N\delta + \frac{N}{Q}t + \frac{N}{Q}\delta + \Delta \qquad (1)$$

Where, N is number of nodes of the modelling RBN, Q is number of logical process (so the number of lightweight agents in each LP is N/Q) and Δ is some delay coefficient of the communication, which could occur by numerous factors, for example by the waiting to process when many LPs send their internal state to

the MLP at the same time. If the ratio of **t** to δ is **a** then the speedup in the ideal case, $\Delta = 0$, will be:

$$Speedup \approx \frac{Nt}{N\delta + \frac{N}{Q}t + \frac{N}{Q}\delta} = \frac{aQ}{Q + a + 1} \qquad (2)$$

Clearly, the speedup depends robustly on the ratio of computation time to communication time. This ratio will be theoretic limit of speedup in our synchronous approach. This also suggests that the simulation on a multiprocessor system does not suffer from communication latency, always give a better result.

5. EXPERIMENTAL RESULTS

Currently, many languages and tool that are used for the distributed and parallel simulation are already available. Most of them use mainly the integrating of agent toolkits into a High Level Architecture (HLA) based distributed simulation and focus mainly on the interoperability between different sequential simulation toolkits not on the gaining speedups (Minson and Theodoropoulos 2004; Wang, Turner, and Wang 2005; Lees, Logan, and Theodoropoulos 2007). Some others approaches tend to obtain performance improvements but in those platforms, the communication between logical processes uses usually TCP/IP protocol which is reliable, FIFO and point- to-point (Pawlaszczyk and Timm 2006; Cicirelli, Furfaro, and Nigro 2007). Thus the broadcast from the MLP to all other LPs is essentially sequential: the MLP send continuously *netStatus* to LPs one by one. We can conjecture the speedup in this case as equation below:

$$Speedup \approx \frac{Nt}{Q*N\delta + \frac{N}{Q}t + \frac{N}{Q}\delta} = \frac{aQ}{Q^2 + a + 1} \qquad (3)$$

Obviously, this *sequential-broadcast* degrades the simulation performance when model is partitioned into many logical processes.

Our experiments were performed in a network of 512MB, Intel 3.0GHZ computers. The LAN has links with 100Mb/s in bandwidth. Logical processes have been implemented with Java language, each one is executed on a separate processor. To "*write*" *subNetState* from LPs to MLP, standard RMI library was used, but we use multicast mechanism to broadcast *netState* from MLP to all others LPs. To ensure that the communication is reliable we use a timeout mechanism and a cyclic redundancy check (CRC) mechanism (Peterson and Brown 1961). The duplicate appearing will be eliminated by checking logical time (Lamport 1978) of the message received (*netState*) with local time at the receiving LP. It is clear that by using of mechanisms for checking every packet arrived, it can reduce the simulation performance but we believe that, in a Local Area Network, this reduction of performance is acceptable. Behaviours of the logical process and

memory logical process are given in algorithm 1 and algorithm 2 respectively.

Algorithm 1: Logical Process
While true Do
    Wait *netState* from MLP
    If CRC value is correct **Then**
      If LogicalTime = LocalTime **Then**
        For each agent i in lightweight-agent list **Do**
          subNetState[i]←agent[i].step(*netState*)
        End for
        LocalTime++
        Send *subNetState* to MLP
      End if
    End if
End while

Algorithm 2 Memory Logical Process
While true Do
    Broadcast *netState* with its CRC value to all LPs
    Clock.Start()
    Repeat
      Wait subNetState from any LPi
      Update i[th] partition of the *newNetState* variable
    Until received *subNetState* variable from all LPs
    Clock.Stop()
    netState ← newNetState
    LogicalTime++
End while

Figure 3 shows the fields of a packet which is broadcasted from MLP to all others LPs. The first two fields are sequence number and CRC value of the data field. The next field identifies the logical time which denotes how far in simulated time the network has simulated and the last field is broadcasted data which is network state (*netData*). Due to limitation on size of sent diagram in Java UDP, the packet size must be less than 64K, including its header, and so if network is too large, MLP will broadcast continuously *netState* variable in some different packets. In those cases, the SegNbr field identifies order of the sent packets. But in current implementation, we didn't use this field, so the largest simulated RBN contain maximum of approximate 2^{19} nodes.

SegNbr	CRC	Time	Data (netState)

Figure 3: Fields of a broadcasted packet

To estimate theoretic speedup of the simulation of large random Boolean network, we compute the ratio of average time an agent need to simulate a step to average time for communicating a single bit data between LP and MLP. We implemented a sequential simulator to simulate RBNs. 500 classical RBNs of 256K-nodes with k is 1, 4, 7 were simulated for 100 steps. The accumulated average simulation time is computed. The results of this experiment were given in figure 4.
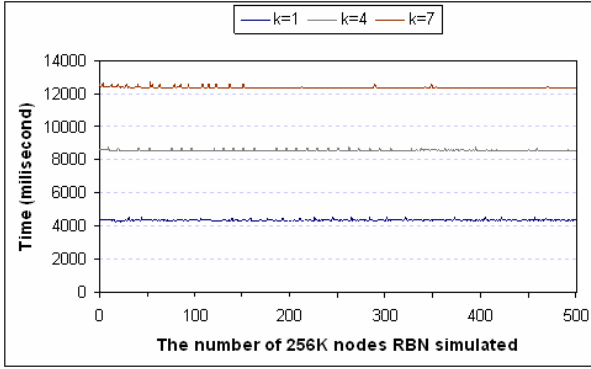
Figure 4: The average time value to simulate a 256K-nodes RBN for 100 steps

The communication time contain the time for writing data to output port at the sending process, the time for transferring data in the link and the time for reading all data from input port at the receiving process. We use two processes executing on the two separates processors: One process sends a volume of data and another process receives this data and then resends immediately it to the sender. The communication time is computed at the first process when it begins sending a volume of data until it receives whole those data. The sending process iterates this procedure in 500 times and the accumulated average time value for communicating a bit data between processes is computed. Figure 5 shows the result of this experiment in the case of volume size being 50Kb. In fact, this average time depends on the buffer size. The smaller buffer size is, the larger average time is.
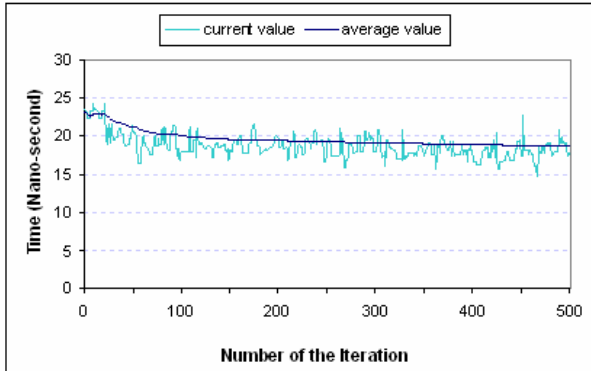

Figure 5: The average time value for broadcasting a bit of data

By using results of two experiments above, we can compute the ideal theoretic speedup according to equation 2. Figure 6 illustrates that result.

To validate these theoretic results, we simulated 500 classical random Boolean networks, each of which is composed 256K nodes. For each network, we simulate it in 500 steps. The number of logical process is equal to the number of processor. Figure 7 shows the speedup of this experiment with different values of parameter k.
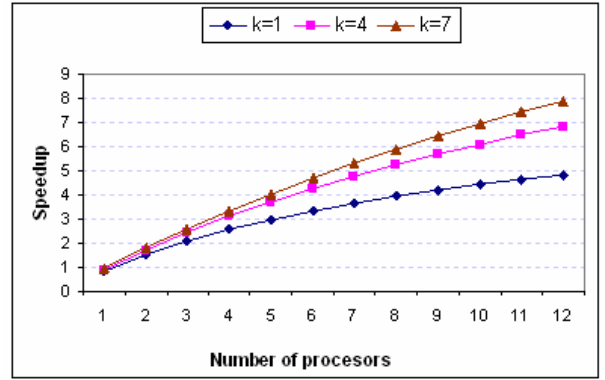

Figure 6: The theoretic ideal speedup for the synchronous simulation of large random Boolean networks

We observe that the real speedup is smaller than the theoretic one, especially when there are more processors. This is mainly due to the considerable bottleneck at the MLP. Nevertheless, in a real share memory multiprocessors system, because each LP writes really to a disjoint memory area, we believe that this bottleneck will not happen.
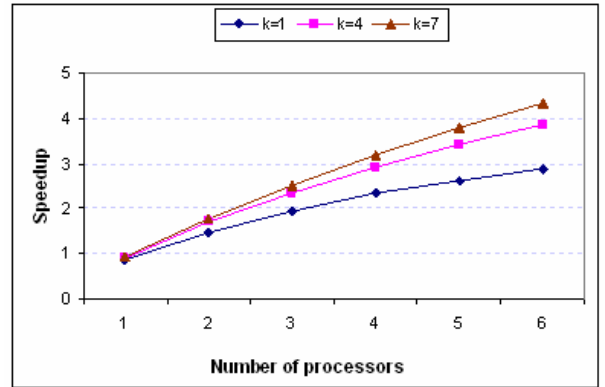

Figure 7: The real speedup for the synchronous simulation of 256K nodes RBN with different connectivities.

6. CONCLUSION

We have introduced classical random Boolean network. A simple multi-agent based approach for implementing large-scale simulation of random Boolean network has been suggested. The theoretic analysis of dependence between the performance and the ratio of computation to communication are presented. Future efforts might be directed toward simulating large random Boolean networks and other large multi agent systems on multi processors systems.

Pratique des Hautes Etudes for providing a proper environment for the initial preparation of this paper.

REFERENCES

Bryant, 1977. *Simulation of packet communications architecture computer systems*. Massachussets Institute of Technology.

Chandy, K.M. and Misra, J., 1979. Distributed simulation: A case study in design and verification of distributed programs. *IEEE Transactions on Software Engineering* SE-5 (5): 440-452.

Cicirelli, F., Furfaro, A. and Nigro, L., 2007. Exploiting agents for modelling and simulation of coverage control protocols in large sensor networks. The *Journal of Systems and* Software 80 (11): 1817-1832.

Derrida, B. and Flyvbjerg, H., 1987. The random map model: A disordered model with deterministic dynamics. *Journal de Physique*, 48: 971-978.

Derrida, B. and Pomeau., Y. 1986. Random networks of automata: a simple annealed approximation. *Europhys. Lett. 2*: 45-59.

Di Paolo, E.A., 2001. Rhythmic and non-rhythmic attractors in asynchronous random Boolean networks. *Biosystems* 59 (3): 85-95.

Ewald, R., Chen, D., Theodoropoulos, G., Lees, M., Logan, B., Oguara, T., and Uhrmacher, A. M., 2006. Performance Analysis of Shared Data Access Algorithms for Distributed Simulation of Multi-Agent Systems. *Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation (PADS'06)*, 29-36.

Flyvberg, H. and N. Kjaer. 1988. Exact solution of the Kauffman model with connectivity one. Journal de Physique A 21 (7): 1695-1718.

Fujimoto, R.M, 1990. Parallel discrete event simulation. *Communication of the ACM* 33 (10): 30-53.

Fujimoto, R.M., 2000. *Parallel and Distributed Simulation Systems*. John Wiley & Sons.

Gershenson, C., 2004. Introduction to Random Boolean Networks. *Workshop and Tutorial Proceedings Ninth International Conference on Simulation and Synthesis of Living Systems (ALife IX),* 160-173.

Harvey, I. and Bossomaier, T., 1997. Time out of joint: Attractors in asynchronous random Boolean networks." *Proceedingsof the Fourth European Conference on Artificial Life (ECAL97)*, 67-75.

Jefferson, D. R. 1985. Virtual time. *ACM Transactions on Programming Languages and Systems*, 7(3): 404-205.

Jefferson, D. R. 1990, August. Virtual time II: Storage management in distributed simulation. *Proceeding of the 9th Annual ACM Symposium on Principles of Distributed,* 75-89.

Kauffman, S., 1969. Metabolic stability and epigenesist in randomly constructed genetic net. *Journal of Theoretical Biology*, 22 (3): 437 - 467.

Kauffman, S., 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press.

Lamport, L., 1978. Time, Clocks, and the Ordering of Events in a Distributed System. *Comm. ACM*, 21 (7): 558- 565.

Lees, M., Logan, B. and Theodoropoulos, G.K., 2007. Distributed simulation of agent-based systems with HLA. *ACM Transactions on Modelling and Computer Simulation (TOMACS)*, 17 (3): 11.

Lin, Y. and Preiss, B.R., 1991. Optimal Memory Management for time wrap parallel simulation. *ACM Transactions on Modelling and Computer Simulation* 1 (4): 283-307.

Logan, B. and Theodoropoulos, G., 2001. The distributed simulation of Multi Agent Systems. *Proceedings of the IEEE*, Volume 89, 174-186.

Macedonia, M. R., Zyda, M.J., Pratt, D.R., Brutzman, D.P. and Barham, P.T., 1995. Exploiting Reality with Multicast Groups. *IEEE Computer Graphics and Applications* 15(5), 38-45.

Minson, R. and Theodoropoulos, G., 2004. Distributing RePast Agent-Based Simulations with HLA. *Proceedings of the 2004 European Simulation Interoperability Workshop*. Paper No. 04E-SIW-046. Edinburgh

Misra, J., 1986. Distributed discrete – event simulation. *ACM Computing Survey*, 18 (1): 39-65.

Nicol, D. M., 1996. Principles of conservative parallel simulation. *Proceedings of the 1996 Winter Simulation Conference*. 128-135.

Pawlaszczyk, D. and Timm, I.J., 2006. A Hybrid Time Management Approach to Agent Based Simulation. *Proceedings of the 29th Annual German Conference on Artificial Intelligence (KI 2006)*, 374-388.

Peterson, W. and Brown, D.T. 1961. Cyclic Codes for Error Detection. *Proceedings of the IRE*, 49(1): 228-235.

Wang, F., Turner, S.J. and Wang, L., 2005. Agent Communication in Distributed Simulations. *Lecture Notes in Computer Science* 3415:11-24.

Wooldridge, M., 2002. *An Introduction to Multi Agent Systems*. John Wiley & Sons.

AUTHORS BIOGRAPHY

PHAM Dang Hai received the engineering diploma in Information Technology from the Hanoi University of Technology, Viet Nam, in 1995 and the diploma of master in Information Technology from the Institut de la Francophonie pour l'Informatique, Viet Nam, in 1997. From 1997 to 2004, he was a lecturer at the Department of computer science, Faculty of Information Technology, Hanoi University of Technology, Viet Nam. He is presently a Phd student at the Ecole Pratique de Hautes Etudes, Paris, France. His current research interest includes Parallel and Distributed simulation, Multi-agent based simulation.

