The 16th International Conference On Modeling And Applied Simulation

September 18 – 20, 2017
Barcelona, Spain

MAS 2017

EDITED BY
Agostino G. Bruzzone
Fabio De Felice
Claudia Frydman
Francesco Longo
Marina Massei
Adriano Solis

Printed in Rende (CS), Italy, September 2017

THE 16TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION, MAS 2017

SEPTEMBER 18 – 20, 2017
BARCELONA, SPAIN

ORGANIZED BY

DIME – UNIVERSITY OF GENOA

LIOPHANT SIMULATION

SIMULATION TEAM

IMCS – INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF SIMULATION

DIMEG, UNIVERSITY OF CALABRIA

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

AUTONOMOUS UNIVERSITY OF BARCELONA

MODELING AND SIMULATION CENTER OF EXCELLENCE (MSCOE)

LATVIAN SIMULATION CENTER – RIGA TECHNICAL UNIVERSITY

LOGISIM

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

MIMOS – MOVIMENTO ITALIANO MODELLAZIONE E SIMULAZIONE

MITIM PERUGIA CENTER – UNIVERSITY OF PERUGIA

BRASILIAN SIMULATION CENTER, LAMCE-COPPE-UFRJ

MITIM - McLeod Institute of Technology and Interoperable Modeling and Simulation – Genoa Center

M&SNet - McLeod Modeling and Simulation Network

Latvian Simulation Society

Ecole Superieure d'Ingenierie en Sciences Appliquees

Facultad de Ciencias Exactas. Ingegneria y Agrimensura

University of La Laguna

CIFASIS: CONICET-UNR-UPCAM

INSTICC - Institute for Systems and Technologies of Information, Control and Communication

National Russian Simulation Society

CEA - IFAC

University of Bordeaux

University of Cyprus

Dutch Benelux Simulation Society

I3M 2017 Industrial Sponsors

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2017 Media Partners

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SIMULATION AND PROCESS MODELING

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF OIL, GAS AND COAL TECHNOLOGY

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SERVICE AND COMPUTING ORIENTED MANUFACTURING

IGI GLOBAL – INTERNATIONAL JOURNAL OF PRIVACY AND HEALTH INFORMATION MANAGEMENT

HALLDALE MEDIA GROUP: MILITARY SIMULATION AND TRAINING MAGAZINE

HALLDALE MEDIA GROUP: THE JOURNAL FOR HEALTHCARE EDUCATION, SIMULATION AND TRAINING

SAGE
SIMULATION TRANSACTION OF SCS

DE GRUYTER
INTERNATIONAL JOURNAL OF FOOD ENGINEERING

EDITORS

AGOSTINO BRUZZONE
MITIM-DIME, UNIVERSITY OF GENOA, ITALY
agostino@itim.unige.it

FABIO DE FELICE
UNIVERSITY OF CASSINO, ITALY
defelice@unicas.it

CLAUDIA FRYDMAN
LSIS, FRANCE
claudia.frydman@lsis.org

FRANCESCO LONGO
DIMEG, UNIVERSITY OF CALABRIA, ITALY
f.longo@unical.it

MARINA MASSEI
DIME, UNIVERSITY OF GENOA, ITALY
massei@itim.unige.it

ADRIANO SOLIS
YORK UNIVERSITY, CANADA
asolis@yorku.ca

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2017

GENERAL CO-CHAIRS

AGOSTINO BRUZZONE, *MITIM-DIME, UNIVERSITY OF GENOA, ITALY*
MIQUEL ANGEL PIERA, *AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN*

PROGRAM CHAIRS

FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
YURI MERKURYEV, *RIGA TECHNICAL UNIVERSITY, LATVIA*

THE 16TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION, MAS 2017

GENERAL CO-CHAIRS

ADRIANO SOLIS, YORK UNIVERSITY, CANADA
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*

PROGRAM CO-CHAIRS

FABIO DE FELICE, *UNIVERSITY OF CASSINO, ITALY*
CLAUDIA FRYDMAN, *LSIS, FRANCE*

MAS 2017 International Program Committee

Naamane Aziz, Laboratoire des Sciences de l'Information et des Systèmes, France
Youssef Bouanan, *Lab Gretha Université Bordeaux*, France
Christos Bouras, University of Patras and CTI&P Diophantus, Greece
Agostino Bruzzone, *University of Genoa, Italy*
Flavio S. Correa Da Silva, University of Sao Paulo, Brazil
Juan Luis Crespo Mariño, *Tecnológico de Costa Rica*
Alessandro Chiurco, *University of Calabria, Italy*
Gerson Cunha, *LAMCE COPPE UFRJ, Brazil*
Saber Darmoull, *King Saud University, Kingdom of Saudi Arabia*
Fabio De Felice, *University of Cassino, Italy*
David Del Rio Vilas, *University of La Coruna, Spain*
Sabeur Elkosantini, King Saud University, Kingdom of Saudi Arabia
Denis M Filatov, Institute of Physics of the Earth- Russian Academy of Sciences, Russian Federation
Ian Flood, University of Florida, USA
Claudia Frydman, *LSIS, France*
Mirsad Hadzikadic, University of North Carolina Charlotte, USA
Yilin Huang, *Delft University of Technology, Netherlands*
Jus Kocijan, *Jozef Stefan Institute, Slovenia*
Janos Sebestyen Janosy, *Centre For Energy Research Hungarian Academy Of Sciences, Hungary*
Pasquale Legato, *University of Calabria, Italy*
Francesco Longo, *MSC-LES, University of Calabria, Italy*
José M. Cecilia , *Universidad Católica San Antonio, Spain*
Marina Massei, *Liophant Simulation, Italy*
Rina Mary Mazza, *University of Calabria, Italy*
Nuno Melão, *Instituto Politécnico de Viseu, Portugal*
Letizia Nicoletti, *Cal-Tek Srl, Italy*
Paulo Moura Oliveira, *Universidade de Tras os Montes e Alto Douro, Portugal*
Antonio Padovano, *University of Calabria, Italy*
Alessandro Pellegrini, *La Sapienza University, Italy*
Diego Crespo Pereira, *University of A Coruna, Spain*
Horacio Emilio Pérez Sánchez, *Universidad de Murcia, Spain*
Antonella Petrillo, *University of Cassino, Italy*
Rosa Rios Prado, *University of A Coruna, Spain*
Francesco Quaglia, *La Sapienza University, Italy*
Maria Celia Santos Lopes, *LAMCE/COPPE/UFRJ – Brasil*
Jean-François Santucci, *University of Corsica, France*
Peer Olaf Siebers, *University of Nottingham, UK Georgios Ch*
Adriano Solis, *York University, Canada*
Gregory Zacharewicz, *IMS Université Bordeaux 1, France*
Frank Werner, Otto-von-Guericke-Universität Magdeburg, Germany
Kuan Yew Wong, Universiti Teknologi Malaysia, Malaysia

Tracks and Workshop Chairs

Inventory Management Simulation
Chairs: Adriano Solis, York University, Canada; Letizia Nicoletti, Cal-Tek Srl, Italy

Production Systems Design
Chair: David del Rio Vilas, Ferrovial Services, Spain

Decision Support Systems Applications
Chairs: Fabio De Felice, Univerisity of Cassino, Italy; Antonella Petrillo, Univerisity of Cassino, Italy

Simulation in Energy Generation and Distribution
Chair: Janos Sebestyen Janosy, Centre for Energy Research Hungarian Academy of Sciences (HAS), Hungary

The Interplay between Behavioural Game Theory and Agent-Based Social Simulation
Chairs: Peer-Olaf Siebers, Nottingham University, UK; Prof Youssef Bouanan, Lab Gretha Université Bordeaux, France

Control and monitoring of Hybrid Renewable Energy Sources and Systems for buildings
Chair: Naamane Aziz, Laboratoire des Sciences de l'Information et des Systèmes, France

Simulation Based Design
Chair: Yilin Huang, Delft University of Technology, Netherlands

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

Workshop on virtual and augmented reality
Chairs: Gerson Cunha, LAMCE/COPPE/UFRJ, Brasil; Maria Celia Santos Lopes, LAMCE/COPPE/UFRJ, Brasil

Simulation and Human Factors Engineering
Chair: Diego Crespo Pereira, University of La Coruna, Spain

Simulation based Optimization
Chairs: Pasquale Legato, University of Calabria, Italy; Rina Mary Mazza, University of Calabria, Italy

High Performance Simulation of biological systems
Chairs: Horacio Pérez-Sánchez, Universidad Católica San Antonio, Spain; José M. Cecilia Universidad Católica San Antonio, Spain

Workshop on Serious Games applied to Security, Crisis Management and Safety
Chairs: Francesco Longo University of Calabria, Italy; Agostino Bruzzone, University of Genoa, Italy

Human-centred and Human-Focused Modelling and Simulation (common track MAS-EMSS)
Chairs: Agostino Bruzzone, DIME, University of Genoa, Italy; Peer Olaf Siebers, University of Nottingham, UK

Towards a universal framework for the Intelligent Social Simulation-based Systems (ISSS)
Chair: Antonio Padovano, University of Calabria, Italy

CHAIRS' MESSAGE

WELCOME TO MAS 2017!