SIMULATION OF THE IMPACT OF THE ENERGETIC CHARACTERS OF TRACTORS AND MACHINES ON THE WORKING EFFICIENCY OF THE SOIL TILLAGE UNITS

Arvids Vilde, Edmunds Pirs

Latvia University of Agriculture, Research Institute of Agricultural Machinery
Latvia University of Agriculture, Research Institute of Agricultural Machinery

Email: vilde@delfi.lv ; edmunds@armuss.lv

ABSTRACT

For soil tillage only those tractor-machine units (aggregates) should be used which ensure their performance with minimum fuel consumption and costs. This can be achieved by aggregates completed with efficient up-to-date tractors and tillage machines that are suitable for local conditions and have optimal design and applications parameters. Analytical relationships, deduced as a result of theoretical studies and confirmed by experimental data, allow simulation performance of soil tillage depending on the energetic characters of tractors and tillage machines that determine optimal parameters, such as the design, working speed and working width of the soil tillage tractor-machine aggregates, providing maximum labour efficiency at minimum fuel consumption and costs. During soil tillage the optimum speed of contemporary energy-saturated high-speed wheeled tractors that ensures maximum labour efficiency with minimum consumption of fuel and means is 7-9 km h^{-1}. It is by 25-30% lower than the speed at which these tractors develop maximum draft power (10-12 km h^{-1}).

Keywords: analytical relationships, soil tillage aggregates, optimal working speed, optimal working width.

1. INTRODUCTION

The efficiency of tractors and machines applied in agriculture is usually estimated as an integrated value including the indices of their intensive and extensive use (Lazarev 1967). The application intensity of tractors and machines, as well as their aggregates is characterized by their working capacity per unit of time. However, in order to obtain more objective estimation data for the used tractor aggregates, their performance should be evaluated by optimal parameters: their working width and their speed (Cesnieks et al. 2003; Vilde 1999; Witney 1996). One of the ways how to raise labour efficiency, to cut the fuel consumption and the production costs, as well as to improve the ecological situation is to improve the tractor loading and aggregation patterns (Witney 1996; Vilde 1976, 1997).

The purpose of this study is, by applying simulation methods, to determine the favourable design of tractor-machine units for soil tillage and their optimal parameters: working speed and width.

2. MATERIALS AND METHODS

The objects of research are: the design of the tractor-machine units (aggregates), their parameters and working regimes. Theoretical and experimental studies are carried out in order to determine the energetic characteristics of tractors and tillage machines and simulation their effect on labour efficiency, energy (fuel) consumption, costs and optimal parameters of soil tillage aggregates (Vilde 2004; Rucins et al. 2005).

The capacity of tractors can be calculated multiplying their draft force and take-off force by their shaft power. The energy requirement for soil tillage and other traction operations (pull technologies) may be determined by the specific static and dynamic resistances (Cesnieks et al. 2003; Vilde 1976, 1998). These methods are applied to optimise the parameters of the tractor-machine aggregates.

The energetic characteristics of the tractors are determined by engine power N_e, it specific power $N^{/}$ related to the unit of the tractor mass and specific fuel consumption g_e, by the tractor draft power N_v and their available changes of the working speed and conditions (Vilde 1976, 1997).

The energetic characteristics of the soil tillage machines with passive working parts are determined by their static and dynamic resistance coefficients (Vilde 1976, 1998).

3. RESULTS AND DISCUSSION

3.1. The energetic characteristics of tractors

The specific fuel consumption of correctly aggregated machines performing agricultural operations does not depend on the capacity of the tractor but on its economy, which is determined by the specific fuel consumption of the engine for the production of a unit of energy g_e, g(kWh)$^{-1}$ and the coefficient of its employment for the useful work k_u (soil tillage). The last coefficient k_u characterises which part of the energy produced by the engine is used up in the technological operations. The lower is the specific fuel consumption of the engine and higher the coefficient k_u of the useful

work (for example, the draft coefficient) (Vilde 1997), the more economic may be the work of the tractor. Therefore, in order to save fuel, the tractors with the most economic engines should be used.

An important factor in fuel consumption is the engine loading. Figure 1 presents generalised curves of the diesel engine loads that show the variations in the values of indices characterising the operation of the engine: the total fuel consumption G, the torsional moment (the moment of rotation) M, the number of crankshaft revolutions n and the specific fuel consumption g_e depending on the effective power N_e developed by the engine (also in percentage) (Vilde 1997). It is obvious from the picture that if the engine loading falls, the specific fuel consumption rises, at first, at a slower rate (up to about 80% loading), but further it increases more and more rapidly (Fig. 1).
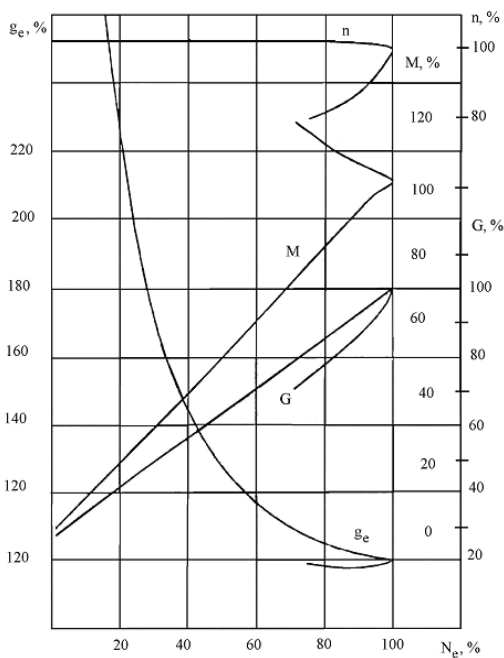

Figure 1: Generalised load curves of the tractor diesel engines: N_e – the efficient power, %; n – the rotational speed of the crankshaft, %; M – the moment of rotation, %; g_e – the specific fuel consumption related to a unit of work of engine, %; G – the total fuel consumption per unit of time, %

Running underloaded engine decreases correspondingly the efficiency of the aggregate, increases the time spent on tilling, as a result of it, salaries, which makes soil tillage still more expensive.

$$g_{ei} = g_{e1} k_{ui}^{-1}, \qquad (1)$$

where g_{e1} – the specific fuel consumption at a total (100%) engine loading.

It follows from what was said before that efficient tractors should be used in order to ensure efficient work, and they should be loaded in a proper way by running them at an optimum speed with machines of proper width (Cesnieks et al. 2003; Vilde 1976, 1997).

It is evident from the graphs (Fig. 1) that the total fuel consumption of the engine per unit of time G_i is a function of the engine loading coefficient k_u:

$$G_i = f(k_{ui}) . \qquad (2)$$

Further, it follows from this correlation that the ratio of the engine loading k_{ui} may be determined by measuring the fuel consumption G_i in a corresponding moment of time and the data saved in the data logger, and their interpretation using an appropriate computer programme.

The next important characteristic of tractors is the productivity of its work. It can be characterised by specific working efficiency w determined as an amount of the performed work W related to a unit of power.

$$w = WN_e^{-1} . \qquad (3)$$

The power N_e of contemporary tractors is generally used for driving the machines with both active and passive working parts:

$$N_e = N_a \eta_a^{-1} + N_p \eta_p^{-1} , \qquad (4)$$

where: N_a and N_p - the power, correspondingly, for driving the machines with active and passive working parts; η_a and η_p – the coefficients of the power transmission efficiency.

The power for driving machines is a function of resistance R and the working speed v:

$$N_a = f(R_a, v) \quad \text{and} \quad N_p = f(R_p, v) . \qquad (5, 6)$$

But there are occasions when a part of the engine power may be used for devices which require a certain driving power irrespective of the working speed of the tractor (for example, the power needed for driving a ventilator, or blowhole). Then there remains less power for driving the machines.

If the tractor is used, for example, only for soil tillage with machines having passive working parts (ploughs, cultivators, etc.), the draft power N_v depends on the draft coefficient η_v of the tractor, which is not a constant value, but depends on the working speed (Vilde 1997, 1976; Kolobov et al. 1972):

$$N_v = N_e \, \eta_v = \eta_{v\,max} \, e^{-c \, (v_o - v)^2}, \qquad (7)$$

where: $e = 2.718$ - (the basis of the natural logarithm);
  v - the working speed of the aggregate;
  v_o - the speed corresponding to the maximum draft capacity, (i.e., the speed corresponding to $\eta_{v\,max}$),;
  c - a coefficient that depends on the physical and mechanical properties of soil and the working capacity (gripping with soil, resistance to movement) of the tractor undercarriage (wheels, caterpillar track);
  $\eta_{v\,max}$ – the maximum draft coefficient of the tractor.

The tractor has the maximal draft power $N_{v\,max}$ at a speed which ensures a maximum draft coefficient.

The values of $N_{v\,max}$, c, v_0 and $\eta_{v\,max}$ were determined by testing.

3.2. The energetic characteristics of the machines used

The efficiency of tractors depends, to a great extent, on the energetic characteristics of the machines used with this tractor.

The amount of energy consumed for soil tillage with the machines having passive operating parts depends on their specific draft resistance (Vilde 1998):

$$E_m = K_1 = k_1{}' + \varepsilon_1 v^2, \qquad (8)$$

where E_m - the specific energy capacity of soil tillage, Nm m^{-2}; K_1 - the specific draft resistance of the machine, N m^{-1}; $k_1{}'$ - a generalised (total) specific static resistance related to a unit of the working width, Nm^{-1}; ε_1 - the dynamic resistance coefficient related to a unit of the working width, N s^2 m^{-3}; v - the working speed of the machine, m s^{-1}.

In terms of the units of measurement used in technical calculations the utilisation of energy obtains the following expression:

$$E_m = 2.778 \cdot 10^{-3} K_1, \quad \text{kWh ha}^{-1}. \qquad (9)$$

To carry out comparative energetic estimation of soil tillage machines, the values of their static and dynamic resistance coefficients are compared, as well as the character of their variations. From the energetic point of view, those machines are better for which the values of the resistance indices are lower.

For the machines with active working parts, in addition to these draft resistance, one must determine the resistance moment (torque moment) too.

3.3. Modes of aggregation, optimisation the parameters of aggregates

Only such aggregates are to be used that ensure the work with a maximum efficiency and minimum fuel consumption. For example, in soil tillage this can be achieved by aggregates of optimum width working at optimum speed. The pure (net) efficiency of the aggregate W is determined by the ratio of the draft efficiency N_v developed by the tractor and the specific efficiency N_1 required to operate the machine (related to 1 m working width) (Vilde 1997, 1976):

$$W = B\,v = v\,k_n\,N_v\,N_1{}^{-1}, \qquad (10)$$

where: k_n - the loading coefficient of the engine (the use of power) ($k_n = 0.75\text{-}0.95$);

B - the working width of the aggregate.

The value of the draft power N_v and of the specific power N_1 required for running the machine varies with the speed of the movement:

$$N_v = N_{v\,max}\,k_{af}\,k_n\,e^{-c\,(v_o - v)^2}; \qquad (11)$$

$$N_1 = 2.778 \cdot 10^{-3}(k_1{}'\,v + \varepsilon_1 v^3), \qquad (12)$$

where: $N_{v\,max}$ - the maximum draft power of the tractor working in the stubble under standard conditions, kW; k_{af} - a coefficient that characterises the impact of soil conditions on the maximum draft power of the tractor.

Since the values of the draft power developed by the tractor and the power that is necessary to run the machine vary with the speed, the labour efficiency W (ha h^{-1}) changes too, as shown in the formula:

$$W = 366.8\,N_{v\,max}\,k_{af}\,k_n\,e^{-c\,(v_o - v)^2}\,(k_1{}' + \varepsilon v^2)^{-1} \qquad (13)$$

By calculating the labour efficiency according to formula (13) at different speeds one can find optimum speed v_{opt} at which the efficiency of the aggregate is the greatest. In this formula the ratio:

$$k_n\,e^{-c\,(v_o - v)^2}\,(k_1{}' + \varepsilon v^3)^{-1} = w \qquad (14)$$

indicates the specific working efficiency w of the soil tillage aggregate determined as an amount of the performed work related to a unit of the tractor draft power.

Knowing the optimum working speed, it is possible to determine from this same relationship (13) the corresponding (optimum) working width of the machine B_{opt} in metres:

$$B_{opt} = 366.8\,N_{v\,max}\,k_{af}\,k_n\,e^{-c\,(v_o - v_{opt})^2}\,(k_1{}' + \varepsilon_1 v^2{}_{opt})^{-1}\,v^{-1}{}_{opt} \qquad (15)$$

The specific fuel consumption Q_0 (kg ha^{-1}) for soil tillage (ploughing, cultivation, harrowing, etc.) can be determined according to the formula (Vilde 1997):

$$Q_0 = 2.778 \cdot 10^{-6} g_e\,\eta_{v\,max}{}^{-1}\,(k_1{}' + \varepsilon_1 v^2)\,e^{c\,(v_o - v)^2} \quad \text{kg ha}^{-1}. (16)$$

Direct costs I are formed from the sum of labour costs a (salaries for the workers with transfers), the expenditure on fuel and lubricants d, the expenses R on repairs of the tractor and machines and technical maintenance, as well as the amortisation deductions A (of the tractor and machines), i.e. deductions for the purchase of new machinery (Vilde 1997):
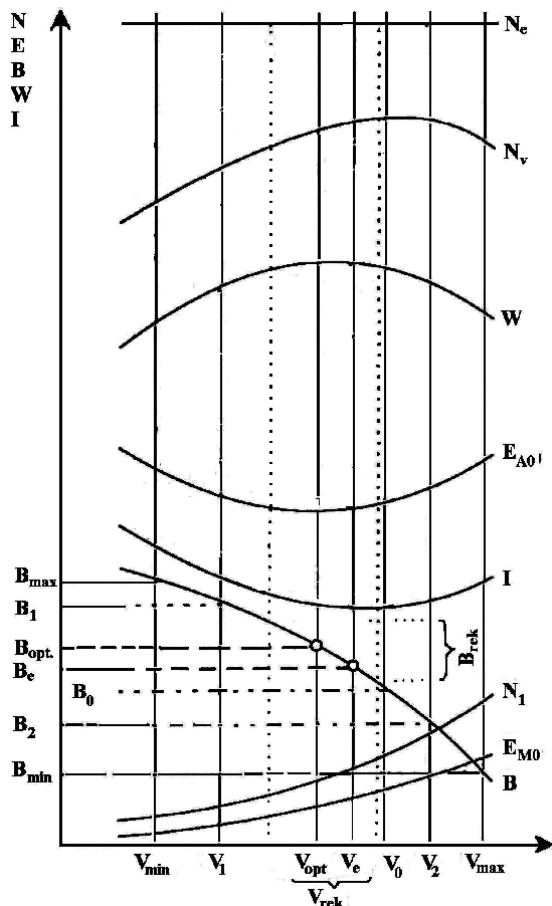
$$I = a + d + R + A. \qquad (17)$$

Figure 2: The scheme of variations of the energetic and economical characteristics of soil tillage aggregates (at a full loading of the tractor engine) depending on the working speed and the width of the machine: N_e – the efficient power of the tractor engine; N_v – the tractor draft power; N_1 – specific power required to run the machine (related to 1 m of its working width); B – the working width of the aggregate; W – the working efficiency of the aggregate; E_{Mo} – the specific energy consumption of the machine (for tilling a unit of area); E_{Ao} – the specific energy consumption of the aggregate (for tilling a unit of area); I – the soil tillage costs; v_{min}, v_{max} – the working speeds (minimum and maximum) corresponding to agrotechnical requirements; v_0 – the speed at which the tractor develops maximum draft power; v_{opt} – the optimum working speed of the aggregate that ensures maximum working efficiency at a minimum energy (fuel) consumption ($v_{opt} < v_0$); v_e-the economic working speed of the aggregate at which the soil tillage costs are minimal; ($v_e > v_{opt}$); B_{opt} – the optimum working width of the aggregate; B_e – the economic working width of the aggregate; v_{rek} – the range of recommended working speeds for the aggregate; B_{rek} – the range of recommended working widths of the aggregate; $v_1...v_2$ – the range of available working speeds of the tractor.