The International Conference on Modeling and Applied Simulation (MAS), started in 2002 and now on its 16th year, is a forum for scholars, scientists, and practitioners to share the latest innovations and advances in applied modeling & simulation (M&S) and associated computer technologies. As well, MAS aims to provide networking opportunities for establishing international collaboration on M&S application projects. The crop of papers submitted and accepted for MAS 2017 comes from a significant number of countries all over the world – Vietnam, Japan, Republic of Korea, United Arab Emirates, Iran, Canada, U.S.A., Brazil, Colombia, Tunisia, Madagascar, France, Germany, Portugal, Spain, Estonia, Czech Republic, the United Kingdom, and Italy.

As we all know, M&S has been applied in many aspects of human endeavor. Various areas of M&S application are covered by a number of the thematic component conferences that have, over the years, become part of the International Multidisciplinary Modeling & Simulation Multiconference (I3M) – harbor, maritime and multimodal logistics (HMS), control and automation (IMAACA), energy, sustainable development and environment (SESDE), defense and homeland security (DHSS), healthcare (IWISH), and food operations and processing (FoodOPS). The MAS 2017 papers span a fairly wide range of M&S application areas – from production, warehousing, inventory, and logistics systems design and improvement, production and operations scheduling, and supply chain quality management to aerospace exploration, performance of biped robots and hybrid vehicles, and augmented and virtual reality training systems, among others.

A successful program requires the dedicated effort of many people: we sincerely thank all the authors who have contributed papers to MAS this year; we are likewise extremely grateful to the MAS 2017 International Program Committee members, particularly the Track Chairs for organizing their respective tracks and encouraging the submission of papers and the reviewers for spending time and effort to read, evaluate, and provide suggestions for improving the quality of the final papers. Special thanks go to faculty and staff of Universitat Autònoma de Barcelona for providing facilities. Last but not least, we thank the Keynote Speakers for their invaluable contribution.

We warmly welcome all conference participants to this 16th staging of MAS in the enchanting and culturally rich city of Barcelona.

Marina Massei
University of Genoa,
Italy

Adriano Solis
York University
Canada

Acknowledgements

I3M 2017 Internal Staff

Matteo Agresta, *Simulation Team, Italy*

Margarita Bagamanova, *Autonomous University of Barcelona, Spain*

Luigi Bruno, *DIMEG, University of Calabria, Italy*

Agostino G. Bruzzone, *DIME, University of Genoa, Italy*

Alessandro Chiurco, *DIMEG, University of Calabria, Italy*

Riccardo Di Matteo, *Simulation Team, Italy*

Caterina Fusto, *CAL-TEK Srl, Italy*

Thimjo Koça, *Autonomous University of Barcelona, Spain*

Francesco Longo, *DIMEG, University of Calabria, Italy*

Gianluca Maglione, *Simulation Team, Italy*

Marina Massei, *DIME, University of Genoa, Italy*

Romualdo Moreno, *Autonomous University of Barcelona, Spain*

Mercedes Narciso, *Autonomous University of Barcelona, Spain*

Letizia Nicoletti, *CAL-TEK Srl, Italy*

Antonio Padovano, *DIMEG, University of Calabria, Italy*

Miquel Angel Piera, Autonomous University of Barcelona, Spain

Marko Radanovic, *Autonomous University of Barcelona, Spain*

Juan Jose Ramos, *Autonomous University of Barcelona, Spain*

Cataldo Russo, *CAL-TEK Srl, Italy*

Nina Schefers, *Autonomous University of Barcelona, Spain*

Marco Vetrano, *CAL-TEK Srl, Italy*

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

INDEX

AHP-K-GDSS: A NEW SORTING METHOD BASED ON AHP FOR GROUP DECISIONS

Alessio Ishizaka[a] Francesco Lolli[b], Rita Gamberini[c], Bianca Rimini[d], Elia Balugani[e]

[b],[c],[d],[e] Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Via Amendola 2, Padiglione Morselli, 42122 Reggio Emilia, Italy
[a] Centre for Operational Research and Logistics, Portsmouth Business School, University of Portsmouth, United Kingdom

[a]Alessio.Ishizaka@port.ac.uk, [b]francesco.lolli@unimore.it, [c]rita.gamberini@unimore.it
[d]bianca.rimini@unimore.it [e]elia.balugani@unimore.it

ABSTRACT

Some public buildings need for energy requalification intervention as they are responsible for a significant share of energy consumption and other related CO_2 emissions. With tight budget constraints choices have to be made.

To solve this problem a group sorting decision support system based on the analytic hierarchy process, the K-means algorithm has been developed. The system aims at sorting alternatives into ordered classes of importance. A case study carried out in an Italian municipality allowed us to verify the validity of our new method in a real setting.

Keywords: Energy requalification; decision support system; AHP; clustering; sorting

1. INTRODUCTION

The Analytic Hierarchy Process (AHP) is a multi-criteria decision-making method developed by Saaty in the 1970s (Saaty 1977). It has been widely used for ranking a finite set of alternatives and for choosing the best alternative from a finite set of alternatives (Ishizaka and Labib 2014). In the paper (Ishizaka, Nemery and al. 2012), AHP was adapted in AHPSort in order to deal also with sorting problems.

A sorting problem aims to assign each alternative into one of the predefined ordered classes (Ishizaka and Nemery 2013b). In the case of problems with a large set of alternatives, AHPSort enables us to avoid the construction of a pairwise comparison matrix including all the alternatives. The alternatives are not compared with each other but only with the profiles representing the classes. Thus, the pairwise comparison matrix is much smaller. In the case of problems where the set of alternatives could change (by either adding or removing an alternative), using AHPSort can avoid modifying the pairwise comparison matrix of the alternatives and recalculating the priorities.

When an alternative is removed, its attached pairwise comparison matrix is also removed but the other pairwise comparison matrices are untouched. When an alternative is added, a new pairwise comparison matrix is added and only the pairwise comparisons of the alternative with the profiles representing the classes need to be provided.

However, sometimes the decision-maker is unable to provide the reference profiles. In this case, Lolli, Ishizaka et al. (2014) have developed Analytic Hierarchy Process-K (AHP-K). It is a hybrid method based on AHP for the evaluation of the weights and the K-means for the sorting of alternatives into K-ordered classes. This paper adapts this multi-criteria decision sorting method used by single decision-maker to group decisions. The aim of the resulting Analytic Hierarchy Process-K-Group Decision Support System (AHP-K-GDSS) is to incorporate all the actors' opinions, to mitigate their subjectivity and to evaluate objective and subjective criteria in one model. This new methodology was developed to solve a real case study of energy requalification of pebble buildings and utilities. The rest of the paper is presented as follow. Section 2 contains the literature review. Section 3 presents the new methodology. Section 4 describes the case study and section 5 concludes the paper.

2. LITERATURE REVIEW

Sorting problems came to the attention of researchers and practitioners later than ranking problems. AHP (Saaty 1977), PROMETHEE (Brans and Vincke 1985), MACBETH (Bana e Costa, De Corte et al. 2012), TOPSIS (Lai, Liu et al. 1994) and ELECTRE III (Roy 1978) are some popular multi-criteria approaches that have been extended to sorting problems, respectively leading to AHP-Sort (Ishizaka, Nemery et al. 2012), AHP-Sort II (Miccoli and Ishizaka 2017), GAHPSort (López and Ishizaka 2017), FlowSort (Nemery and Lamboray 2008), GAIA-Sort (Nemery, Ishizaka et al. 2012), MACBETHSort (Choudhary and Shankar 2012), TOPSIS-Sort (Sabokbar, Hosseini et al. 2016), ELECTRE-Tri (Yu 1992) and its variant ELECTRESort (Ishizaka and Nemery 2014), ELECTRE Tri-C (Almeida-Dias, Figueira et al. 2010)(Almeida-Dias, Figueira et al. 2010), ELECTRE Tri-nC (Almeida-Dias, Figueira et al. 2012), ELECTRE Tri-nB (Fernández, Figueira et al. 2017). Most sorting approaches assume a single DM or a group acting as one (Ka 2011), and the

goal of finding an agreed classification method is therefore neglected in the case of multiple DMs. However, most decisions are taken by several decision-makers (Ishizaka and Nemery 2013). Multiple criteria sorting in the context of group decision-making is a challenging field of research (Gothwal and Saha 2015), with applications to several operational settings.

Two recent multi-criteria decision-making approaches address group sorting problems. The first one was proposed by Wang and Chen (2006), based on intuitive fuzzy outranking relations among the alternatives. The second one (Lolli, Ishizaka et al. 2015) represents the extension of FlowSort (Nemery and Lamboray 2008) to group decisions. This paper present an extension to group decision for AHPSort-K.