The scheme of variations of the energetic and economical characteristics of the soil tillage aggregates depending on the working speed and width of the machine are given in Figure 2.

It is evident from Figure 2 that the specific power required to operate the machine N_l and the specific energy requirement E_{Mo} increase with the increase in the speed of work v. Increasing the speed, the draft power N_v developed by the tractor rises reaching its maximum, after that it decreases. The specific energy requirement E_{Ao} of the aggregate needed for tilling soil falls with the rise in speed and reaches its minimum, then it rises again (at a full loading of the engine). The soil tillage costs I vary in a similar way. At an optimum speed when the aggregate reaches maximum efficiency the energy requirement and, consequently, fuel consumption and costs of soil tillage are minimal (at a full loading of the engine).

It is also clear from Figures 1 and 2 that, in case the tractor is not loaded (it works with a small width and at a low speed), the efficiency of the aggregate decreases but the energy (fuel) requirement for soil tillage and costs increase correspondingly. Therefore one should work in the range of recommended speeds v_{rek} with the corresponding working widths B_{rek} of the machine at a full loading of the engine.

The optimum speed of contemporary energy-saturated high-speed wheeled tractors that ensures maximum labour efficiency with minimum consumption of fuel and means during soil tillage is 7-9 km h[-1]. It is by 25-30% lower than the speed at which these tractors develop maximum draft power (10-12 km h[-1]) (Vilde 1997, 1976).

The economic working speed at which the costs of soil tillage are the lowest is a little greater (10-15%) than the optimum one. The more expensive the machine is, and the cheaper the labour force and fuel are, the greater is this speed (Cesnieks et al. 2003).

The best aggregates under Latvian conditions are the mounted (hang-up) aggregates, including the wide multi-sectional aggregates, during the operation of which it is possible to transfer their extra weight (in order to perform technological operations) to the tractor using the automatic control system of the tractor hydraulic hitch-up device, hydraulic loaders or other analogous means (the support of the frontal part of the machine on the wheels of the tractor) (Vilde 1997).

The results obtained by simulating the operation of the soil tilling units are confirmed in experimental tests. They show a close relationship among the parameters of aggregation, the mode of operation (speed, engine loading), the efficiency of the soil tilling aggregates, the specific fuel consumption and the expenses.

As an example, variations of energetic and economic indices of ploughing rocky soils with ploughs of different working widths depending on the speed are shown in Figure 3.

It is evident from diagrams (Figure 3) that at a low working speed ($v \approx 5$ km h[-1]) with a plough of insufficient working width the labour productivity of the aggregate decreases almost by half, but the specific fuel consumption increases by 20-30% (3-4 kg ha[-1]), and the costs rise by 20-30% (3.70-7.40 € ha[-1]), in

contrast to the work at an optimum speed with a machine of an optimum width.

It is evident from diagrams (Figure 3), too, that, by their energetic characteristics (the values and alteration of the specific draft resistance K_l), both of the tested ploughs are almost equivalent. However, in case the energetic estimation is given, as it is often erroneously done, according to the specific fuel consumption at a speed which does not load the tractor engine, one can draw a false conclusion that the 2-body plough ATA-2-40 with the KAUR-40 AGS bodies (with the specific fuel consumption of ~20 kg ha^{-1}) is much worse than the other 3-body plough PGP-3-35 with the Kverneland bodies No 8 (with the specific fuel consumption of ~17 kg ha^{-1}) at approximately the same speed ~ 1.5 m s^{-1}.
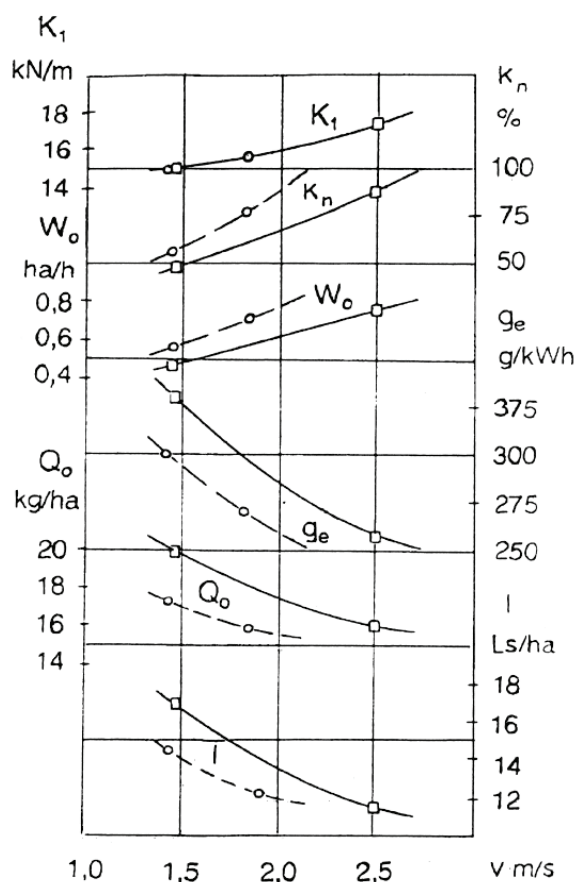

Figure. 3. Variations in energetic and economic characteristics of ploughs for rocky soils with different working widths depending on the speed of ploughing:

K_l – the specific resistance of the plough, kN m^{-1};

k_n – the loading of the tractor engine, %;

W_0 – the efficiency of the direct work, ha h^{-1};

g_e – the specific fuel consumption of the engine, g (kWh)$^{-1}$;

Q_o – the specific fuel consumption, kg ha^{-1};

l – the ploughing costs, Ls ha^{-1} (1Ls ≈ 1.5 €);

□ - a 2-body plough ATA-2-40 with the KAUR-40 AGS bodies;

○ - a 3-body plough PGP-3-35 with the Kverneland bodies No. 8.

The field tests of the improved plough PGP-3-35 on the peasants' farms show the following agro-technical and economical results:

- By replacing the bodies PGC 61000 of the plough PGP-3-35 with the KAUR-40 AGS bodies (the working width of each body 50 cm) and rebuilding it into a 2-body variant (PGP-2-50 with the working width 1.0 m) for more rational work with the tractor MTZ-82 (55 kW), the fuel economy is 15-20%.

- Increasing the working width of the plough bodies by 50 cm (PGP-3-50 with the working width 1.5 m) the fuel consumption of the tractor MTZ-952 (66 kW) decreases by 25-30%.

- There is a possibility of qualitative (up to 96%) and cheap (without previous shredding) introduction of the siderate (catch crop) plants and long-stalk stubble into soil thus saving 1.4-1.8 man h ha^{-1}, fuel 6-8 kg ha^{-1}, the financial resources – 8…11 USD ha^{-1}.

- Rebuilding the ploughs reduces the expenditure of the ploughshares by 30%.

The other bigger ploughs of the PGP family have been rebuilt by a similar scheme (for PGP-7-40 the working width increases from 2.8 m to 3.2…3.6 m).

Similar results are obtained with the cultivation aggregates (Vilde 1998).

The experiments, carried out with the soil tillage tractor-machine aggregates, confirmed the results obtained by simulating their operation.

CONCLUSIONS

1. Analytical relationships deduced as a result of theoretical studies and confirmed by experimental data allow simulation of the operation of soil tillage tractor-machine units in order to determine their optimal parameters, such as the design, working speed and working width, providing maximum labour efficiency at a minimum fuel consumption and costs.

2. The best soil tillage aggregates under the Baltic conditions are tractors with mounted machines, also wide multi-section aggregates during the operation of which it is possible to transfer extra weight of the machines (in order to perform technological operations) to the tractor using the automatic control system of its hydraulic hitch-up device.

3. Improvements of 3-7 body ploughs for rocky soils by equipping them with high-speed semi-helicoidal bodies having an increased operating width up to 50 cm, and the corresponding increase in the total working width of the whole plough for the work with more powerful tractors are technically and agro-technically possible and profitable. This reduces the specific draft resistance of ploughs by 14…26%, raises their efficiency by 25…42%, decreases the specific fuel consumption by 16…24% (3…4 kg ha^{-1}), and lowers correspondingly the ploughing costs.

REFERENCES

Cesnieks, S., Vilde, A., Rucins, A., Cesnieks, A., 2003. Economical efficiency of the parameter optimisation of soil tillage aggregates. *International Scientific Conference "Motor Vehicle, Logistics, Alternative Fuels." Proceedings April 24, 2003.* Jelgava, Latvia, 146–151.

Rucins, A. and Vilde A., 2005. Modelling forces acting on the plough body. *Simulation in Wider Europe. 19th European Conference on modelling and Simulation ECMS 2005 June 1-4, 2005 Riga, Latvia.* Riga, 414-419.

Vilde, A., 1997. Cost-Efficient Soil Tillage. Ulbroka, Latvia, 49 p.

Vilde, A., 2004., Mechanical and mathematical foundations for modelling the dynamics of soil tillage machine operating parts. *TEKA Commission of Motorization and Power Industry in Agriculture, Volume IV.* Polish Academy of Sciences Branch in Lublin. Lublin, Poland, 228–236.

Vilde, A., 1999. Energetic and economic estimation of soil tillage systems. *Wydawnictwo Akademii Rolniczej Szczecin (eds.), Folia Universitatis Agriculturae Stetinensis 195, Agricultura 74.* Poland, 213-222.

Vilde, A., 1998. Energetic estimation of soil tillage machines by testing. Theoretical motivation and methods. *Proceedings of the Latvia University of Agriculture, Vol. 13 (290)* Jelgava, 39-47.

Vilde, A., 1976. Lieljaudas traktoru izmantošana. Rīga: Liesma, 144 lpp. (In Latvian). Using high-power high-speed tractors.

Witney, Brian., 1996. Choosing and using farm machines. *Land Technology Ltd, Edinburgh.* Scotland UK, 412 p.

Лазарев, А.А., 1967. Состояние и пути дальнейшего внедрения комплексной механизации в сельскохозяйственное производство республики. *Труды Латв. НИИМЭСХ, т. I.* Рига, 262–313. (In Russian). Situation and ways for subsequent complex mechanisation of agricultural production in republic.

Колобов, Г.Г., Парфенова, А.П., 1972. Тяговые характеристики тракторов. Москва: Машиностроение, 153 с. (In Russian). The traction characteristics of tractors.

AUTHORS BIOGRAPHY

ARVIDS A. VILDE was born in 1929 in Bauska, Latvia, in a farmer's family. After finishing the secondary school he entered the Latvia Academy of Agriculture, where he studied Agricultural Machinery and obtained the degree of an engineer in 1954. Subsequently he obtained the following degrees: a Candidate of Technical Sciences (SU) in 1965, a Doctor of Technical Sciences (SU) in 1986, a Doctor Habil. Sc. Ing. in 1992. He worked for 5 years as the chief engineer in the field of agricultural machinery and production. Since 1960 up to this time he is engaged in the leading research groups in Latvia Research Institute of Agricultural Machinery. The research branches: soil dynamics in tillage, including the dynamic properties of soil; the technology, machines and units for soil tillage and sugar beet growing, their rationalisation and perfection; the use of big high-speed tractors and machines; energy requirements for field crop production and the ways of its reduction; normative requirements for tractors and machines, their working load and fuel consumption. Now he is leading the research in Precision Agriculture using the GPS, and in the plant spacing simulation by sowing seeds at exact intervals, as well as in the simulation of forces acting on the plough body in order to determine its draft resistance and optimal parameters.

A. Vilde has received several medals and diplomas at the Exhibition of Economic Achievements in Moscow. In 1985 he received the Lavian State Prize. He was named a Merited Inventor of Latvia in 1990, International Man of the Year for 2000-01 and a Latvian State Emeritus Scientist 2001. He is an expert of promotion councils and a publicist who has written more than 800 publications including eighteen monographs. He enjoys orchards and stenography. With his late wife Velta, he has four children and eleven grandchildren. His e-mail address is vilde@delfi.lv

EDMUNDS A. PIRS was born on June 27, 1977 in Tukums. In 1984 he started school at Rainis Gymnasium in Tukums. In 1992 he continued his studies at State Kandava Agricultural Technical School, speciality of agricultural engineering, from which he successfully graduated in 1996. In September, 1996 he entered Latvia University of Agriculture (LUA), the motor transport speciality. During his studies E. A. Pirs successfully passed his production practices and worked at the Baltic Motors and Auto JS service centres of cars, as well at an agricultural enterprise in Norway. In 2001 he successfully graduated from the LUA obtaining the higher education of the first degree with a Bachelor's Degree in Engineering Sciences. In September, 2001 E. A. Pirs continued studies for the Master's Degree at Latvia University of Agriculture, the motor transport speciality. During the academic vacation from May till November, 2004 he worked on a vegetable farm in the State of Michigan, the USA. In March 2005 he started working as a mechanical engineer at the ARMUSS trading enterprise of agricultural machinery. He has attended several courses on tractor machinery in Italy and Great Britain. In 2006 E. A. Pirs graduated from the LUA obtaining the Master's Degree in Engineering Sciences. In 2007 he was promoted in office as the Manager of the Service Department. Since January 1, 2007 he is engaged, at the same time, as an assistant at the Research Institute of Agricultural Machinery, LUA. In June 2007 E. A. Pirs started his doctoral studies at

Latvia University of Agriculture. The syllabus of his doctoral studies is Agricultural Engineering Science. E. A. Pirs is married. His e-mail address is edmunds@armuss.lv

A SHORT GUIDE TO CLIENT'S SATISFACTION FROM A SIMULATION MODEL

Michal Stec[a]

[a]SimulateFirst, Stefan Meier Str. 167, 79104 Freiburg, Germany

[a]stec@simulatefirst.com

ABSTRACT

After 10 years of applying simulation I come to the following conclusion: the top four features of a simulation model, which guarantee client's satisfaction are: ease of use, good animation, automatic model generation and full results.

I show a decent way, which has a relatively high chance to lead every simulation project to a success. The only success factor that matters is client satisfaction, measured by the number of projects done for returning clients.

Keywords: simulation model, animation, ease of use, automatic model generation

1. INTRODUCTION

All the discrete simulation enthusiasts, me including, know how great the simulation technique is and how much it can potentially offer. However, manufacturing industry has been less than successful in the implementation of simulation as a decision support tool (Heavey 2004). This means practically that it is not easy to find and convince a client to a simulation project.

One could expect that every simulation project is better than the previous one. Thus, with time, experience und support from the simulation organizations and other simulation bodies every simulation user should be well advanced in this field after just a few years. However, simulation technique is still not that widely used in the present world.

My short explanation to that dilemma is that manufacturing clients do not want to try this technique on another project. I mean, they might have used it once but they do not want to use it anymore. Or maybe until another simulation consultant brings some hard facts and convinces them that he can do it in a new, better way. Clients may not even clearly state why they refuse simulation, because the formulated goals of the past simulation projects have been even met. However, they feel that they have not gained enough from simulation. They noticed that the technique is difficult to deal with, they do not trust the results and above all it is expensive.

In this article I am not announcing that I found the golden rule about how to carry out simulation projects. Actually, such rules and guidelines for simulation models have been already described in numerous books,

to mention one of the most popular: Law et al (2000) and the summary of how to build credible models in Law (2005). In this article I present a decent practical way, which builds on that simulation knowledge, but which additionally increases a chance of client satisfaction from the simulation project. Client satisfaction is the most important success factor, measured by the number of projects done for returning clients. Only this helps to spread a good word about simulation and guarantees stable development of simulation service.

2. SEEK CLIENT'S SATISFATION

I find that clients are modest in defining goals for a simulation project. After a short presentation they are impressed with examples of my previous models and they think how great the technique is. When we talk about a pilot project they quickly decide that they do not need fancy animation. Instead, they just want a small model with i.e. five machines, which should provide only some simple results. They promise that the model will be extended in the following phases, if only their goals are fulfilled.

Client's assumption is that the simple simulation model leads to a short project timeline, thus with relatively small budget. What he does not know is that the simulation consultant needs to make a relatively big step in learning the client's world, collecting data and drawing processes. The time of such a pilot project is not necessarily short, or if declared short, then there are high chances that it is underestimated, so it lasts longer than planned. After that time the client is no longer satisfied with what he ordered. He expects more.

My top four features that clients always want, although they may not say it up front, are:

- to play with the model (ease of use),
- to see a comprehensive animation,
- to rearrange the model automatically and
- to read the full results down to the smallest detail.

These proposed features make every model a successful one. They simply manifest the power of the simulation tool to the client. The remainder of this paper gives a clue regarding each of these features.

3. EASE OF USE

A simulation model is in fact a standalone computer application, which is designed and developed to bring important findings to the client, fast and meaningful. There are countless applications for making our lives easier and which provide a particular benefit to the user. Simulation could be one of them too. The client would like to use it as he uses Microsoft Office for example. In that case the client would like to use the model even on frequent basis. He expects then that the input data is clearly organized, the animation model doesn't require long time to test one scenario and that the reports are detailed enough to answer his questions. In that case a simulation model could be used even by a simulation novice, but with a good idea what needs to be simulated.

A very good approach to "ease of use" simulation tool is adding a front-end interface to the simulation tool for data collection and report analyses. Such interface could be done for example in Excel from Microsoft. Excel is commonly used by many if not most of decision makers. This fact has been already recognized by many simulation users and simulation software vendors many years ago. So, why not use it as a data keeper for all simulation projects, instead of defining the data in the internal simulation environment. This way the client has one application fewer too learn and he feels like he knows it already for a long time.

	A	B	C	D	E	F	G	H	I
1	Machine	a (m/s2)	a (m/s2)	v (m/s)	v (m/s)	grip	put	from	to
2		a x	a z	v x	v z	s	s	pos id	id
3	Crane 1	2.00	1.00	3.00	1.00	3	3	1	4
4	Crane 2	2.00	1.00	3.00	1.00	3	3	3	6
5	Crane 3	4.00	1.00	4.00	1.00	3	3	6	15
6	Crane 4	2.00	1.00	3.00	1.00	3	3	14	20
7	Crane 5	3.00	1.00	3.00	1.00	3	3	20	25
8	Crane 6	2.00	1.00	3.00	1.00	3	3	25	28
9									
10									

Figure 1: A data table in Excel: machine

A data table in an Excel sheet doesn't have to be fixed to its original size, as it would be in the case of a parameter form in some simulation environment. The tables in Excel can grow wide and long enough for most applications. And all this is done without any work on the development side.

Another advantage of Excel is that it can be used for rapid prototyping of a simulation application. Thus, before sending an offer to a client for a simulation project, the Excel simulation cockpit can be already roughly defined. The client feels then how he will use the simulation model once it is ready. He can press the buttons that will trigger procedure exporting the data to the simulation software. He may see the temporary output data that will be updated after each simulation run. The client sees then what input data is considered and what output will be read from the model. Then he may come up with his own ideas, which can be quickly added to this prototype. This way the two parties, the client and the simulation engineer, come gradually closer to the common understanding of the modeled environment.

How the tables in Excel are organized depends on the model requirements and data hierarchy. I find it useful to draw flowcharts and class diagrams in simplified UML (unified modeling language), but with practice simple data structures can be defined directly in Excel. The more complex the data structures are, the better they should be planned.

Such simulation cockpit becomes a front-end application (1st tier), from which data is transferred to the simulation model (2nd tier). That simulation model may exist then even as a black box to the client, in order to hide its complexity away from him. On the other hand that model should stand out like a neat and well organized structure, in case some quick changes have to be done in front of the client.

Of course, what is easy from the client's point of view does not have to be that straightforward to implement in a simulation model. However, a properly carried out model conceptualization phase (for example in UML) will most likely simplify all client's requirements regarding the project and will help to design a valid simulation model, which receives client's acceptance fairly soon.

The data arrangement in data tables and the relations between the tables are important for the success of a simulation model. All data should be easily accessible from the model level whenever required. This simplifies model construction and enables manual or automatic manipulation with a great deal of parameters. Thus, it makes sense to design the data structure in a way, which resembles a simple database. The rows in tables have their ID, and particular features are referred also with ID. These are, in turn, elaborated in separate tables.

UML is widely used for application development (Fowler 1999). There have been a number of articles written on using UML in simulation (Barijs 2001), a practical tool for model conceptualization phase, providing a dynamic perspective to the information available from modeling tools. It is clear that UML is not "magic" software that will help convince your client to the simulation project. Instead, this tool helps you, as a simulation expert, to plan and design the project structure, so it includes all what the client is asking for. It enables gathering all the data tables, objects of the simulation model and process flows in one place. At this stage it should be possible to answer many design questions, i.e. how are machines, operators and products related to each other? How are they defined in the data tables? Finally how will they be regenerated automatically? This allows to uncover many false assumptions and design pitfalls. This saves time and puts the project on the right track from day one.

The classes from the UML diagram in figure below could be easily represented as tables in the Excel cockpit. This UML diagram has been drawn in Microsoft Visio. Other widely used UML package is Visual Paradigm.
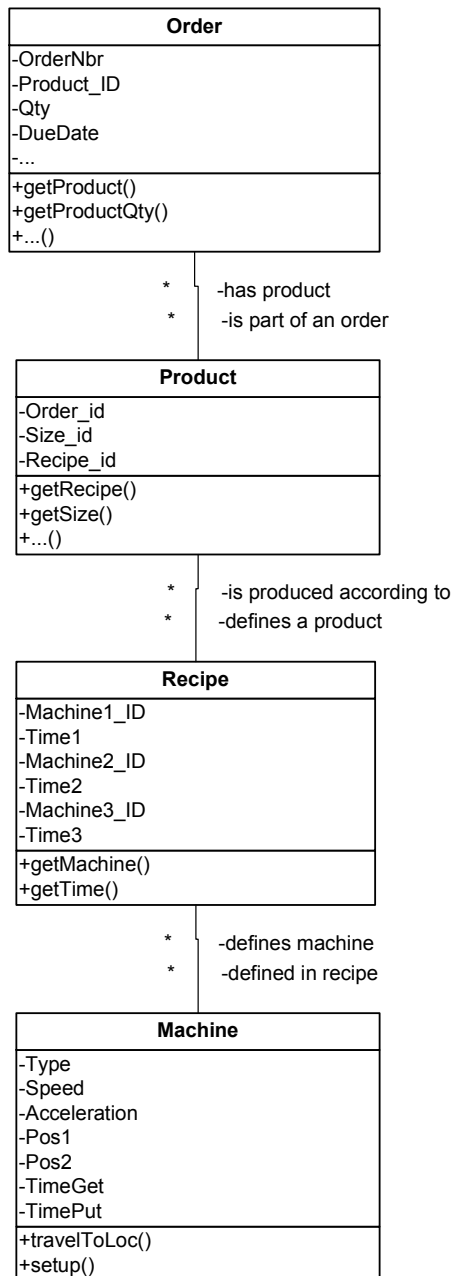
Order
-OrderNbr
-Product_ID
-Qty
-DueDate
-...
+getProduct()
+getProductQty()
+...()

*   -has product
*   -is part of an order

Product
-Order_id
-Size_id
-Recipe_id
+getRecipe()
+getSize()
+...()

*   -is produced according to
*   -defines a product

Recipe
-Machine1_ID
-Time1
-Machine2_ID
-Time2
-Machine3_ID
-Time3
+getMachine()
+getTime()

*   -defines machine
*   -defined in recipe

Machine
-Type
-Speed
-Acceleration
-Pos1
-Pos2
-TimeGet
-TimePut
+travelToLoc()
+setup()

Figure 2: A UML class diagram.

4. ANIMATION

Once a simulation model receives the complete data from the simulation cockpit then it is immediately ready for a simulation run.

It makes sense to help the client understand how the data and logic work together in the model, so he grasps quickly the power of the complex calculations in the simulation model. Explaining the code at this point would make an opposite effect, loose of interest on his side. Instead, the animation does exactly what the spoken explanation would aim for: the client identifies himself with the model; he may run it in a slow pace and check if the model does what required. Without animation the client has little chance to make sure that the model is reliable and does what he expects.

Some models have 2D animation some 3D. Needless to say, the 3D animation is preferred as the client may easier compare the model with what he knows or had seen. The 2D animation is often too abstract, which for the first time viewer is not easy to follow.

The model in figure below is a production machine with a number of cranes moving products to consecutive positions, where the products are processed chemically. The movement of these cranes is not trivial with multiple dependencies of one crane on another. This 3D animation proved here very useful for the client, indeed.
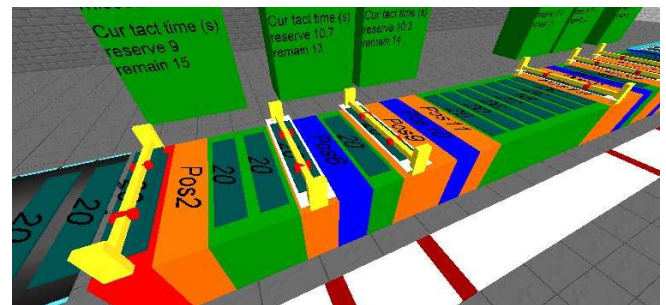

Figure 3: 3D animation of a production machine.

When designing animation for a simulation model it is recommended to ask the client and yourself frequently: what is important in the model, which parameters have to be shown there and how? A good animation usually suffices to get an approval and meet satisfaction level from the client. For example each crane in my example model has a full information table (see figure 4), with throughput, number of serviced parts and other output data.

There are simulation packages on the market, which offer 3D animation built into the simulation environment. Instead of post processing animation type, it becomes rather an integral part of a simulation model. Each time a module is added to the model, it comes with a 3D representation, either drawn from basic 3D shapes or imported from a VRML file. Thus, 3D animation should not take much project time anymore, and it gives a concrete value to the client, which is hard to overestimate.

5. AUTOMATIC MODEL GENERATION

At first the client tries various parameters in experimenting with the model. He changes process times, failure rates and setup times. Soon after that the client asks for adding another machine. And then he hears that this option needs another day to be implemented.

There are obvious extensions to each model, which with little effort come to one's mind: number of resources, areas sizes or paths for traveling objects (Stec 2006). Some of them are simple parameters for existing simulation modules, but other ones are subject to automatic generation of a module in the simulation model.

Figure 4: Information table on machine performance in the 3D model.

At first this may look difficult, in some simulation environments this may not be even possible. But when the simulation environment utilizes modules, which can be fully controlled from code, then short programs written in the proprietary simulation script will do miracles. Moreover, with some practice this appears to be even a better modeling technique than the manual "drag and drop" of the modules followed by manual definition of the module parameters. It is because once the scripts are written they can be reused multiple times.

The Excel cockpit described in the section 3 of this paper has numerous data tables defined for user input. Assuming that there is a "machine" table in the cockpit, the client has a possibility to enter the parameters for the predefined machines in the model, whereas each machine is defined in one row of the table. But then the user may simply want to add another machine to the table, because it looks as easy as filling in another row with another machine data. If so, then why not allowing him for adding it to the model on the fly?

A machine is defined as an object or a module. It has a number of parameters such as location, size, speeds, accelerations, processing time, and many others. The machine has also built in operations like setup, processing and similar. All these parameters and operations can be set and triggered from the simulation script level, thus diminishing the need for manual drag and drop of the modules.

The parameters and type of machine operations to perform are defined in the machine table in the Excel cockpit (see figure 1). Thus, all what is needed is a simple VBA interface in Excel to transfer the data to the simulation model, which, if necessary, will also recreate missing machines there.

Automatic model generation improves greatly model maintenance. This enforces that the model structure is well thought over. Adding statistics or another parameter is no longer an issue. It may require changing only one line of code and regenerating the model.

In the example shown in figure 3, the machine was supposed to run with 30 positions and 6 handling cranes. The simulation experiments shown that actually 32 positions and 5 cranes would produce much better result. Adding the new positions, removing one crane and updating a number of connection and parameters would require normally an hour and a few tests. But, thanks to the automatic model generation, the new simulation model was built within a minute. Needless to say, the client was thrilled.

Last but not least simulation is perceived by many clients as an expensive tool. That expense reduces significantly with every consecutive use. Thus, the client will be always more eager to spend his money on simulation project if the model could be reused on another logistic center, another shop floor in another plant or new factory layout.

The ease of use factor plays again a major role. The easier it is to build a virtual plant in the simulation tool, the less effort it takes to convince the decision maker to use the tool again.

6. FULL RESULTS

After the initial validation of the model based on the model animation, a client may want to check what happens in the model once it is simulated over a long period. The client certainly does not want to test it with real time animation then. Thus, the results collections come here in place very handy. In addition to the usual statistics like min, max, avg, stdev and half width, the user may want to see a plot with what actually happened minute after minute in the simulation. This way he gains an additional confidence in the simulation tool. He might for example see what a particular crane in the model was doing. This could be explained for example by a state diagram like this one below.
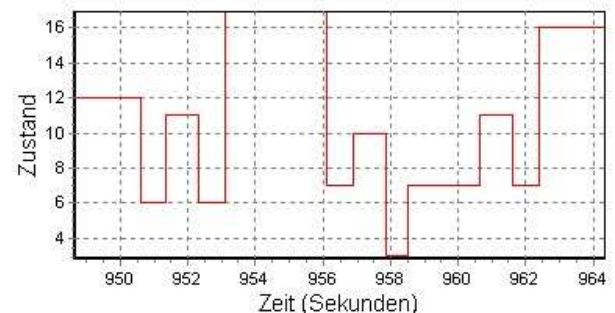

Figure 5: State diagram.

Another useful method for reading detailed results is Gantt chart. The one in figure 6 shows resource

occupation by a product over the whole simulation time. One color in the chart below represents one product ID. One can zoom in and out in this chart to the required detail level.
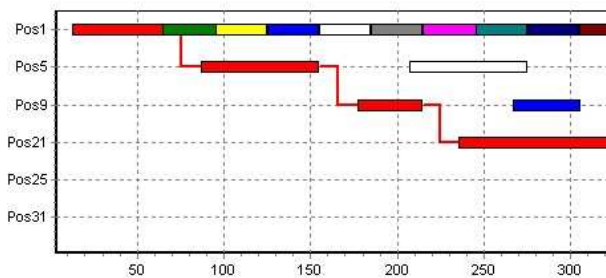

Figure 6: Gantt chart.

Such diagrams give detailed information about the simulation run and answer client's question in a very short time.