3. METHODOLOGY

The developed Group Decision Support System (GDSS) consists of two stages. The first stage (Lolli, Ishizaka et al. 2016) involves sorting alternatives into ordered classes (Lolli, Ishizaka et al. 2016). The second stage involves selecting the alternatives to be funded under the constraints of a limited budget.

3.1. Sorting stage with AHP-K-GDSS

Step 1: Pair-wise comparison between criteria
The I criteria are pair-wise compared in order to calculate their relative weights w_i^k with the eingevalue method (Ishizaka 2014), where $i = 1, \dots, I$, is the criterion and $k = 1, \dots, K$ is the decision-maker. The K priority vectors are thus obtained as the eigenvectors associated with the highest eigenvalues for each comparison matrix. Consistency analysis must be performed (Saaty 1980) to ensure the judgmental consistency of the matrices, that is to say compliance with the transitivity rule in terms of a consistency ratio lower than a typical threshold value of 0.1.

Step 2: Score assignment
Each alternative is evaluated directly (e.g. assignment of a score on a scale 1-20) or indirectly (e.g. with pairwise comparison matrices) on each criterion.

Step 3: Individual Ranking
The universal priority of alternative a_j for decision-maker k is thus obtained by the weighted sum of $v_{i,j}^k$:

$$P_j^k = \sum_{i=1}^{I} w_i^k \times v_{i,j}^k \qquad (1)$$
where $j = 1, \dots, J$ and $k = 1, \dots, K$

Let $\overline{P_j^k}$ be the ordered array of P_j^k, i.e. the individual ranking for decision-maker j. Hence, different K rankings are now available.

Step 4: Universal ranking
The weights of the decision-makers K (i.e. u^k) are assigned by the owner of the decision-making process such that $\sum_{k=1}^{K} u^k = 1$ in accordance with their

experiences, skills, etc. The universal ranking is obtained by a further weighted sum of P_j^k (1) as follows:

$$P_j = \sum_{k=1}^{K} u^k \times P_j^k \qquad \text{with } j = 1, \dots, J \qquad (2)$$
where $\overline{P_J}$ is the ordered vector of P_j, i.e. the universal ranking.

Step 5: Universal Sorting
In order to sort the alternatives into N classes, the K-Means algorithm is applied in the universal ranking of Step 4.

The aim is to create compact and well-separated N classes of alternatives for $\overline{P_J}$ by minimising the sum of the squared distances between the centroid of each class (i.e. the mean point in the case of Euclidean distance) and the items in the class. This step allows all the alternatives to be classified into classes with different degrees of priority.

Step 6: Individual Sorting for the Veto application
This optional further step aims to compare the universal sorting with the individual sorting. In order to find the individual sorting, the K-Means algorithm is now applied to $\overline{P_j^k}$, where $k = 1, \dots, K$.

An alternative a_j may be classified into a universal class C_n, very distant from the individual class C_n^k of a particular DM. This divergent opinion can now be taken into account with a veto. That is to say, the alternative is downgraded or upgraded in order to classify it into a universal class closer to the individual one. In sum, the veto applied to the DMs avoids the full compensation of the weighted sum (Step 4) by opportunely modifying the final sort on the basis of the individual sort.

4. CASE STUDY
4.1. Introduction
The proposed decision support system was validated in an Italian municipality of about 30,000 inhabitants. Energy is becoming a precious resource (Ishizaka, Siraj et al. 2016). The aim of the study was to choose between 34 interventions of energy requalification $j = 1, \dots, 34$

Three council members served as decision-makers, with the mayor leading the decision process. In particular, DM1 is a budget representative, DM2 is responsible for social policies and DM3 is an environmental expert.

The mayor assigned weights to the DMs on the basis of their experience and skills. In this case, the weights were set equal, i.e. $u^1 = u^2 = u^3 = 0.33$. The decision-makers were asked to directly assign scores on a scale of 1-20 to the alternatives as regards to the criteria

4.2. Criteria
A brainstorming performed by the decision-makers gathered the following criteria: Annual CO2 savings (c1), Annual monetary savings (c2), Financial payback time (c3), Comfort improvement (c4), Image toward citizens (c5), Educational value (c6), Local employment development (c7), Increase in energy self-sufficiency

(c8). Each DM pairwise compare them and the derived weights are given in the table 1.

Table 1: Weights of the criteria

	DM1	DM2	DM3
c1	0.071	0.054	0.459
c2	0.231	0.023	0.12
c3	0.331	0.018	0.082
c4	0.048	0.153	0.056
c5	0.024	0.385	0.02
c6	0.033	0.223	0.028
c7	0.157	0.105	0.038
c8	0.106	0.038	0.197
CR	0.03	0.09	0.06

4.3. Results

The individual rankings (Step 3) were achieved by multiplying the scores by the weights of the criteria (Table 1) for each DM. The global rankings were then obtained by aggregating these individual rankings through a weighted sum over DMs (Step 4), with DMs equally weighted in this case.

The K-means algorithm was launched on the global ranking in order to achieve the global sorting into three ordered clusters named C1, C2 and C3 from the most to least preferred (Step 5). As the K-means is a greedy algorithm, the final partition depends on both the number of iterations and the starting partition. In order to verify the robustness of the solution, the algorithm was launched 20 times with 100 iterations, each time starting from a different random partition. It has been found that the solution does not change over launches, and can therefore be considered robust. In order to apply the veto system, the global sorting approach has to be compared with the individual sorting, which was also obtained with the K-means algorithm (Step 5). Table 3 shows the global rankings of the alternatives, with global priorities reported in brackets, along with the global sorting. Furthermore, the last column the final classes after the veto application (Step 6), where the symbol "-" indicates that no change in sorting occurred. Since the number of classes and decision-makers is three in this case, the veto is simply expressed by the following two conditions: i) if an alternative is globally sorted as C1 (C3), but at least one DM classifies it as C3 (C1), then it is reclassified into C2; and ii) if an alternative is globally sorted as C2, but two DMs classify it as C3 (C1), then it is reclassified into C3 (C1). In our case, only the alternative 17 was reclassified into C2 for condition i).

Table 1: Sorting of the alternatives

Alternatives	Global ranking and priority	Global sorting	Veto
1	30 (7.73)	C3	-
2	34 (7.10)	C3	-
3	18 (8.71)	C3	-
4	24 (8.14)	C3	-
5	33 (7.48)	C3	-
6	25 (8.03)	C3	-
7	31 (7.70)	C3	-
8	26 (7.88)	C3	-
9	27 (7.78)	C3	-
10	21 (8.50)	C3	-
11	19 (8.54)	C3	-
12	20 (8.53)	C3	-
13	22 (8.49)	C3	-
14	9 (10.24)	C2	-
15	7 (10.31)	C2	-
16	11 (9.96)	C3	-
17	1 (12.32)	C1	C2
18	16 (8.94)	C3	-
19	29 (7.74)	C3	-
20	17 (8.91)	C3	-
21	12 (9.90)	C3	-
22	14 (9.45)	C3	-
23	10 (10.01)	C3	-
24	5 (11.36)	C2	-
25	28 (7.75)	C3	-
26	32 (7.69)	C3	-
27	3 (11.69)	C2	-
28	2 (12.10)	C1	-
29	4 (11.68)	C2	-
30	23 (8.41)	C3	-
31	8 (10.27)	C2	-
32	6 (10.60)	C2	-
33	13 (9.58)	C3	-
34	15 (9.35)	C3	-

The alternative in C1 will be funded. The alternatives in C2 will be funded depending on their global ranking and funds available. The alternatives in C3 will not be funded.

5. CONCLUSIONS

A new AHP-based group sorting method has been defined with the aim of classifying a set of alternatives into a predefined number of ordered classes, without recourse to limiting profiles defined by decision-makers. They were simply asked to assign scores to the alternatives, while the k-means clustering algorithm was used to classify the alternatives into well-separated clusters. This automatically solves a sorting problem when decision-makers are not confident in providing the

limiting profiles, and the generation of a predefined number of not-empty classes has to be forced. The individual sorting was also achieved with the aim of allowing the application of a veto system. This is the first AHP-based group sorting approach that has been developed.

The application of this group decision support system to a real case study has validated the approach and confirmed its usefulness in real settings, where transparent procedures are required in decision-making groups.

Topics for further research may include distance-based models for veto application, as well as fuzzy or evidence-based theories for dealing with the uncertainty of decision-makers' judgements. An index of compactness to measure the quality of the sorting could also be developed.

REFERENCES

Almeida-Dias J., Figueira J., Roy B., 2010. Electre Tri-C: A multiple criteria sorting method based on characteristic reference actions. European Journal of Operational Research, 204(3), 565-580.

Almeida-Dias J., Figueira J., Roy B., 2012. A multiple criteria sorting method where each category is characterized by several reference actions: The Electre Tri-nC method. European Journal of Operational Research, 217(3), 567-579.

Bana e Costa C., De Corte J.-M., Vansnick, J.-C., 2012. MACBETH. International Journal of Information Technology & Decision Making, 11(02), 359-387.

Brans, J.-P., Vincke P., 1985. A preference ranking organisation method. Management science, 31(6), 647-656.

Choudhary D., Shankar R., 2012. An STEEP-fuzzy AHP-TOPSIS framework for evaluation and selection of thermal power plant location: A case study from India. Energy, 42(1), 510-521.

Fernández E., Figueira J., Navarro J., Roy B., 2017. ELECTRE TRI-nB: A new multiple criteria ordinal classification method. European Journal of Operational Research 263(1), 214-224.

Gothwal S., Saha R., 2015. Plant location selection of a manufacturing industry using analytic hierarchy process approach. International Journal of Services and Operations Management, 22(2), 235–255.

Ishizaka A., 2014. Comparison of fuzzy logic, AHP, FAHP and hybrid fuzzy AHP for new supplier selection and its performance analysis. International Journal of Integrated Supply Management, 9(1), 1-22.

Ishizaka A., Labib A., 2014. A hybrid and integrated approach to evaluate and prevent disasters. Journal of the Operational Research Society, 65(10), 1475–1489.

Ishizaka A., Nemery P., 2013a. A Multi-Criteria Group Decision Framework for Partner Grouping when sharing Facilities. Group Decision and Negotiation, 22(4), 773-799.

Ishizaka A., Nemery P., 2013b. Multi-Criteria Decision Analysis. Chichester (United Kingdom), John Wiley & Sons Inc.

Ishizaka A., Nemery P., 2014. Assigning machines to incomparable maintenance strategies with ELECTRE-SORT. Omega, 47(0), 45-59.

Ishizaka A, Nemery P., Pearman C., 2012. AHPSort: an AHP based method for sorting problems. International Journal of Production Research, 50(17), 4767-4784.

Ishizaka A., Siraj S., Nemery P., 2016. Which energy mix for the UK? An evolutive descriptive mapping with the integrated GAIA-AHP visualisation tool. Energy, 95, 602–611.

Ka B., 2011. Application of Fuzzy AHP and ELECTRE to China dry port location selection. The Asian Journal of Shipping and Logistics, 27(2), 331-353.

Lai Y.-J., Liu T.-Y., Hwang C.-L. (1994). "TOPSIS for MODM." European Journal of Operational Research 76(3): 486-500.

Lolli F., Ishizaka A., Gamberini R., 2014. New AHP-based approaches for multi-criteria inventory classification. International Journal of Production Economics, 156(0), 62-74.

Lolli F., Ishizaka A., Gamberini R., Rimini B., 2016. Solving the group sorting problem with an AHP-based approach. Proceedings of OR58: The OR Society Annual Conference, pp. 96-106, Portsmouth (Hampshire, UK).

Lolli F., Ishizaka A., Gamberini R., Rimini B., Messori M. 2015. FlowSort-GDSS – A novel group multi-criteria decision support system for sorting problems with application to FMEA. Expert Systems with Applications, 42(17–18), 6342-6349.

López C., Ishizaka A., 2017. GAHPSort: A new group multi-criteria decision method for sorting a large number of the cloud-based ERP solutions. Computers in Industry, 92–93, 12-24.

Miccoli F., Ishizaka A., 2017. Sorting municipalities in Umbria according to the risk of wolf attacks with AHPSort II. Ecological Indicators, 73, 741-755.

Nemery P., Ishizaka A., Camargo A., Morel L., 2012. "Enriching descriptive information in ranking and sorting problems with visualizations techniques. Journal of Modelling in Management, 7(2), 130-147.

Nemery P., Lamboray C. 2008. FlowSort: a flow-based sorting method with limiting or central profiles. TOP, 16(1), 90-113.

Roy B., 1978. ELECTRE III: algorithme de classement base sur une présentation floue des préférences en présence de critères multiples. Cahiers du CERO, 20(1), 3-24.

Saaty T., 1977. A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15(3), 234-281.

Saaty T., 1980. The Analytic Hierarchy Process. New York, McGraw-Hill.

Sabokbar H., Hosseini A., Banaitis A., Banaitiene N., 2016. A novel sorting method topsis-sort: An

application for tehran environmental quality evaluation. Ekonomie a Management, 19(2), 87-104.

Wang E., Chen J., 2006. Effects of internal support and consultant quality on the consulting process and ERP system quality. Decision Support Systems, 42(2), 1029-1041.

Yu W., 1992. ELECTRE TRI: Aspects méthodologiques et manuel d'utilisation, Université Paris-Dauphine. 65- 80.

A JAVA LIBRARY FOR EASING THE DISTRIBUTED SIMULATION OF SPACE SYSTEMS

Alberto Falcone, Alfredo Garro

Department of Informatics, Modeling, Electronics, and Systems Engineering (DIMES), University of Calabria, via P. Bucci 41C, 87036, Rende (CS), Italy

{alberto.falcone, alfredo.garro}@dimes.unical.it

ABSTRACT
The space flight domain is one of the numerous fields that involve experts belonging to different scientific domains such as mathematical, physical, aerospace and software engineering. Many research efforts are focusing on the definition of methods, tools and software libraries, mainly aiming at providing a robust and flexible way for defining, building and simulating complex systems in space so as to understand, predict and optimize their behavior. In this context, the paper presents a space flight dynamics library, named *Java Space Dynamics Library (JSDL)*, which offers high fidelity models and algorithms to manage space systems according to the SISO Space Reference FOM standardization initiative.

Keywords: Modeling and Simulation, Space Flight Dynamics, Distributed Simulations, High Level Architecture (HLA)

1. INTRODUCTION
Due to the increasing complexity of space systems, and thus of the related engineering problems (Falcone, Garro, and Tundis 2014; Fortino et al. 2007; Garro et al. 2015; Garro and Falcone 2015), there is a consistent investment in the development of new methods, tools and software libraries able to provide a robust and flexible way for defining, building and simulating them (Falcone et al. 2016; Fortino et al. 2006; Ido 2012; Rogovchenko-Buffoni et al. 2014; San-Juan et al. 2011). These available, commercial and noncommercial, solutions support one or more of the phases in the development of space systems such as flight mechanics, propulsion, orbit controls and data analysis; however, none of them seems capable of providing complete coverage of the whole development process in a flexible way (Pulecchi and Lovera 2006).
In this context, there is an increasing need for efficient and flexible solutions capable of covering all the steps in the design and develop of space systems, especially for supporting system modeling and simulation where modularity, flexibility and reusability are key features to provide (Falcone et al. 2016; Falcone et al. 2015; Pulecchi and Lovera 2006).

To contribute to fill this lack, the paper presents the *Java Space Dynamics Library (JSDL)* project, emphasizing its flexibility and showing the set of services provided to define and build space systems such as satellites and spacecrafts. The rest of the paper is structured as follows: related works are discussed in Section 2; Section 3 presents the *Java Space Dynamics Library (JSDL)* whose architecture and provided services are discussed in Section 4 and 5 respectively. Finally, in Section 6 conclusions are drawn and future research directions are delineated.

2. RELATED WORK
There are several research efforts on the development of methods, tools and libraries in the astrodynamics field, mainly aiming at providing a robust and flexible way for defining, building and simulating complex systems in space. The most applicable solutions have been developed after the mid-1960's when space missions were the attention of media and computers become prevalent in academia and industry.
The *Java Astrodynamics Toolkit (JAT)* is an open source library of reusable components, distributed under the GNU General Public License (GLP). It is implemented in the Java language and helps developers to create their own application programs and solve problems in astrodynamics, mission design, spacecraft navigation, guidance and control. It provides functionalities that allow the rapid development of spacecraft simulations including 2D and 3D visualization capabilities. Possible applications of JAT include: (i) Design and analysis of space missions, including trajectory optimization; (ii) Simulation of spacecraft navigation, guidance and control as well as its visualization in a 3D environment; and, (iii) Simulation of the motion for basic rigid and flexible spacecraft dynamics (Gaylor, Page, and Bradley 2006).
Another software library that enables developers to effectively define and manage elements in space is *Orbits Extrapolation Kit (Orekit)* (CS Communication & Systémes 2017). *Orekit* is implemented in the Java language and aims at providing accurate and efficient low level standard astrodynamical models (e.g., time, frames, orbital parameters, orbit propagation, attitude

and celestial bodies) and algorithms (e.g., time conversions, propagations and pointing) for the development of flight dynamics applications. It is designed to be easily used in very different contexts, from quick studies up to critical operations. It was developed in 2002 at CS Systémes d'Information and was officially released as an open source software, under the Apache License Version 2.0, in 2008 (CS Communication & Systémes 2017).