There are other diagrams and reporting tools available, which are even more meaningful and faster to grasp than these shown above. It is important then to select those, which suit one's style and experience. It would be best to enclose such a report tool into a module in the simulation software and just insert it to a simulation model when required.

If the client is fully informed about what happened in the simulation model then he can confidently draw valid conclusions from it. And this certainly increases his appreciation for this simulation technique.

7. GO ABOVE LIMITATIONS

I assume that every simulation tool is good in at least two of these features mentioned in this paper: providing parameters for playing with the model and full result set. But what if it is really hard to get good animation in your simulation tool? What if automatic building a model is simply out of question?

In that case it is the features that stand out that must be taken advantage of. Good simulation skills, good sense of what is really important in the simulation project will always help to prove the concept to the client.

But in the long run there is a higher chance of success of a simulation model with good animation against a model with poor one. 3D CAD tools are already commonly used, and the decision makers are used to such graphics. Thus, if your simulation environment allows only 2D then it will probably not make the expected impression on your client. Upgrading your simulation software to match the decent ones on the market, would be here well recommended.

Also a simulation tool with limited model generation capabilities (i.e. lack of detailed programming script) will be losing against a tool where every single operation in the model can be triggered from the code.

There is really no need to say that the simulation does not allow for something or to complain that the client wants too much. The simulation is really a

powerful tool, one just has to open his mind and seek for solutions to the client's questions. Maybe it is just the simulation software that is not capable anymore, and needs to be replaced.

All the points mentioned in this article make the client in favor of using simulation again in the near future. This, in turn, helps to spread a good word about simulation.

REFERENCES

Barjis, J., Shishkov, B., 2001. UML based business systems modeling and simulation. *Proceedings of the Fourth International Eurosim Congress,* pp. 1-6. June 26-29, Delft (The Netherlands).

Fowler, M., Scott, K., 1999. *UML Distilled: Applying the Standard Object Modeling Language.* London, Adddison Wesley Longman Inc.

Heavey, C., McNally, P., 2004. Developing simulation as a desktop resource. *Int. J. Computer Integrated Manufacturing*, Vol. 17, 437-450.

Law, A.M., Kelton, W.D., 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.

Law, A., 2005. How to Build Valid and Credible Simulation Models. *Proceedings of Winter Simulation Conference,* pp. 24-32. Dec 4-7, Orlando, (FL, USA).

Stec, M., 2006. Fast Creation of Realistic and Efficient Free Path Network within a Simulation Model of a Shop Floor and a Supply Chain. *Proceedings of Winter Simulation Conference*, pp. 1531-1536. Dec 3-6, Monterey (CA, USA).

AUTHORS BIOGRAPHY

Michal Stec is a self employed simulation engineer, and the owner of SimulateFirst, a simulation company in Freiburg, Germany. He has received M. Eng. in Informatics from Wroclaw University of Technology, Poland and M. Sc. Eng in Industrial Engineering from National University of Ireland in Galway. He had worked for 6 years as a consultant for IT (Accenture) and discrete simulation (Simulation und Automation Technologies AG), managing successfully a number of simulation projects for known brands. In 2006 he set up SimulateFirst (www.simulatefirst.com), which offers simulation services and which meets clients' satisfaction in numerous applications in production, construction and logistics. He participates actively in simulation conferences; he published and presented articles at Winter Simulation Conference (2006) and at ASIM Symposium "Simulation in Production and Logistics" in Kassel (2006).

SIMULATION OF THE GOMS KEYSTROCK LEVEL MODEL USING DEVS

Ali Mroue[a], Jean Caussanel[b]

(a)(b)Laboratory of sciences of informations and of systems
LSIS UMR 6168 University of Paul Cezanne, Aix-Marseille III
Av. escadrille de Normandie Niemen 13397 Marseille Cedex 20, France

(a)ali.mroue@lsis.org,   (b) jean.caussanel@lsis.org

ABSTRACT
In this paper, we describe our approach for the simulation of GOMS model. The GOMS models characterized human interaction as a goal-directed search, or as general problem solving, in which the shortest route to any goal was considered the most efficient. DEVS is a formalism used to represent Discrete Event System Specifications; it can represent complex systems in an effective way.The DEVS formalism is a powerful tool for model simulation. In order to simulate GOMS models in an effective way, we propose a method to transform GOMS models to DEVS models in order to simulate and improve them. In this paper we will propose the transformation of the Keystroke-Level Model for GOMS into DEVS.

Keywords: DEVS,Goms KeyStrock, HCI

1. INTRODUCTION

In a design at the conceptual level, usability problems are among the most difficult to detect as design proceeds. The designers are unable or may be reluctant to make changes at the conceptual level either for time deadline or because the effort required. Operational prototypes frequently evolve into the final user interface. Evaluating the usability of an interactive system during the early stages of design, e.g. at the conceptual level, is therefore an important task. Predictive usability metrics are an emerging and promising approach to assessing and improving the quality of user interface designs.

One of the important approaches that can be used to predict the usability is GOMS family models, which allows the study of the usability of the interface easly with minimum cost and time, and allow detecting problems before start while still in the design process. There is no existing real simulator for GOMS models, and if it exists it will be in program form, and models will be attached to its application. Another important thing is that GOMS has many limitations, and we are working in improving it, being one of these limits the fact that GOMS models don't take into account the stress, fatigue, etc. of the user.

2. GOMS MODELS

The acronym GOMS stands for Goals, Operators, Methods, and Selection Rules (Card, Newell and Moran 1983). GOMS is a behaviour description model, that lets model the behaviour at different levels of abstraction, from task level to physical actions. GOMS uses as a starting point the Model Human Processor principle of rationality that attempts to model and predict user behavior.

Its essential contribution is a formal structure that allows organizing the design process.
The design method that induces GOMS is done on two axes. In the analysis of task (since determines the behavior). In the predictive evaluation of user behavior in the task.

2.1. GOMS ELEMENTS

Goals, Operators, Methods and Selection Rules (GOMS) are an approach to modeling human computer interaction.

GOMS models consist of descriptions of the methods required to accomplish a specific goal.
Methods are a sequence of operators and sub-goals to achieve a goal. If there exists more than one method to accomplish a goal then selection rules are used to choose which method to use.

- Goals are tasks the system's user wants to accomplish. For example, "Create Folder, Delete Word". A goal can have a hierarchical structure,this means that the achievement of a goal may require to accomplish one or more sub-goals.
- Operators are actions allowed by the software or actions that used are executing. An operator is an atomic level action that can't be composed, and it's characterized by its execution time. The execution of the operators causes change in the mental state of the users or in the environment state. There are two types of operators mental and physical, For example the operators "press enter", "point to the word", etc . . . . Are the physical operators.

326

The model also includes mental operators, such as "thinking", etc. . . . .

- Methods refer to the process that allows one to accomplish a goal. Methods are possible sequences of operators and sub-goals that completely accomplish goals. For example to delete the word just typed, one could press the backspace key repeatedly until all characters have been erased, or i could move one's hand to the mouse, double-click the word and press the delete key. Each of these is a method.

- Selection Rules are used when there exists more than one method that can accomplish the same goal. They are the possible rules by which the user decides which method to choose in order to accomplish a particular goal. A rule has the form:
  If <condition>
  Then use the method M;

Goms has been used in many applications

- Telephone operator (CPM-GOMS)

- CAD systems (NGOMSL)

- Text editing using the mouse (KLM)

2.2. GOMS VARIATION

There are four different models of GOMS: CMN-GOMS, KLM, and NGOMSML CPM-GOMS.
CMN-GOMS stands for Card, Moran and Newell GOMS, is the original version of the GOMS technique in human computer interaction. This technique requires a strict goal-method-operation-selection rules structure.
KLM stands for Keystroke-Level Model based on CMN-GOMS developed by Card, Moran and Newell.

NGOMSML stands for Natural GOMS Language, developed by David Kieras (Kieras 2006). "An NGOMSL model is in program form, and provides predictions of operator sequences, execution time, and time required to learn the methods. Like CMN-GOMS, NGOMSL models explicitly represent the goal structure, and thus can represent high-level goals like collaboratively writing a research paper" (John and Kieras 1996).

CPM-GOMS stands for Cognitive, Perceptual, and Motor and the project planning technique Critical Path Method. CPM-GOMS was developed in 1988 by Bonnie John (John and Kieras 1996). CPM-GOMS does not make the assumption that operators are performed serially, and hence it can model the multitasking behavior that can be exhibited by experienced users. The technique is also based directly on the model human processor a simplified model of human response.

In this paper we will introduce the equivalent of the Keystroke Level Model GOMS in DEVS. Consequently we will only detail the keystroke-level model of GOMS.

2.3. KEYSTROCK-LEVEL MODEL

KLM is a simplified version of CMN-GOMS. It eliminates the goals, methods and rules of selection, leaving only the primitive operators. It contains several claims simplification, it uses operators at the level of primitive keystrokes for prediction. This model seeks to predict efficiency (time taken by expert users doing routine tasks) by breaking down the user's behavior into a sequence consisting of the five primitive operators.

The analyst indicates the method for accomplishing the goal. This method will be composed only of a sequence of strike level primitive operators.

The KLM contains five types of operators:
K - To press a key or button.
P - To point with a mouse to a target on a display device.
H - To home hands on keyboard or other device.
D - To draw a line segment on a grid.
M - To mentally prepare to do an action or closely related series of primitive actions
R - To symbolize the system's response time during which the user has to wait for the system.

Each of these operators has an estimation of their execution time. The time needed to execute a task or to accomplish a goal is the sum of all the time elapsed in the execution of every class of operators.
$Texec = TK + TP + TH + TD + TM + TR$

As we said, every operator is characterized by it's execution time:
TK = (total duration of the tests) / (number of keys hit without error)
20 s worst typist, 0.28 s average typist (40 wpm), 0.08 s best typist (155 wpm)

Pointing is determined by Fitts's Law:
$TP = TP = (K0 - TK) + I \log_2(D/L + 0.5)$

In general $TP \simeq 1.1$ s for all pointing tasks.
$TD = 0.9 \times n \ 0.16 \times l + n$ (n = number of segments to draw, l = sum of the segments' height)

The homing time is estimated by a simple experiment in which the user moves his/her hand back and forth from the keyboard to the mouse. TH = 0.4s

The M operator represents the time to prepare mentally for the next step in the method primarily to retrieve that step (the thing on has to do) from long-term memory. A step is a chunk of the method, so the M operators divide the method into chunks. The resulting estimate (from Card & Moran's paper of 1978) was

1.35 sec_TM = 1.35. TR = 0 if $n \leq t$, TR = n - t if $n \succ t$
n: time required by the system to process the command
t: Is the time required by the user to execute an operator during the processing of the command.

The KLM technique includes a set of five heuristic rules for placing mental operators in order to account

for mental preparation time during a task that requires several physical operators.

0: Insert M

- In front of K's which are not part of an argument string.
- In front of all P's selecting a command.

1: Remove M between fully anticipated operators.

- PMK→PK

2: if a string of MKs belongs to a cognitive unit delete all M but the first.

- 4564.23:MKMKMKMKMKMKMK→MKKK KKKK

3: if K is a redundant terminator then delete M in front of It.

- press Enter, press Enter: MKMK→MKK

4a: if K terminates a constant string (command name) delete the M in front of it.

- cd enter: MKKMK→MKKK

4b: if K terminates a variable string (parameter) keep theMin front of it.

- cd class enter:
  MKKKMKKKKKMK→MKKKMKKKKKM K

2.3.1. USE OF KLM

To use KLM we must perform the three following steps

1. Encode using all physical operators (K, P, H, D(n,l), R(t)).
2. Apply KLM rule number "0" and introduce the M.
3. Apply KLM rule number "1, 2, 3 and 4" to remove the M.

Example:
Goal: Selection of a text
Method:
Grab the mouse
Point the mouse to the desired location
Select

First Step:
H[mouse] P[mouse] K[mouse-button] H[keyboard]

Second Step (Rule 0):
H[mouse] M P[mouse] M K[mouse-button] H[keyboard]

Third Step (Rules 1): H[mouse] M P[mouse] K[mouse-button] H[keyboard]

As a result we obtain: Tmethod = 2TH + TP + TK + TM.

3. DEVS FORMALISM

DEVS (Zeigler 1984) is a formalism used to represent Discrete Event System Specifications; it can represent complex system in an effective way. DEVS model is a powerful simulation model. It is a modular formalism for deterministic and causal systems. It allows for component-based design of complex systems. Several specific platforms for DEVS models simulation can be found.

A DEVS model may contain two kinds of DEVS components: Atomic DEVS and Coupled DEVS. An Atomic DEVS does not contain any component in it. It only has a mathematical specification of its behavior. A Coupled DEVS is a modular composition of one or more Atomic and Coupled DEVS.

An atomic DEVS model has the following structure:

$D = <XD, YD, SD, \delta extD, \delta intD, \lambda D, taD>$

Where:

XD: is the set of the input ports and values
YD: is the set of the output ports and values
SD: the sequential state set
$\delta extD: Q \times X \to S$, is the external transition function where $Q = (s, e) \to s \in S, 0 \le e \le ta(s)$ is the total state set.
$\delta intD: S \to S$, is the internal transition function
$\lambda D: S \to Y$, is the output function
$taD: S \to R+0, \infty$ (non-negative real), is the time advance function

A coupled model, also called network of models, has the following structure:

$N = <X, Y, D, \{Md /d \in D\}, EIC, EOC, IC>$

X and Y definitions are identical to XD and YD of an atomic model. The inputs and outputs are made up of ports. Each port can take values and has its own field of values.

$X = \{ (\rho,\mu) / \rho \in IPorts, \mu \in X\rho \}$
$Y = \{ (\rho,\mu) / \rho \in OPorts, \mu \in Y\rho \}$

D is the set of the model names involved in the coupled model.

Md is a DEVS model. The variables representing the inputs and the outputs of the model will be indexed by the model identifier. Hence the following notation:

$Md = <Xd, Yd, Sd, \delta extd, \delta intd, \lambda d, tad>$

The inputs and the outputs of the coupled model are connected to the inputs and outputs of the included models.

$EIC = \{((N, a), (d, b))/a \in IPorts, b \in IPortsd\}$

The set of the coupled model input ports ipN associated with the input ports ipd of the models D are the components of the coupled model.

There is the same situation for the output ports.

$EOC = \{((N, a), (d, b))/a \in OPorts, b \in OPortsd\}$

Inside the coupled model, the outputs of a model can be coupled with the inputs of the other models. An output of a Model cannot be coupled with one of its inputs.

IC = {((i, a), (j, b))/i,j ∈ D, i, j,a ∈ OPortsi, b ∈ IPorts j }.

4. GOMS TO DEVS

We chose to transform GOMS models to the DEVS formalism for many reasons some of these reason are:

1. They have many common properties. GOMS Goal has hierarchical structure; DEVS have a hierarchical structure "Coupled Model". GOMS has an atomic element named operator that can't be decomposed and has an execution time property, the same for DEVS models, which have an atomic element named state that has its time life period assigned etc.

2. DEVS is very powerful in simulation, and we have a powerful simulator for DEVS models named LSISDME.

GOMS has some limitations, like for example that it doesn't take into account the fatigue of the user. So in the work presented in (Seck, Frydman and Giambiasi 2004).