European Space Agency (ESA) engineers have been developing several spacecraft simulation tools that form the *Mission - Customer Furnished Item (CFI) Software (Mission CFI)*. It includes the following products (ESA 2017):

- The *Earth Observation CFI (EOCFI)* software, which is a collection of multiplatform precompiled C libraries for timing, coordinate conversions, orbit propagation, satellite pointing calculations, and target visibility calculations, specifically parametrized and configured for EO satellites;

- The *EO Orbit and Attitude Adapter (EO Adapter)*, which is part of the Earth Observation Mission Software Suite. It is a tool/library to generate Orbit and Attitude files compliant with EOCFI format using data extracted from one or more binary files, for example files containing Telemetry packets including Orbit and Attitude information;

- The *Envisat CFI* software, which is a collection of multiplatform precompiled C libraries for timing, coordinate conversions, orbit propagation, satellite pointing calculations, and target visibility calculations, specifically parametrized and configured for the Envisat satellite.

The JSDL project presented in this Section stems from the SISO Space Reference FOM standardization initiative carried out by the SISO Space Reference FOM (SRFOM) Product Development Group (PDG) (Möller et al. 2016). JSDL aims at supporting the development of complex space systems by providing high fidelity models and algorithms to manage them. Differently from proprietary and commercial solutions that require tool-specific knowledge and training, JSDL is an open source project released under the open source policy Lesser GNU Public License (LGPL) and can be freely and easily customized and/or extended to cover specific domain aspects. This license allows anybody to build both commercial and noncommercial applications without restrictions or limitations from the use of JSDL. In the following sections the JSDL project is described in details by highlighting its architecture and functionalities.

3. THE JSDL PROJECT

Java Space Dynamics Library (JSDL) is a low-level space dynamics library that facilitates the design and development of space systems, such as space vehicles and satellites. The open source nature of the library allows developers to investigate and customize the architecture and functionalities defined in the source code to fit their own needs.

The JSDL has been designed and developed in the context of the research activities carried out within the SMASH-Lab (System Modeling And Simulation Hub - Laboratory) of the University of Calabria (Italy) working in cooperation with the SISO Space Reference FOM (SRFOM) Product Development Group (PDG) (Möller et al. 2016). The primary goal of JSDL is to provide high fidelity models and algorithms needed for defining space systems that are as accurate and robust as those provided by existing commercial and government software. It is fully implemented in the Java programming language and provides a consistent set of functionalities for developing and running complex elements in space such as, time scales, reference frames, orbital parameters, orbit propagation, and attitude.

The JSDL provides to developers the following resources: (i) the *technical documentation* that describes the library with its philosophy and mission; (ii) the *user guide* to support developers in the use of the library; and (iii) a set of *reference examples* that show how to create space systems.

In the following, the attention is focused on the architecture and services provided by the library.

4. ARCHITECTURE OF THE JSDL

The JSDL library depends only on the Java Standard Edition version 7 (or above), Apache Commons Math (Apache Commons 2017) version 3.6 and JDateTime (Jodd Components 2017) version 3.8 libraries at runtime. The JSDL provides a set of services, each of which defines some Java classes and interfaces that enable specific functionalities. The JSDL architecture is shown in Figure 1.
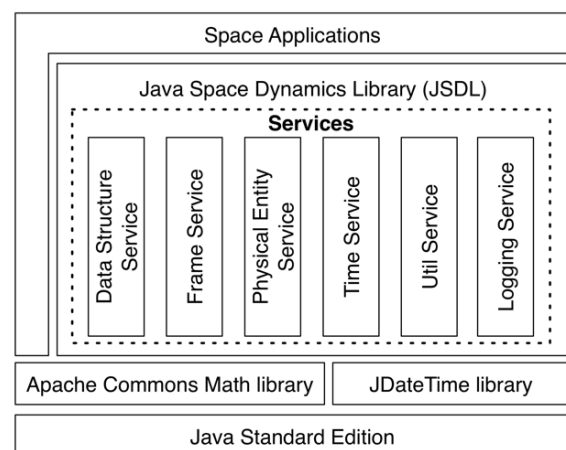

Figure 1: Architecture of the JSDL library.

Space Applications. Contains the space applications that are built using the functionalities provided by the JSDL. An application can interact with the Apache Commons Math and JDateTime directly or through the JSDL library.

Java Space Dynamics Library (JSDL). It is the core library for creating Space applications. It provides a set of features useful for modeling objects in space. The complexity of the features provided is hidden behind an intuitive set of APIs.

Apache Commons Math library. It is a standard library of lightweight, self-contained mathematics and statistics components addressing the most common practical problems not immediately available in the Java programming language (Apache Commons 2017).

JDateTime library. It is a library that offers a very precise way to track dates and time. It uses well-defined and proven astronomical algorithms for time manipulation (Jodd Components 2017).

In the following Sections, the six JSDL services with their UML Class diagrams are described in detail.

5. SERVICES OF THE JSDL

5.1. Data Structure Service

The *Data Structure Service* defines functionalities that ease working with complex data structures. It provides a very useful set of data structures (tree and queue) to build and manage *Reference Frames* and *Physical Entities* with their transformations.

The structure of the *Data Structure Service* is shown in Figure 2 by using a UML Class Diagram.
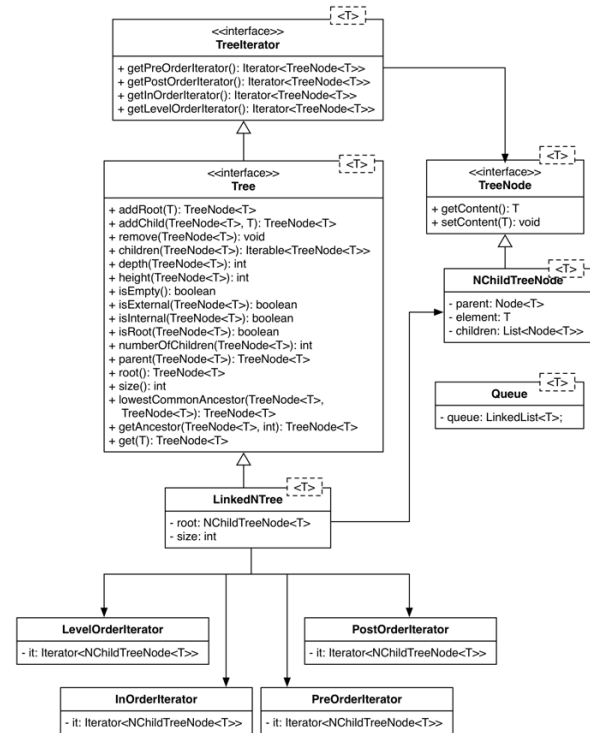
Figure 2: The architecture of the Data Structure Service.

The *LinkedNTree* is a generic class that stores elements hierarchically where each element has a parent element and zero or more children elements. It implements the Tree interface that defines some functionalities to handle a tree such as *height()*, *depth()*, *root()* and *size()*. Moreover, all the common traversal schemes for trees

are provided: *LevelOrderIterator*, *PreOrderIterator*, *InOrderIterator* and *PostOrderIterator*.

The *Queue* class provides a queue data structure that follows the First-in First-out (FIFO) strategy. Elements can only be added to the end (enqueue) and only be removed from the front (dequeue). The queue has been implemented by using a Java standard *LinkedList* and provides two methods *enqueue()* and *dequeue()* to perform each task respectively.

5.2. Frame Service

Reference frame is a fundamental concept for representing when and where a physical entity exists in time and space (Falcone et al. 2014; Möller et al. 2016). This representation is referred to as the state of the entity. In order to represent the state of something, it is necessary to express that state with respect to some time scale and some referent coordinate system. This combination of time and coordinate system is referred as a *Space-Time Coordinate* or *Reference Frame* (Möller et al. 2016). The structure of the Frame Service is shown in Figure 3 by using a UML Class Diagram.
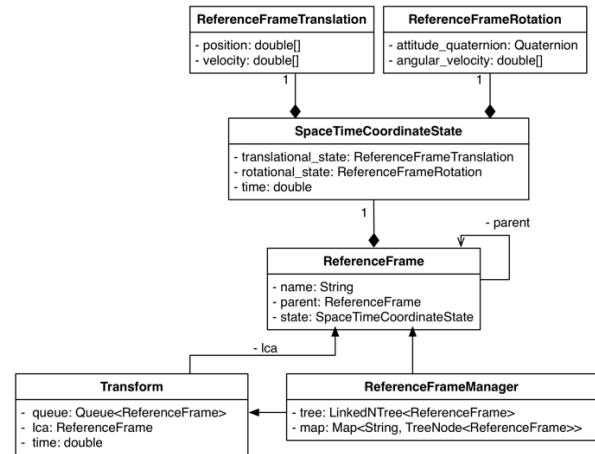
Figure 3: The architecture of the Frame Service.

The *Frame Service* provides functionalities to handle Reference Frames. It includes the fundamental *ReferenceFrame* class that represents a single frame. Each Reference Frame, as defined in the SISO Space Reference FOM (Möller et al. 2016), is composed of three attributes: (i) *name*, which represents the unique name of the reference frame; (ii) *parent,* which is the parent Reference Frame. If it is *NULL*, the Reference Frame is the root frame; and (iii) *space-time coordinate state*, which defines through the *SpaceTimeCoordinateState* Class a four-dimensional representation of the space-time coordinate state with respect to its parent reference frame (Möller et al. 2016). It consists of:

- *Translational state information*, which provides through the *ReferenceFrameTranslation* class a position vector $\vec{v}$ from the origin of the parent reference frame to the origin of the reference frame. It

also provides a velocity vector $\vec{v}$ for the motion of the reference frame with respect to the parent frame. Both of these vectors are expressed with respect to the parent reference frame. These vectors can be used to describe the translational position and motion of a frame with respect to its parent;

- *Rotational state information*, which provides through the *ReferenceFrameRotation* class an attitude quaternion $\tilde{q}$ that describes the attitude of the reference frame with respect to its parent frame. It also provides an angular velocity vector $\vec{w}$ that describes the rotational motion of the reference frame with respect to the parent frame expressed in the subject frame's coordinates. $\tilde{q}$ and $\vec{w}$ can be used to describe the attitude and rotational motion of a frame with respect to its parent.
- *Time*, which contains information about the time t to which the *space-time coordinate state* corresponds.

As shown in Figure 4, all Reference Frames are organized as a tree that is formed from a single base root node with directed paths from an arbitrary number of child nodes.
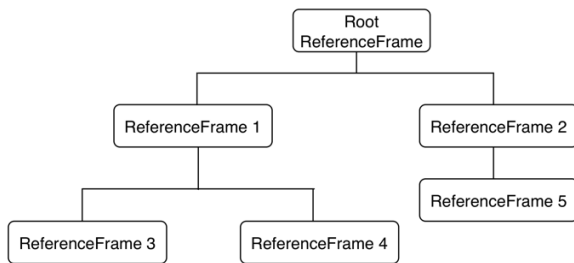
Figure 4: Tree of ReferenceFrames.

These child nodes can then have directed paths from other arbitrary sets of child nodes.

The translational and rotational information can be used to transform a generic vector expressed in a given reference frame $\vec{r}_{child}$ into a vector expressed in its parent frame $\vec{r}_{parent}$. In turn, the vector $\vec{r}_{parent}$ now expressed in the parent frame can be expressed in the parent's parent frame or in another child frame of the parent frame. Chaining together sequences of transformations using the relationships established in the reference frame tree allows for transformation between any pair of frames in the reference frame tree.

Transformations are defined and managed by the *Transform* and *ReferenceFrameManager* classes. In particular, a transformation is computed by merging individual transforms while walking the shortest path between them. The walking/merging operations are handled transparently by the library. Developers only need to select the frames, provide the date and ask for the transformation, without knowing how the frames are related to each other. Transformations are defined as operators that when applied to the coordinates of a

vector expressed in the *initial Reference Frame*, provide the coordinates of the same vector expressed in the *final Reference Frame*.

Equation 1 gives the transformation of a position vector expressed in a child reference frame into a position vector expressed in the parent reference frame (Kuipers 2002),

$$\vec{r}_{parent} = \vec{r}_{0\_parent} + \tilde{Q}(\vec{r}_{child}) \quad (1)$$

where $\vec{r}_{child}$ is the position vector expressed in the child reference frame, $\tilde{Q}(\vec{r}_{child})$ is the quaternion rotation operator associated with the attitude quaternion $\tilde{q}$ that defines the attitude of the child reference frame with respect to the parent reference frame; $\vec{r}_{0\_parent}$ is the vector giving the position of child reference frame origin with respect to the parent reference frame origin expressed in parent reference frame coordinates; $\vec{r}_{parent}$ is the position vector of the entity expressed in parent reference frame coordinates.

With reference to the $\tilde{Q}(\vec{r}_{child})$ operation, it is the canonical way of multiplying a quaternion $\tilde{q}$ by a vector $\vec{x}$ as given by expression (2),

$$\tilde{Q}(\vec{x}) = \tilde{q} \cdot \vec{x} \cdot \tilde{q}^* \quad (2)$$

where $\tilde{q}^*$ is the conjugate of $\tilde{q}$.

The relative motion between a child reference frame and a parent reference frame is provided by the velocity $\vec{v}$ and angular velocity $\vec{w}$ vectors. Equation 3 gives the velocity of an entity expressed in the parent reference frame given the velocity of the entity expressed in the child reference frame (Kuipers 2002),

$$\vec{v}_{parent} = \vec{v}_{0\_parent} + \tilde{Q}(\vec{v}_{child} + (\vec{w}_{child} \times \vec{r}_{child})) \quad (3)$$

where $\vec{v}_{child}$ is the velocity vector of an entity expressed in the child reference frame, $\vec{w}_{child}$ is the angular velocity vector of the child frame with respect to the parent frame and expressed in child frame coordinates, $\vec{v}_{o\_parent}$ is the velocity of the child frame with respect to the parent frame expressed in parent frame coordinates, and $\vec{v}_{parent}$ is the velocity of an entity expressed the parent reference frame.

In most cases, the position and velocity relationships are sufficient. However, acceleration is sometimes needed and is included for completeness. Equation 4 gives the acceleration of an entity expressed in the parent reference frame given the acceleration of the entity expressed in the child reference frame (Kuipers 2002),

$$\vec{a}_{parent} = \vec{a}_{0\_parent} + \tilde{Q}\left(\vec{a}_{child} + (\vec{w}_{child} \times (\vec{w}_{child} \times \vec{r}_{child})) + (2\vec{w}_{child} \times \vec{v}_{child}) + (\vec{\alpha}_{child} \times \vec{r}_{child}) \right) \quad (4)$$

where $\vec{a}_{child}$ is the acceleration of an entity expressed in the child reference frame, $\vec{\alpha}_{child}$ is the angular

acceleration of the child frame with respect to the parent frame and expressed in child frame coordinates, $\vec{a}_{0\_parent}$ is the acceleration of the child frame with respect to the parent frame expressed in parent frame coordinates, and $\vec{a}_{parent}$ is the acceleration of an entity expressed in the parent reference frame.

Concerning reverse transformations; using the child to parent vector transformation equations above defined along with some vector and quaternion algebra, the resulting equation 5 gives the transformation of a position vector expressed in a parent reference frame into a position vector expressed in the child reference frame (Kuipers 2002),

$$\vec{r}_{child} = \tilde{Q}^*\left(\vec{r}_{parent} - \vec{r}_{0\_parent}\right) = -\vec{r}_{0\_child} + \tilde{Q}^*\left(\vec{r}_{parent}\right) \tag{5}$$

where $\tilde{Q}^*\left(\vec{r}_{parent}\right)$ is the conjugate quaternion rotation operator associated with the attitude quaternion $\tilde{q}$ that defines the attitude of the child reference frame with respect to the parent reference frame, and $\vec{r}_{0\_child}$ is the vector giving the position of child reference frame origin with respect to the parent reference frame origin expressed in child reference frame coordinates (Kuipers 2002).

$$\vec{v}_{child} = \tilde{Q}^*\left(\vec{v}_{parent} - \vec{v}_{0\_parent}\right) - \left(\vec{w}_{child} \times \vec{r}_{child}\right) = -\vec{v}_{0\_child} - \left(\vec{w}_{child} \times \vec{r}_{child}\right) + \tilde{Q}^*\left(\vec{v}_{parent}\right) \tag{6}$$

Similar relationships can be derived for velocity (Equation 6) and acceleration (Equation 7) (Kuipers 2002).

$$\vec{a}_{child} = \tilde{Q}^*\left(\vec{a}_{parent} - \vec{a}_{0\_parent}\right) - \left(\vec{w}_{child} \times \left(\vec{w}_{child} \times \vec{r}_{child}\right)\right) - \left(2\vec{w}_{child} \times \vec{v}_{child}\right) - \left(\vec{\alpha}_{child} \times \vec{r}_{child}\right) = -\vec{a}_{0\_child} - \left(\vec{w}_{child} \times \left(\vec{w}_{child} \times \vec{r}_{child}\right)\right) - \left(2\vec{w}_{child} \times \vec{v}_{child}\right) - \left(\vec{\alpha}_{child} \times \vec{r}_{child}\right) + \tilde{Q}^*\left(\vec{a}_{parent}\right) \tag{7}$$

5.3. Physical Entity Service

The structure of the Physical Entity Service is shown in Figure 5 by using a UML Class Diagram.

PhysicalEntity is the highest-level object class in the JSDL entity hierarchy. This class provides attributes to describe an entity's location in time and space. It also contains attributes to uniquely identify it individually from all other physical entities.