3. a model for fatigue and stress of users is defined under DEVS. So the transformation of GOMS models to DEVS will help us to extend GOMS in order to add the model of fatigue and stress to GOMS.

5. TRANSFORMATION

In this paper we present only the transformation of KLM into DEVS models.

After studying the two approaches "DEVS-KLMGOMS" we propose the following rules for transformation:

- A KLM-GOMS model Corresponds to an atomic DEVS model that has 1 input and 1 output.
- The Goal is the name of the DEVS model
- The number of phases in this model is equivalent to the number of operators + two phases, one for the initial phase and the other for the end phase. The Life time of each phase is equivalent to the execution time of the corresponding operator.
- The phases are connected together by internal transitions.

Applying these rules to a KLM-GOMS example will give:

KLM GOMS:
Goal:
Select Text
Method:
H[mouse]MP[mouse]K[mouse button]H[keyboard]

The above KLM example mentioned is equivalent in DEVS to (Figure 1). Which is an atomic DEVS model having 2 initial phases "init and End phases" plus "n" other phases. The n represent the number of operators used in the model etc…
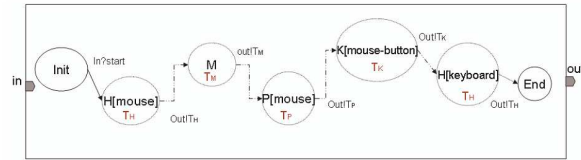

Figure 1: DEVS "Select Text" KLM Goal

6. RESULTS

By giving the equivalence for GOMS keystrock level model in DEVS, we have added many features to its initial representation.

First of all the proposed keystrock DEVS representation gives an effective way for simulating keystrock model. This simulation can be done automatically in an easy way using our laboratory simulator "LSISDME".

Secondly, any change or add of new operators can be done easily without any modification or recompilation of the simulation tool "LSISDME". Other simulators have to be modified and recompiled in order to take into account new operators or new functions.

Another important point for the transformation to DEVS is that we have now the capability to add existing models such as the stress model defined by (Seck, Frydman and Giambiasi 2004) and take into account this human factor. Note that coupling the Keystrock and the stress models will extend the original one (Keystrock Level GOMS) and solve some of the disadvantages of GOMS models.

7. CONCLUSION

In this paper we introduced GOMS models and the DEVS formalism; we explained the reasons of our choice for the DEVS formalism for simulation GOMS Model. We have detailed the KeyStrock Level Model, and we defined all its' elements in the DEVS formalism. In this work we succeeded in finding a new simulation tool for the Keystrock Level Model that is effective and gives good results. This transformation into DEVS has many advantages, first we have obtained an effective simulation tool for Keystrock level models, and secondly we can easily interpret these models and extend them in order to add many other modules. We can also do the simulation in interactive mode, and implement the core of the DEVS simulator in a user training applications. The current prospects relate to complete transformation of all GOMS model family to the DEVS formalism in order to simulate it, and then extend it so as to take in consideration the user stress and fatigue, Also we will model the task of search of information on the web and simulate it.

REFERENCES

Card, S. K.; Newell, A. and Moran, T. P. 1983. *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum Associates, Inc. Mahwah, NJ, USA.

Kieras, D.E. 2006. *A Guide to GOMS Model Usability Evaluation using GOMSL and GLEAN4.* University of Michigan

John B. E. and Kieras D. E. 1996. *The GOMS family of user interface analysis techniques: comparison and contrast.* ACM Transactions on Computer-Human Interaction, 3(4):320–351.

Seck M., Frydman C., and Giambiasi N. 2004. *Using DEVS for modeling and simulation of human behavior.* In Lecture Notes in Computer Science special issue DEVS modelling and simulation, volume 3397, page 692, février 2005. Pub. Springer Verlag - Ed. Tag Gon Kim - ISSN : 0302–9743 - ISBN : 3–540-24476-X

Zeigler B. P. 1984. *Theory of Modelling and Simulation.* Krieger Publishing Co., Inc. Melbourne, FL, USA.

SIMULATION OF NATURAL PHENOMENA BY CELLULAR AUTOMATA WITH THE LIBAUTOTI LIBRARY: AN APPLICATION TO GEOLOGICAL MODELLING

William Spataro[a], Giuseppe Spingola[b], Giuseppe Zito[c], Donato D'Ambrosio[d],
Rocco Rongo[e], Maria Vittoria Avolio[f] , Salvatore Di Gregorio[g]

[a] [b] [c] [d] [f] [g]Department of Mathematics and HPCC, University of Calabria, Italy
[e] Department of Earth Sciences and HPCC, University of Calabria, Italy

[a]spataro@unical.it, [b]g.spingola@gmail.com, [c]giuseppe.zito@gmail.com, [d]d.dambrosio@unical.it, [e]rongo@unical.it,
[f]avoliomv@unical.it, [g]dig@unical.it,

ABSTRACT

This work presents the application of libAuToti, an open-source parallel library for implementing models based on the Cellular Automata approach, for the simulation of geological processes. In particular, we describe the implementation of two models for lava flows and debris flows, as defined by the SCIARA and SCIDDICA models, respectively. Cellular Automata are parallel computing models which are continuously gaining attention from the Scientific Community for their potentiality and efficiency. libAuToti permits a straightforward and simple implementation of Macroscopic Cellular Automata models, which are appropriate for the simulation of spatial extended dynamical systems. Experiments have demonstrated the elevated computational efficiency of the library, executed both on an HPC machine and a standard multi-core PC, confirming the reliability of the library and goodness of simulation results.

Keywords: natural phenomena simulation, cellular automata, lava flows, debris flows, parallel computing

1. INTRODUCTION

Cellular Automata (CA) are discrete dynamical systems, widely utilized for modelling and simulating complex systems, whose evolution can be described in terms of local interactions. Well known examples are Lattice Gas Automata and Lattice Boltzmann models (Succi 2004), which are particularly suitable for modelling fluid dynamics at a microscopic scale. However, many natural phenomena are difficult to be modelled at such scale, as they generally evolve on very large areas, thus needing a macroscopic level of description. Moreover, they may be also difficult to be modelled through standard approaches, such as differential equations (McBirney and Murase 1984), and Macroscopic Cellular Automata (MCA) (Di Gregorio and Serra 1999) can represent a valid alternative.

Among the above mentioned phenomena, lava flows and debris flows may involve serious dangers for people security and property, and their forecasting could significantly decrease these hazards, for instance by simulating paths and evaluating the effects of control works (e.g. embankments or channels).

As regards MCA, *libAuToti* (Spataro et al. 2008) has proven to be an efficient and flexible simulation parallel library. Written in C++, it allows for a simple and concise definition of both the transition function and the other characteristics of the MCA model definition. Moreover it was projected for both sequential and parallel execution, both on shared and distributed memory machines (thanks to the adoption of the Message Passing paradigm for the inter-processes communications), by completely hiding parallel implementation issues to the user.

In the following, MCA are briefly presented and the main characteristics of the *libAuToti* library illustrated. Two examples of application are also shown, which regard the implementation of the MCA basic lava flow model SCIARA and the MCA landslide model SCIDDICA version "T". A general discussion on the results and on the future perspective of the library concludes the paper.

2. MACROSCOPIC CELLULAR AUTOMATA: THE SCIARA AND SCIDDICA MODELS

As previously stated, CA are dynamical systems, discrete in space and time. They can be thought as a regular *n*-dimensional lattice of sites or, equivalently, as an *n*-dimensional space (called cellular space) partitioned in cells of uniform size (e.g. square or hexagonal for *n*=2), each one embedding an identical finite automaton. The cell state changes by means of the finite automaton transition function, which defines local rules of evolution for the system, and is applied to *each* cell of the CA space at discrete time steps. The states of neighbouring cells (which usually includes the central cell) constitute the cell input. The CA initial configuration is defined by the finite automata states at time *t*=0. The global behaviour of the system emerges, step by step, as a consequence of the simultaneous application of the transition function to each cell of the cellular space.

When dealing with the modelling of spatial extended dynamical systems, MCA can represent a valid choice especially if their dynamics can be

described in terms of local interaction at macroscopic level. Well known examples of successful applications of MCA include the simulation of lava (Crisci et al. 2004) and debris flows (Di Gregorio et al. 1999) forest fires (Trunfio 2004), agent based social processes (Di Gregorio et al. 2001) and highway traffic (Di Gregorio et al. 1996), besides many others.

By extending the classic definition of Homogeneous CA, MCA facilitate the definition of several aspects considered relevant for the correct simulation of the complex systems to be modelled. In particular, MCA provide the possibility to "decompose" the CA cell state in "substates" and to allow the definition of "global parameters". Moreover, the dynamics of MCA models (especially those developed for the simulation of complex macroscopic physical systems such as debris or lava flows) is often "guided" by the "Minimisation Algorithm of the Differences" (cf. Di Gregorio and Serra 1999), which translates in algorithmic terms the general principle for which natural systems leads towards a situation of equilibrium. Refer to Di Gregorio and Serra (1999), for a complete description of the algorithm, besides theorems and applications.

2.1. The MCA lava flow model SCIARA

SCIARA is a family of bi-dimensional MCA lava flow models, successfully applied to the simulation of many real cases, such as the 2001 Mt. Etna (Italy) Nicolosi lava flow (Crisci et al. 2004), the 1991 Valle del Bove (Italy) lava event (Barca et al. 1993) which occurred on the same volcano and employed for risk mitigation (D'Ambrosio et al. 2006). In this work, the basic version of SCIARA (Barca et al. 1993) was considered and its application to the 2001 Nicolosi (Sicily) event shown.

SCIARA considers the surface over which the phenomenon evolves as subdivided in square cells of uniform size. Each cell changes its state by means of the transition function, which takes as input the state of the cells belonging to the von Neumann neighbourhood. It is formally defined as

$$SCIARA = < R, X, Q, P, \sigma >$$

where:

- R is the set of points, with integer coordinates, which defines the 2-dimensional cellular space over which the phenomenon evolves. The generic cell in R is individuated by means of a couple of integer coordinates (i, j), where $0 \leq i < i_{max}$ and $0 \leq j < j_{max}$.
- $X = \{(0,0), (0, -1), (1, 0), (-1, 0), (0, 1)\}$ is the von Neumann neighbourhood relation, a geometrical pattern which identifies the cells influencing the state transition of the central cell. The neighbourhood of the generic cell of coordinate (i, j) is given by

$$V(X, (i, j)) =$$
$$= \{(i, j)+(0,0), (i, j)+(0, -1), (i, j)+(1, 0),$$
$$(i, j)+(-1, 0), (i, j)+(0, 1)\} =$$
$$= \{(i, j), (i, j-1), (i+1, j), (i-1, j), (i, j+1)\} \quad (1)$$

- Q is the set of cell states; it is subdivided in the following substates:

  - Q_z is the set of values representing the topographic altitude (m);
  - Q_h is the set of values representing the lava thickness (m);
  - Q_T is the set of values representing the lava temperature (K°);
  - Q_o^5 are the sets of values representing the lava outflows from the central cell to the neighbouring ones (m).

  The Cartesian product of the substates defines the overall set of state Q:

  $$Q = Q_z \times Q_h \times Q_T \times Q_o^5$$

- P is set of global parameters ruling the CA dynamics:

  - $P_T = \{T_{vent}, T_{sol}, T_{int}\}$, the subset of parameters ruling lava viscosity, which specify the temperature of lava at the vents, at solidification and the "intermediate" temperature (needed for computing lava adherence), respectively;
  - $Pa = \{a_{vent}, a_{sol}, a_{int}\}$, the subset of parameters which specify the values of adherence of lava at the vents, at solidification and at the "intermediate" temperature, respectively;
  - p_c, the cooling parameter, ruling the temperature drop due to irradiation;
  - p_r, the relaxation rate parameter, which affects the size of outflows.

- $\sigma : Q^5 \rightarrow Q$ is the deterministic cell transition function. It is composed by four "elementary processes", briefly described in the following:

  σ1. *Outflows computation*. It determines the outflows from the central cell to the neighbouring ones by applying the minimisation algorithm of the differences; note that the amount of lava which cannot leave the cell, due to the effect of viscosity, is previously computed in terms of adherence. Parameters involved in this elementary process are: P_T and P_a.

  σ2. *Lava thickness computation*. It determines the value of lava thickness by considering the mass exchange among the cells. No parameters are involved in this elementary process.

  σ3. *Temperature computation*. It determines the lava temperature by considering the

temperatures of incoming flows and the effect of thermal energy loss due to surface irradiation. The only parameter involved in this elementary process is p_c.

σ4. *Solidification*. It determines the lava solidification when temperature drops below a given threshold, defined by the parameter T_{sol}.

2.2. The MCA debris flow model SCIDDICA

SCIDDICA is a family of bi-dimensional MCA debris flow models, successfully applied to the simulation of many real cases, such as the 1988 Mt. Ontake (Japan) landslide (Di Gregorio et al. 1999), and the 1998 Sarno (Campania, Italy) disaster (D'Ambrosio et al. 2003). In this work, the basic version "T" of SCIDDICA (Avolio et al. 2000) was considered and its application to the 1992 Tessina (Italy) landslide shown. For a more in-depth knowledge of debris flow modelling using MCA approach, please refer to (D'Ambrosio et al. 2003)

SCIDDICA considers the surface over which the phenomenon evolves as subdivided in square cells of uniform size. Each cell changes its state by means of the transition function, which takes as input the state of the cells belonging to the von Neumann neighbourhood. It is formally defined as

$$SCIDDICA = < R, X, Q, P, \sigma >$$

where:

- R is the set of points, with integer coordinates, which defines the 2-dimensional cellular space over which the phenomenon evolves. The generic cell in R is individuated by means of a couple of integer coordinates (i, j), where $0 \leq i < i_{max}$ and $0 \leq j < j_{max}$.

- $X = \{(0,0), (0, -1), (1, 0), (-1, 0), (0, 1)\}$ is the von Neumann neighbourhood relation, a geometrical pattern which identifies the cells influencing the state transition of the central cell. The neighbourhood of the generic cell of coordinate (i, j) is given by

$$V(X, (i, j)) =$$
$$= \{(i, j)+(0,0), (i, j)+(0, -1), (i, j)+(1, 0),$$
$$(i, j)+(-1, 0), (i, j)+(0, 1)\} =$$
$$= \{(i, j), (i, j-1), (i+1, j), (i-1, j), (i, j+1)\} \quad (2)$$

- Q is the set of cell states; it is subdivided in the following substates:
    - Q_z is the set of values representing the topographic altitude (i.e. elevation);
    - Q_h is the set of values representing the debris thickness;
    - Q_o^5 are the sets of values representing the debris outflows from the central cell to the neighbouring ones (recall that the central cell is part of its neighbourhood).

    The Cartesian product of the substates defines the overall set of state Q:

$$Q = Q_z \times Q_h \times Q_o^5$$

- P is set of global parameters ruling the CA dynamics:
    - p_a is the parameter which specifies the thickness of the debris that cannot leave the cell due to the effect of adherence;
    - p_z is the critical altitude, defining two regions of different rheological behaviours of the flow; it is related to parameters p_f;
    - p_f is the "friction angle" parameter, which empirically defines the minimum slope between two cells needed for debris motion - its value is related to that of the parameter p_z;
    - p_r is the relaxation rate parameter, which affects the size of outflows (cf. section above).

- $\sigma : Q^5 \to Q$ is the deterministic cell transition function. It is composed by two "elementary processes":
    - $\sigma_1 : (Q_z \times Q_h)^5 \times p_a \times p_z \times p_f \times p_r \to Q_o^5$ determines the outflows from the central cell to the neighbouring ones by applying the minimisation algorithm of the differences. In brief, a preliminary control eliminates those cells for which the slope with respect the central cell is less than p_f. Moreover, the amount of debris which cannot leave the cell, due to the effect of viscosity, is simply modelled by means of the parameter p_a. Thus, by means of the minimization algorithm, outflows $q_o(0,i)$ ($i=0,1,...,4$) are evaluated and the substates Q_o^5 accordingly updated. Note that $q_o(0, 0)$ represents the amount of debris that does not flow out of the central cell. Eventually, a relaxation rate factor, $p_r \in]0,1]$, can be considered in order to obtain the local equilibrium condition in more than one CA step. This can significantly improve the realism of model as, in general, more than one step may be needed to displace the proper amount of debris from a cell towards the adjacent ones. In this case, if $f(0,i)$ ($i=1, ..., 4$) represent the outgoing flows towards the 4 adjacent cells, the resulting outflows are given by $q_o(0,i)=f(0,i)\cdot p_r$ ($i=1, ..., 4$), while the amount of debris remaining in the central cell is obtained as:

$$h_r(0) = q_0(0,0) = h(0) - \sum_{i=1}^{4} q_0(0,i) \quad (3)$$

- $\sigma_2 : (Q_o^5)^5 \to Q_h$ determines the value of debris thickness inside a cell by considering mass exchange in the cell neighbourhood: $h(0) = \sum_i (q_o(0,i) - q_o(i,0))$, where $i = 0,1,...,4$ and $q_o(i,0)$ represents the inflow from the ith adjacent cell. The substate Q_h is accordingly updated. Note that no

parameters are involved in this elementary process.

3. THE LIBAUTOTI LIBRARY FOR MACROSCOPIC CELLULAR AUTOMATA

The MCA approach permits to straightforwardly define a simulation model of a complex system, such as a lava flow. Moreover, the MCA models of complex systems often need to be the most efficient possible since, depending also on the size of input data, each simulation can last days or even weeks (cf. D'Ambrosio et al. 2006). As a consequence, the developer must implement proper optimization strategies (e.g. cf. Walter and Worsch 2004) and, when mandatory, parallelize the program (e.g. by means of MPI – Message Passing Interface).

Unfortunately, the world of development tools for MCA suffers of a lack of open-source software and often the developer must pay attention to all the low levels details of the implementation, such as memory allocation/de-allocation, I/O management and so on.

To overcome these difficulties, the developer of a MCA model could decide to uneffortlessly implement her/his own simulation model by scratch using a high level language (such as C++ or FORTRAN) or consider a valid (but not-free) software solution, namely the CAMELot CA simulation environment. Advantages of the last solution include the use of a proprietary language (i.e. the CARPET one, cf. Spezzano and Talia 1998) for model definition, integrated 2D/3D viewer and the possibility to run the simulation in parallel by means of the Message Passing paradigm, in a completely transparent manner to the user (i.e. she/he does not need to care of parallel issues such as process allocation on nodes, data partitioning, global-reduction operations, etc). However, disadvantages include the not extensibility of the library, lack of adequate debug facilities and reduced 3D visualization capabilities.

As consequence of all these last considerations, the *libAuToti* library can represent a valid solution. Differently to CAMELot, *libAuToti* is not an integrated simulation environment, but an ANSI C++ (thus portable) library, which intends to offer the main features of CAMELot, as the possibility to simply define the model by ignoring low level details, manage input and output data and execute the simulation both sequentially or in parallel by adopting the Message Passing paradigm.

Differently to CAMELot, the developer produces a standard C++ program by including the *libAuToti* library header file, with all the advantages that this approach permits. Among these, the possibility to choose the preferred development environment with the related debug facilities (not very functional in CAMELot), and the possibility to easily introduce further features such as an optimization module or a 3D viewer. In the following examples, CA model development is illustrated by considering the *libAuToti* library on the SCIARA lava flows and SCIDDICA debris flows models. Furthermore, computational

results of tests carried out on an Alpha Server SC45 supercomputer and on an Intel dual-core PC are illustrated and commented.

3.1. A libAuToti implementation of SCIARA

As stated above, the library has been tested by considering the Macroscopic Cellular Automata model SCIARA for lava flow simulation. In the following, a seamless parallel implementation of SCIARA is presented. In particular, Figure 1 and 2 illustrate the CA definition and transition function, respectively. Besides common C++ libraries such as `iostream`, `math` or `vector` (here omitted for briefness), the program includes the `libAuToti.h` header that allows to employ the library. In particular, the object `sciara` is defined, representing a MCA with the possibility to perform a parallel simulation. As it can be easily seen, the user does not need to concentrate in parallel issues, such as the CA space decomposition or message passing of boundary "ghost" cells between nodes: all these operations are transparently carried out by the library. Moreover, once the program has been compiled, the user has to simply decide how many processing elements the CA has to run on (e.g. `mpirun -np 4 sciara` for automatically executing sciara in parallel on 4 processing nodes). Note that, in order to perform a sequential simulation, it will be sufficient to define an object of type `SeqSimulationCA_2D`.

```
// header files here...
#include "libAuToti.h"

int main(int argc, char** argv)
{
   ParallelSimulationCA_2D sciara;
   sciara.setDim(410,296);

   //flow substates
   sciara.addSubstate(SUBSTATE_FLOAT);// 0
   sciara.addSubstate(SUBSTATE_FLOAT);// 1
   sciara.addSubstate(SUBSTATE_FLOAT);// 2
   sciara.addSubstate(SUBSTATE_FLOAT);// 3
   //elevation substate
   sciara.addSubstate(SUBSTATE_FLOAT );//4
   //lava thickness substate
   sciara.addSubstate(SUBSTATE_FLOAT); //5
//lava temeperature substate
   sciara.addSubstate(SUBSTATE_FLOAT); //6


   sciara.setRadius( 1 );

   // von Neumann neighbourhood
   sciara.addNeighboor( 0,-1 );
   sciara.addNeighboor( 1, 0 );
   sciara.addNeighboor( -1, 0 );
   sciara.addNeighboor( 0, 1 );

   sciara.addParameter(1373); //vent temp.
   sciara.addParameter(1153); //solidif. temp.
   sciara.addParameter(1265); //interm. temp.
   sciara.addParameter(0.7); //vent adherence
   sciara.addParameter(7); //solidif. adher.
   sciara.addParameter(1.5); //interm. adher.
   sciara.addParameter(1.4e-14); //cool. par.
   sciara.addParameter(0.6); //relaxation

   sciara.initialize(argc, argv);
```

```
sciara.readFromFile( "./nicolosi.txt", 4);
sciara.readFromFile( "./vent_pos.txt", 5);

sciara.setNumStep( 18000 );

sciara.run();

sciara.finalizeSimulation();
return 0;
}
```

Figure 1: The SCIARA CA definition by using the *libAuToti* free MCA library.

The member function `setDim` defines the cell space dimensions, while `setRadius` and `addNeighboor` the CA the neighbourhood radius and relation, respectively.

Table 1: List of all allowed type for the substates definition together with their corresponding C++ types.

libAuToti substate type	Corresponding C++ type
SUBSTATE_CHAR	char
SUBSTATE_SHORT	short
SUBSTATE_INT	int
SUBSTATE_LONG	long
SUBSTATE_FLOAT	float
SUBSTATE_DOUBLE	double

As required by SCIARA seven substates are defined thanks to the `addSubstate` member function. Table 1 lists all the allowed substates type and their identifiers. Once a substate is defined, a numerical handle is automatically defined in order to refer the substate in the program. However, to simply refer to a substate, the user can declare an enumerative type; for the case of SCIARA it could be the following:

```
enum Subs{
    FLOW0 = 0,  //outflow Qo towards the N cell
    FLOW1,  //outflow Qo towards the E cell
    FLOW2,  //outflow Qo towards the W cell
    FLOW3,  //outflow Qo towards the S cell
    ELEV,   //cell elevation a.s.l Qz
    TEMP,   //cell temperature (K°)
    THICK   //lava thickness Qh inside cell
};
```

In this way, it is possible to refer to a substate through a symbolic identifier, instead of a numerical one.

As requested by the considered version of SCIARA, the von Neumann neighbourhood was defined, thanks to the `addNeighboor` member function. Note that, similarly for the case of substates, a numerical handle is implicitly assigned to each cell of the neighbourhood, except for the central cell that the library directly refers (cf. below). The library also permits the definition of more "sophisticated" neighbourhoods, as Moore's one (both square or hexagonal) by the way required by the latest releases of SCIDDICA (D'Ambrosio et al. 2003, Iovine et al. 2005).

Parameters are defined by means of the `addParameter` member function. As before, a numerical handle is automatically defined by the library, and the user can define an enumerative type to better manage parameters setting and usage; it could be the following (note that the first 8 correspond to SCIARA parameters shown in Section 2.1):

```
enum Params{
    TEMP_V = 0,    //temp. at vent
    TEMP_S,        //temp. at solidification
    TEMP_I,        //interm. temp
    ADHERENCE_V,   //adherence at vent
    ADHERENCE_S,   //adherence at solid.
    ADHERENCE_I,   //interm. adherence
    COOL,          //cooling parameter
    RELAXATION,    //relaxation
    IS_VENT,       //cell is a crater?
    VENT_THICK     //crater lava thickness
};
```

Parameter setting and retrieval are managed by the two self-explanatory member functions `setParameter` and `getParameter`, respectively. Moreover, input data is automatically performed by the function `readFromFile`, which reads files in standard ASCII Grid format.

Eventually, the simulation is performed (for 18000 steps) by means of the function `run()`.

Figure 2 shows the *libAuToti* implementation of the SCIARA transition function, defined as an inline member function. Recall that, by CA definition, this function is called at *each* step. Substates identifiers, together with the neighbours ones, are utilised as arguments of `getSubstate*`, which returns the value of a certain substate for the specified neighbouring cell. Here the symbol `*` can be one of the types listed in Table 1.

```
inline      void      parallelSimulationAC2D      ::
transitionFunction()
{
// set lava width for crater
if(getSubstateFloat(ELEV)==getParameter(IS_CRATER))
    updateCellSubstate(THICK,
            getParameter(vent_thick));

switch (STEP%2){
    case 1: calc_flows(this);
        break;
    case 0: calc_thickness(this);
        calc_temperature(this);
        calc_quote(this);
        break;
    }
}
```

Figure 2: The SCIARA transition function definition by using the *libAuToti* free MCA library.

For instance, the following portion of code verifies if the value of the substate `ELEV` of type `float`, is equal to the `IS_CRATER` parameter, meaning that the cell itself represents a crater:

```
if(getSubstateFloat(ELEV)==getParameter(IS_CRAT
ER))
```

Analogously, if one wants to check the value of a neighbouring cell, she/he has to simply add the neighbour index as second argument to the

`getSubstateFloat` function. The following example could test if the value of the substate `THICK` of the neighbour number 2 is equal to 0:

```
if (getSubstateFloat( THICK, 2 ) == (float)0 )
```

Equivalently, the function `updateCellSubstate` updates the value of a given substate for the central cell. For example, the following portion of code sets the new value for the substate `THICK` for the central cell to the value of the `vent_thick` parameter (i.e. the amount of lava that is discharged by the crater at each CA step).

```
updateCellSubstate(THICK,
    getParameter(vent_thick));
```

Besides the setting of the value of lava thickness for the crater, the transition function subsequently calls alternatingly the `calc_flow` function and the `calc_thickness`, `calc_temperature` and `calc_quote` functions (here all omitted for brevity), depending on the fact the CA step, whose value is given by the internal variable `STEP`, is odd or even. All four functions correspond to the elementary processes σ_1, σ_2, σ_3 and σ_4, respectively (cf. Section 2.1). In this way, a single SCIARA step corresponds to 2 executions of the *libAuToti* transition function (double-step transition function).
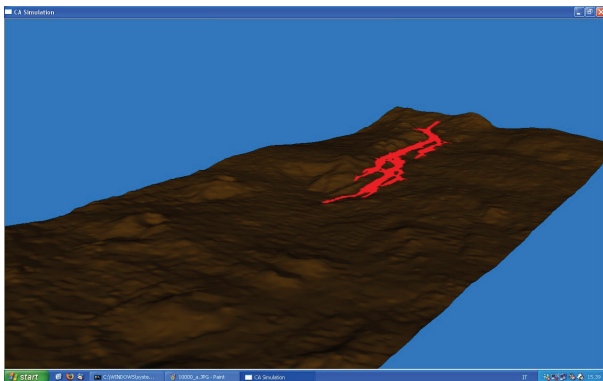
Figure 3: The SCIARA MCA application to the simulation of the 2001 Mt Etna event at Nicolosi. The figure, shows a snapshot of the CA configuration of the system after 18000 steps with the lava path in red, utilizing the libAuToti OpenScenegraph 3D viewer.

As stated above, though an integrated environment surely ensures many advantages, a benefit of using a separate library is the more flexibility which, for instance, allows the developer to easily extend the software features with the aid of further libraries. This is the case of the SCIARA and SCIDDICA implementations, where the computational model was written by utilizing the *libAuToti* library and two 3D visualization modules developed with OpenGL and with the OpenScenegraph visual toolkit. Figure 3 shows a snapshot of the SCIARA model simulation on the

Nicolosi 2001 lava event at the end of the simulation with the latter viewer.

At the same way, it would be possible add a further module, e.g. for model optimization purposes, by using the PGAPack (PGApack web reference 1998) free library for Parallel Genetic Algorithms (PGAs) (cf. also Iovine et al. 2005 for an application of PGAs for optimizing parameters of a MCA model).

3.2. A libAuToti implementation of SCIDDICA

In this paragraph, a seamless parallel implementation of SCIDDICA is presented. In particular, Figure 4 and 5 illustrate the CA definition and transition function, respectively. As in the SCIARA libAuToti implementation, the program includes the `libautoti.h` header that allows to employ the library. In particular, the object `sciddica` of type `ParallelSimulationCA_2D` is defined, representing a MCA with the possibility to perform a parallel simulation.

```
// header files here...
#include "libAuToti.h"

int main(int argc, char** argv)
{
    ParallelSimulationCA_2D sciddica;
    sciddica.setDim(410,296);

    //flow substates
    sciddica.addSubstate(SUBSTATE_FLOAT);// 0
    sciddica.addSubstate(SUBSTATE_FLOAT);// 1
    sciddica.addSubstate(SUBSTATE_FLOAT);// 2
    sciddica.addSubstate(SUBSTATE_FLOAT);// 3
    //elevation substate
    sciddica.addSubstate(SUBSTATE_FLOAT );//4
    //width substate
    sciddica.addSubstate(SUBSTATE_FLOAT); //5

    sciddica.setRadius( 1 );

    // von Neumann neighbourhood
    sciddica.addNeighboor( 0,-1 );
    sciddica.addNeighboor( 1, 0 );
    sciddica.addNeighboor(-1, 0 );
    sciddica.addNeighboor( 0, 1 );

    sciddica.addParameter(0.5); //adherence
    sciddica.addParameter(900); //crit. elevat.
    sciddica.addParameter(0.3); //fric. param.
    sciddica.addParameter(0.6); //relaxation

    sciddica.initialize(argc, argv);
    sciddica.readFromFile( "./tessina.txt", 4);
    sciddica.redFromFile( "./source.txt", 5);

    sciddica.setNumStep( 10000 );

    sciddica.run();

    sciddica.finalizeSimulation();
    return 0;
}
```

Figure 4: The SCIDDICA CA definition by using the libAuToti free MCA library.