Physical entities have two intrinsically associated reference frames: (i) a *structural frame*; and (ii) a *body frame*. These are not registered in the reference frame tree but are used to place and orient the entity in space with respect to a reference frame in the tree. The origin of the *structural frame* is located at some arbitrary but known point on the entity (Möller et al. 2016). The *body frame* origin is at the entity's *center of mass* and is located with respect to the entity's structural reference frame by a vector from the origin of the structural reference frame to the center of mass of the entity. This vector is expressed in the entity's structural reference frame. The orientation of the entity's *body frame* with respect to the entity's structural reference frame is defined by an attitude quaternion.
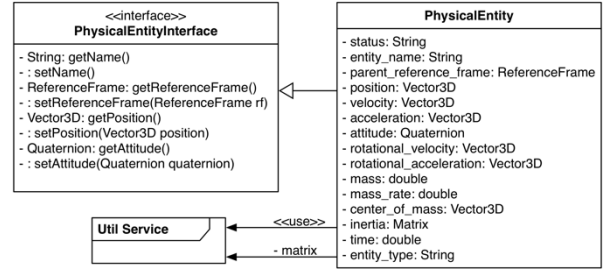
Figure 5: The architecture of the Physical Entity Service.

The *Physical Entity Service* is designed to provide functionalities for space objects such as satellites, asteroids and vehicles. The core attributes defined in the *PhysicalEntity* class includes the position and orientation with respect to a defined parent reference frame, which must be a reference frame instance in the reference frame tree, and a time tag in a defined time scale. This information is sufficient to unambiguously represent an entity in time and space.

5.4. Time Service

The *Time Service* allows to manage epochs, time scales, time units and to compare time instants. The structure of the *Time Service* is shown in Figure 6 by using a UML Class Diagram.
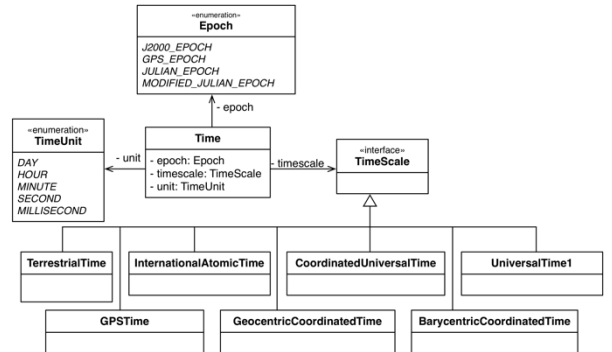
Figure 6: The architecture of the Time Service.

The principal class is *Time* that represents a unique instant in time defined by specifying a point in a specific epoch (e.g., *J2000, GPS* and *Julian epoch*), time scale and time unit (Möller et al. 2016). The *TimeScale* interface defines a set of predefined time scales:

- *Universal Time (UT)*. It is a time standard based on Earth's rotation, defined as the Mean Solar Time at the Royal Observatory in Greenwich, England. There are three variations of Universal Time. UT0 is the observed mean solar time. UT1 is UT0 corrected for polar

motion, the motion of the Earth's rotational axis over the surface of the Earth, and UT2 that is corrected for seasonal variations but today it is considered obsolete.

- *International Atomic Time (TAI)*. It was introduced in 1972 and represents a high-precision atomic coordinate time standard based on the notional passage of proper time on Earth's geoid (Guinot 1986). This time scale is accurate enough to observe relativistic effects for clocks in motion or accelerated by a local gravity field. One advantage of using TAI is that it is a continuous uniform time scale. Specifically, the rate of time passage for TAI is constant unlike the Earth rotation based scales. This means that the Earth rotation based time scales diverge from TAI over time due to the variations in the Earth's rotation. TAI is exactly 36 seconds ahead of UTC. The 36 seconds results from the initial difference of 10 seconds at the start of 1972, plus 26 leap seconds in UTC since 1972.

- *Coordinated Universal Time (UTC)*. It is a 24-hour time standard that is used to synchronize world clocks. UTC is defined by the International Telecommunications Union Recommendation (ITU-R TF.460-6), Standard-frequency and time-signal emissions (Recommendation I., 460-6 2002) and is based on International Atomic Time (TAI) with leap seconds added at irregular intervals to compensate for the slowing of Earth's rotation. Leap seconds are inserted as necessary to keep UTC within 0.9 seconds of universal time, UT1 (Department T.S., United States Naval Observatory).

- *Global Positioning System Time (GPS Time)*. GPS Time is the uniform time scale with a starting epoch at midnight between Saturday January 5th and Sunday January 6th, 1980 (1980 January 6, 00:00:00 UTC). GPS Time counts in weeks and seconds of a week from this instant. The GPS week begins at the transition between Saturday and Sunday. The days of the week are numbered sequentially, with Sunday being 0, Monday 1, Tuesday 2, etc. The GPS time scale begins at the GPS starting epoch with GPS week 0. Within each week, the time is usually denoted as the second of the week (SOW). This is a number between 0 and 604,800 (60 x 60 x 24 x 7). Sometimes SOW is split into a day of week (DOW) between 0 and 6 and a second of day (SOD) between 0 and 86400. While GPST is a uniform time scale, it does have rollover. To limit the size of the numbers used in the data and calculations, the GPS Week Number is a ten-bit count in the range 0-1023, repeating every 1024 weeks. As a result, the week number 'rolled over' from 1023 to 0 at 23:59:47 UTC on Saturday, 21st August 1999. This was before midnight UTC because every GPS week contains exactly 604,800 seconds, to keep the calculations consistent. The 13 intervening leap seconds had put UTC behind GPS system time. The next GPS week rollover occurs on April 6th, 2019.

- *Terrestrial Time (TT)*. It is an astronomical time standard defined by the International Astronomical Union (IAU) used widely for geocentric and topocentric ephemerides. TT is defined to run at the same rate as TAI seconds but with an offset of 32.184 seconds. This offset is based on preserving continuity with other historical dynamic time scales.

- *Geocentric Coordinated Time (TCG)*. It is a coordinate time standard defined in 1991 by the International Astronomical Union (IAU). It is primarily used for theoretical developments based on the Geocentric Celestial Reference System (GCRS). TCG is a relativistic time scale and since the reference frame for TCG is not rotating with the surface of the Earth and not in the gravitational potential of the Earth, TCG ticks faster than clocks on the surface of the Earth by a factor of $6.97 \cdot 10^{-10}$ seconds. TCG, Barycentric Coordinated Time (TCB) and Terrestrial Time (TT) are de-fined in a way that they have the same value on January 1st 1977, 00:00:00 TAI (JD 2443144.5 TAI).

- *Barycentric Coordinated Time (TCB)*. It is a time scale, defined in 1991 by the International Astronomical Union (IAU), primarily used for theoretical developments based on the Barycentric Celestial Reference System (BCRS). TCB is a relativistic time scale and since the reference frame for TCB is not influenced by the gravitational potential caused by the Solar system, TCB ticks faster than clocks on the surface of the Earth by $1.55 \cdot 10^{-8}$ seconds. TCB, Geocentric Coordinated Time (TCG) and Terrestrial Time (TT) are defined in a way that they have the same value on January 1st 1977, 00:00:00 TAI (JD 2443144.5 TAI).

5.5. Util Service

The *Util Service* defines a number of useful functionalities, primarily transformations ones that are useful for working with *Physical Entities* in space. This service should not be considered merely a utility one that is separate from the rest of JSDL; in fact, JSDL depends directly on several of the classes defined in it. Indeed, it provides services needed to define both *Reference Frame* and *Time* objects with their standard conversions.

The structure of the *Util Service* is shown in Figure 7 by using a UML Class Diagram.

The *Matrix* class represents a mathematical matrix. It provides methods for creating matrices, operating on

them arithmetically and algebraically, and determining their mathematical properties such as trace, rank, inverse and determinant.
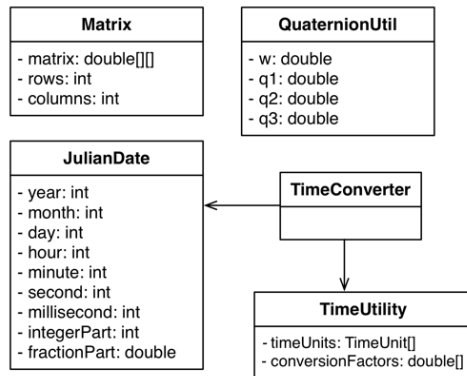

Figure 7: The architecture of the Util Service.

The *QuaternionUtil* class provides classical methods to manage quaternions such as conjugate, inverse and norm. The *JulianDate* class represents a Julian Date, which is a universal time used by all astronomers to ensure that observations are based on a universal astronomical time. It corresponds to the day, hour and minute of the observation and is the interval of time in days since noon at Greenwich on 1 January 4713 BC.

Finally, the *TimeConverter* and *TimeUtility* allow to perform time conversions. Moreover, it is possible to easily convert a *JulianDate* to a standard Java *Calendar* object to have a date/time representation of it through the use of the *toCalendar(JulianDate jd)* method defined in the *TimeConverter* class. For example, the Truncate Julian Date (TJD) 17131.83333333334 can be converted in a Calendar object with value 2015 April 19, 20:00:00 UTC.