The member function `setDim` defines the cell space dimensions, while `setRadius` and `addNeighboor` the CA the neighbourhood radius and relation, respectively. Moreover, as required by SCIDDICA, six substates are defined thanks to the `addSubstate` member function.

As requested by the considered version of SCIDDICA, the von Neumann neighborhood was defined, thanks to the `addNeighboor` member function.

As in the SCIARA libAuToti implementation, parameters are defined by means of the `addParameter` member function.

Moreover, input data is automatically performed by the function `readFromFile`, which reads files in standard ASCII Grid format. Eventually, the simulation is performed (for 10000 steps) and results saved (even in ASCII Grid format) by means of the function `run()`.

Figure 5 shows the libAuToti implementation of the SCIDDICA transition function, defined as an inline member function. Substates identifiers, together with the neighbours ones, are utilised as arguments of `getSubstate*`, which returns the value of a certain substate for the specified neighbouring cell. Here the symbol `*` can be one of the types listed in Table 1.

```
inline void parSimulationAC2D::transitionFunction()
{
if(getSubstateFloat(ELEV)>getParameter(DETACHMENT))
    setParameter(FRICTION,0.3);
else
    setParameter(FRICTION,0.5);

if (STEP == 0)
    if (getSubstateFloat(WIDTH) > (float)0)
            updateCellSubstate(ELEV,
                              (getSubstateFloat(ELEV)-
                               getSubstateFloat(WIDTH))
                              );
switch (STEP%2){
    case 1:  calc_flows(this);
                break;
    case 0:  calc_width(this);
                break;
    }
}
```

Figure 5: The SCIDDICA transition function definition by using the libAuToti free MCA library.

For instance, the following portion of code verifies if the value of the substate `WIDTH` of type `float`, is greater than 0 for the central cell:

```
if ( getSubstateFloat( WIDTH ) > (float)0 )
```

Analogously, if one wants to check the value of a neighbouring cell, she/he has to simply add the neighbour index as second argument to the `getSubstateFloat` function. The following example tests if the value of the substate `WIDTH` of the neighbour number 2 is equal to 0:

```
if (getSubstateFloat( WIDTH, 2 ) == (float)0 )
```

Equivalently, the function `updateCellSubstate` updates the value of a given substate for the central cell. For example, the following portion of code sets the new value for the substate WIDTH for the central cell to the value `new_width`.

```
updateCellSubstate(WIDTH, new_width );
```

Besides the setting of the p_f parameter (cf. FRICTION in Figure 5), the transition function performs a preliminary elaboration at step 0 in order to properly handle input data. Subsequently, it calls one of the two functions `calc_flow` or `calc_width` (here omitted for brevity), depending on the fact the CA step, whose value is given by the internal variable `STEP`, is odd or even. They correspond to the elementary processes σ_1 and σ_2, respectively (cf. Section 2.2). In this way, a single SCIDDICA step corresponds to 2 execution of the libAuToti transition function (double-step transition function). Accordingly, the SCIDDICA simulation steps were indeed exactly 5000.
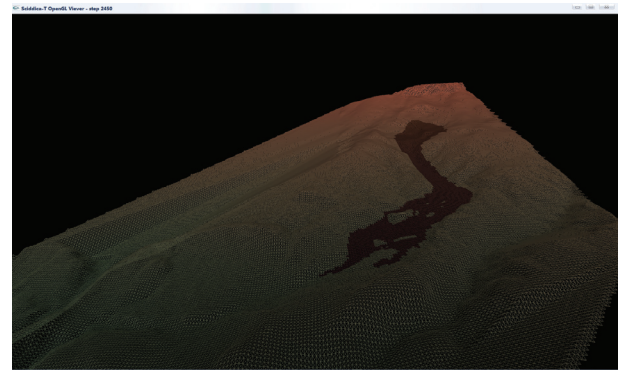

Figure 6: The SCIDDICA MCA application to the simulation of the Tessina Landslide. The figure shows a snapshot of the CA configuration of the system after 5000 steps utilizing the libAuToti OpenGL 3D viewer

3.3. Computational results

As concerns computational results, the SCIARA and SCIDDICA models – as implemented by adopting the *libAuToti* library – were tested on two parallel computers: a 24 processor HP Alpha Server SC45 and an Intel T2500 2Ghz dual-core PC.
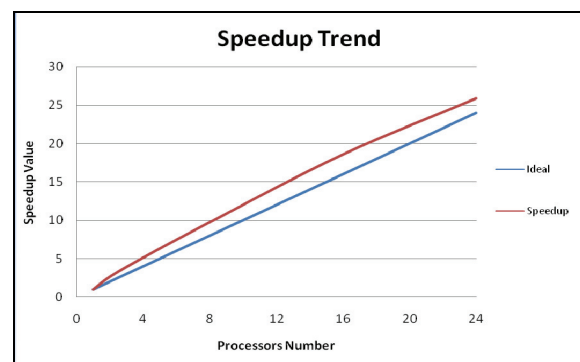

Figure 7: Speed-up of *libAuToti* for the model SCIARA carried out on a 24 processor Alpha Server SC45. Superlinear speedup is evident up to 24 processors.
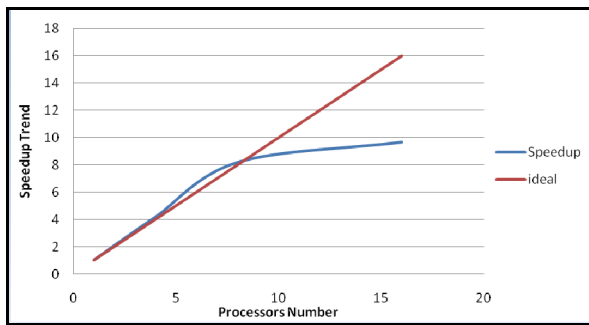
Figure 8: Speed-up of libAuToti for the MCA model SCIDDICA. Results refer to experiments carried out on 16 processors of the Alpha Server SC45. Superlinear speedup is evident up to 8 processors.

In particular, the high performing computer is composed of 6 SMP nodes, each with 4 Alpha 64-bit 1250 Mhz EV68 processors and 16 MB cache. Figure 7 shows the speed-up diagram for experiments carried out on the high performing machine, referred to the SCIARA model. In this case, super-linear speed-up was achieved up to 24 processors, probably due to the limited dimensions of the CA space and cache memory effects. Figure 8 depicts the speed-up diagram, referred to the SCIDDICA model experiments performed adopting 16 processors. Even in this case results, though slightly worse than the previous experiments, show the good scalability of the parallel system, together with super-linear effects up to 8 processors. Eventually, experiments carried out on the low cost machine – the Intel T2500 2Ghz dual-core PC, here omitted for briefness – have been positive. In particular, with respect with the first case, results have shown an even better super-linear effect when executed on both cores. Still, this is probably due both to the limited dimensions of the CA space and for the greater cache-effect (i.e. cache is physically closer to the CPU in multi-core machines) of the dual core architecture with respect to the supercomputer.

However, in all cases, computational results evidenced the goodness of the library in exploiting the considered parallel systems, confirming that even low-cost off-the-shelf machines can be fruitfully employed.

4. CONCLUSIONS

We have presented examples of implementation of the *libAuToti* C++ parallel library for Scientific Computing of two models based on the Macroscopic Cellular Automata (MCA) approach. In particular, the two considered models regard SCIARA, for lava flow simulation, and SCIDDICA, for debris flow modeling. The library, particularly suitable for MCA modeling, is easily extensible and easily combinable with different libraries, such as optimization or visualization modules.

Moreover, computational performances, achieved on a 24 processor Alpha Server SC45 supercomputer and on an Intel dual-core processor based machine have demonstrated the goodness of the library. Future developments of the library will regard the

implementation of a load balance strategy (e.g. based on the scattered decomposition of the computational domain), further library core optimization and the integration of an interactive console (e.g. similar to that of GNUPlot or Octave) to better control the phases of the simulation. Eventually, a full integration with interactive OpenGL or OpenScenegraph scientific 3D viewers will be tackled in the future. The library is freely downloadable from its official website, at the URL http://autoti.mat.unical.it. The library's name, *libAuToti*, is gaily taken from Prof. Salvatore "Toti" Di Gregorio, inventor of the MCA approach.

REFERENCES
Avolio, M.V., Di Gregorio, S., Mantovani, F., Pasuto, A., Rongo, R., Silvano, S., Spataro, W., 2000. Simulation of the 1992 Tessina landslide by a cellular automata model and future hazard scenarios, *International Journal of Applied Earth Observation and Geoinformation* 2:41-50.

Barca, D., Crisci, G.M., Di Gregorio, S., Nicoletta, F., 1994. Cellular Automata for simulating lava Flows: A method and examples of the Etnean eruptions. *Transport Theory and Statistical Physics* 23:195-232.

Crisci, G.M., Rongo, R., Di Gregorio, S., Spataro, W., 2004. The simulation model SCIARA: the 1991 and 2001 lava flows at Mount Etna. *Journal of Volcanology and Geothermal Research*, 132:253-267.

D'Ambrosio D., Di Gregorio, S., Iovine, G., Lupiano, V., Rongo, R., Spataro, W., 2003. First simulations of the Sarno debris flows through Cellular Automata modeling. *Geomorphology*, 54:91-117.

D'Ambrosio, D., Rongo, R., Spataro, W., Avolio, M.V., Lupiano, V., 2006. Lava Invasion Susceptibility Hazard Mapping Through Cellular Automata. *Proceedings of ACRI 2006*, pp. 452-461. September 20-23, Perpignan, France.

D'Ambrosio, D.; W. Spataro; and G. Iovine. 2006. Parallel genetic algorithms for optimising cellular automata models of natural complex phenomena: an application to debris-flows. *Computers and Geosciences-UK*, 32, 861-875.

Di Gregorio, S., Festa, D.C., Rongo, R., Spataro, W., Spezzano, G., Talia, D., 1996. A microscopic freeway traffic simulator on a highly parallel system. In: D'Hollander, E.H., Joubert, G.R., Peters D. Trystam, F.J., eds. *Parallel Computing: State-of-the-Art and Perspectives*, Amsterdam:Springer, 69-76.

Di Gregorio, S., Mele, F., Minei, G., 2001. Automi Cellulari Cognitivi. Simulazione di evacuazione, Proceedings of Input - Conferenza Nazionale

Informatica Pianificazione Urbana e Territoriale, Democrazia e Tecnologia, cdrom, June 23-26, Isole Tremiti, Italy.

Di Gregorio, S., Rongo, R., Siciliano, C., Sorriso-Valvo, M., Spataro, W, 1999. Mount Ontake landslide simulation by the cellular automata model SCIDDICA-3, *Physics and Chemistry of the Earth Part A* 24:97-100.

Di Gregorio, S., Serra, R., 1999. An empirical method for modelling and simulating some complex macroscopic phenomena by cellular automata. *Future Generation Computer Systems* 16:259-271.

Iovine, G., D'Ambrosio, D., Di Gregorio, S., 2005. Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects. *Geomorphology* 66:287-303.

McBirney, A.R., Murase, T., 1984. Rheological properties of magmas. *Annual Review of Earth Planetary Sciences*, 12:337-357.

PGAPack web reference: www-fp.mcs.anl.gov/CCST/ research/reports_pre1998/comp_bio/stalk/pgapack. html.

Spataro, W., D'Ambrosio, D., Spingola, S., Zito, G., Rongo, R., 2008. LibAuToti, A Parallel Cellular Automata Library for Simulation: An example of Application to Landslides. *Proceedings of Summer Computer Simulation Conference*, June 16-19, Edinburgh, UK).

Spezzano, G., Talia, D., 1998. Designing Parallel Models of Soil Contamination by the CARPET Language, *Future Generation Computer Systems* 13:291-302.

Succi, S., 2004. *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*. London, Oxford:University Press.

Trunfio, G.A., 2004. Predicting Wildfire Spreading Through a Hexagonal Cellular Automata Model, *Proceedings of ACRI 2004*, pp. 725-734. October 23-25, Amsterdam, The Netherlands.

Walter, R., Worsch, T., 2004. Efficient Simulation of CA with Few Activities. *Proceedings of ACRI 2004*, pp. 101-110. October 23-25, Amsterdam, The Netherlands.

AUTHORS BIOGRAPHY

WILLIAM SPATARO was born in London, UK. He is an Assistant Professor in Computer Science at the Department of Mathematics at the University of Calabria (Italy). His principal interests regard Parallel Computing, Cellular Automata, Modelling and Simulation of complex phenomena and Genetic Algorithms. His e-mail address is: spataro@unical.it and his Web-page can be found at http://www.mat.unical.it/spataro.

DONATO D'AMBROSIO was born in Cosenza, Italy. He is an Assistant Professor in Computer Science at the Department of Mathematics at the University of Calabria (Italy). His interests concern Genetic Algorithms, Parallel Computing, Simulation of Complex Natural Phenomena and Computer Graphics. His e-mail address is: d.dambrosio@unical.it and his Web-page can be found at http://www.mat.unical.it/~donato.

ROCCO RONGO was born in Naples, Italy. He is an Assistant Professor in Computer Science at the Department of Earth Sciences at the University of Calabria (Italy). His major interests regard Parallel Computing, Modelling and Simulation of Geological processes by Cellular Automata and GIS software. His e-mail address is: rongo@unical.it and his Web-page can be found at http://dister.unical.it/rongo/.

GIUSEPPE SPINGOLA was born in Praia a Mare, in Calabria (Italy). His Computer Science Degree thesis regarded the first release of the libAutoti library. He presently is working with the Department of Mathematics and the Department of Chemical Engineering of the University of Calabria for code parallelization of simulation models and on further releases of the libAuToti library. His email address is g.spingola@gmail.com.

GIUSEPPE ZITO was born in Torino, Italy. His Computer Science Degree thesis regarded the first release of the libAutoti library. He currently is collaborating with the Department of Mathematics and the Department of Chemical Engineering of the University of Calabria for code parallelization of simulation models on further releases of the libAuToti library. His email address is giuseppe.zito@gmail.com.

MARIA VITTORIA AVOLIO was born in Cosenza, Italy. Mathematician and Researcher Assistant at the Department of Mathematics and at the University of Calabria, Italy, she is involved basically in Modelling and Simulation of lava, landslide and pyroclastic flows by Cellular Automata methods. She is the beloved wife of William Spataro, who fell in love with her during her PhD course in Computer Science and Mathematics at the same University. Her e-mail address is: avoliomv@unical.it.

SALVATORE DI GREGORIO was born in Castellamare del Golfo, Sicily, Italy. He is Full Professor in Computer Science at the Department of Mathematics at the University of Calabria (Italy). Together with Professor Gino Mirocle Crisci, he leads the Empedocles Research group at the same University. His research interests concern Computational Models of Complex Systems and Parallel Computing. His e-mail address is: dig@unical.it.

AUTHOR'S INDEX