5.6. Logging Service

The *Logging Service* provides functionalities useful to both track down any problems or errors occurred during its use, and understand how the JSDL core services work. This information is stored into the *jsdl_trace.log* file.

The structure of the *Logging Service* is shown in Figure 8 through the use of a UML Class Diagram.

Figure 8: The architecture of the Logging Service.

6. CONCLUSION

In the space flight dynamics domain, many research efforts are focusing on the definition of methods, tools and software libraries, mainly aiming at providing a robust and flexible way for defining, building and simulating complex systems in space.

As discussed in the paper, due to the increasing complexity of space systems and thus of the related engineering problems; new methods, tools and software

libraries have been developed in each of these organizations primarily for specific needs and later generalized so as to make them modular, flexible and reusable. The available, commercial and noncommercial, solutions support one or more of the phases in the development of space systems such as flight mechanics, propulsion, orbit controls and data analysis, however none of them seems capable of providing complete coverage of the whole development process of space simulations. To overcome this issue, the *Java Space Dynamics Library (JSDL)* has been created.

The *Java Space Dynamics Library (JSDL)* stems from the SISO Space Reference FOM standardization initiative carried out by the SISO Space Reference FOM (SRFOM) Product Development Group (PDG) (Möller et al. 2016). JSDL is still evolving and aims at supporting the development of complex space systems by providing high fidelity models and algorithms to manage them.

REFERENCES

Apache Commons, 2017. Apache Commons Math home page. Available from: https://commons. apache.org/proper/commons-math/ [accessed 20 Feb 2017].

Rogovchenko-Buffoni, L., Tundis, A., Hossain, M.Z., Nyberg, M., and Fritzson, P., 2014. An integrated toolchain for model based functional safety analysis. Journal of Computational Science, 5(3), pp. 408-414.

CS Communication & Systémes, 2017. Orbits Extrapolation Kit (Orekit) home page. Available from: https://www.orekit.org/ [accessed 20 Feb 2017].

Department T.S., United States Naval Observatory. Leap Seconds home page. Retrieved 17 July 2011. Available from: http://tycho.usno.navy.mil/ [accessed 20 Feb 2017].

ESA, 2017. The Mission CFI software home page. Available from: http://eop-cfi.esa.int/index.php/ mission-cfi software [accessed 20 Feb 2017].

Falcone, A., Garro, A., and Tundis, A., 2014. Modeling and Simulation for the performance evaluation of the on-board communication system of a metro train. In Proceedings of the 13th International Conference on Modeling and Applied Simulation (MAS 2014), pp. 20-29. September 10-12, Bordeaux (France).

Falcone, A., Garro, A., Longo, F., and Spadafora, F., 2014. Simulation exploration experience: A communication system and a 3D real time

visualization for a moon base simulated scenario. In Proceeding of the 18th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, DS-RT '2014, pp. 113-120, October 1-3, Toulouse, France. IEEE Computer Society.

Falcone, A., Garro, A., Taylor, S. J. E., Anagnostou, A., Chaudhry, N. R., and Salah, O., 2016. Experiences in simplifying distributed simulation: The HLA Development Kit Framework. Journal of Simulation, 1-20. ISSN: 1747-7786, DOI: 10.1057/s41273-016-0039-4. Palgrave Macmillan UK.

Falcone, A., Garro, A., Anagnostou, A., Chaudhry, N. R., Salah, O., and Taylor, S. J. E., 2015. Easing the development of HLA Federates: the HLA Development Kit and its exploitation in the SEE Project. In Proceedings of the 19th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, DS-RT '2015, pp. 50-57. October 14-16, Chengdu, China. IEEE Computer Society.

Fortino, G., Garro, A., Mascillaro, S., and Russo, W., 2007. ELDATool: A Statecharts-based Tool for Prototyping Multi-Agent Systems. In Proceeding of the 8th AI*IA/TABOO Joint Workshop "From Objects to Agents": Agents and Industry: Technological Applications of Software Agents, WOA '2007, pp. 14-19, September 24-25, Genova, Italy.

Fortino, G., Garro, A., Russo, W., Caico, R., Cossentino, M., and Termine, F., 2006. Simulation-driven development of multi-agent systems. In Proceedings of the 4th International Industrial Simulation Conference, ISC '2006, pp. 17-24, June 5-7, Palermo, Italy. EUROSIS.

Gaylor, D., Page, R. and Bradley, K., 2006. Testing of the java astrodynamics toolkit propagator. In AIAA/AAS Astrodynamics Specialist Conference and Exhibit, p. 6754.

Garro, A., and Falcone, A., 2015. On the integration of HLA and FMI for supporting interoperability and reusability in distributed simulation. In Proceedings of the Symposium on Theory of Modeling and Simulation - DEVS Integrative M&S Symposium, DEVS 2015, Part of the 2015 Spring Simulation Multi-Conference, SpringSim 2015, pp. 9-16. April 12-15, Alexandria, USA,. Society for Computer Simulation International.

Garro, A., Falcone, A., Chaudhry, N. R., Salah, O. A., Anagnostou, A., and Taylor, S. J., 2015. A prototype HLA development kit: results from the 2015 simulation exploration experience. In Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, pp. 45-46, June 10-12, London, UK. IEEE Computer Society.

Guinot, B., 1986: Is the international atomic time tai a coordinate time or a proper time? Celestial mechanics 38(2), pp. 155–161.

Ido, H., 2012. Space Flight Dynamics as a Service.

Jodd Components, 2017. JDateTime home page. Available from: http://www.jodd.org/doc/jdate time.html [accessed 20 Feb 2017]

Möller, B., Garro, A., Falcone, A., Crues, E. Z. and Dexter, D. E., 2016. Promoting a-priori interoperability of HLA-based Simulations in the Space domain: the SISO Space Reference FOM initiative. Proceedings of the 20th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications DS-RT 2016, pp. 100-107. September 21-23, London, UK. IEEE Computer Society.

Kuipers, J., 2002. Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality.

Pulecchi, T. and Lovera, M., 2006. A modelica library for space flight dynamics. In Proceedings of the 5th International Modelica Conference.

San-Juan, J. F., Lara, M., López, R., López, L. M., Folcik, Z. J., Weeden, B. and Cefola, P. J., 2011. Using the DSST semi-analytical orbit propagator package via the NonDyWebTools/ AstroDyWebTools open science environment. RdM, 6(1), 2π.

Recommendation, I.: 460-6, 2002. Standard-frequency and time-signal emissions (questionitu-r 102/7). ITU-R Recommendations: Time Signals and Frequency Standards Emission, Geneva, International Telecommunications Union, Radio-communication Bureau.

AUTHORS BIOGRAPHY

Alberto Falcone

Alberto Falcone is a PhD student in Information and Communication Engineering for Pervasive Intelligent Environments at University of Calabria (Italy). In 2016, he was Visiting Researcher at NASA Johnson Space Center (JSC), working with the Software, Robotics, and Simulation Division (ER). He is a member of the Executive Committee as Student Team Coordinator of the Simulation Exploration Experience (SEE) project.

Alfredo Garro

Alfredo Garro is an Associate Professor of Computer and Systems Engineering at the Department of Informatics, Modeling, Electronics and Systems Engineering (DIMES) of the University of Calabria (Italy). In 2016, he was Visiting Professor at NASA Johnson Space Center (JSC), working with the Software, Robotics, and Simulation Division (ER). He is vice chair of the Space Reference Federation Object Model (SRFOM) Product Development Group (PDG) of SISO. He is the Technical Director of the "Italian Chapter" of INCOSE. He is involved as an IEEE senior member in the activities of the IEEE Computer Society, IEEE Reliability Society and IEEE Aerospace and Electronic Systems Society.

EFFECT OF ANKLE JOINT POSITION ON BIPED ROBOT WALKING BEHAVIOUR

Van-Tinh Nguyen[a], Ngoc-Linh Tao[b], Hiroshi Hasegawa[c]

[a], [b], [c]Graduate School of Engineering and Science
Shibaura Institute of Technology, Japan
[a]School of Mechanical Engineering
Hanoi University of Science and Technology, Vietnam

[a]nb16508@shibaura-it.ac.jp, [c]h-hase@shibaura-it.ac.jp

ABSTRACT

This paper addresses the effect of ankle joint position on the walking behavior of a biped robot. The mentioned foot structure consists of a tiptoe and a big toe inspired by the human foot which have a crucial role on moving stability. The study subject is a small robot called Kondo KHR-3HV, belonging to the Kondo Kagaku Company. Due to the small size of the robot and considering a reduction in energy consumption in toe mechanism, a passive joint using torsion spring was selected as a toe joint. The gait generation method, for finding the proper position of ankle joint, is used by varying the ankle joint position. There are two requirements of robot design: go straight and stay within setting conditions. The paper is implemented by two stages. First, the biped robot locomotion is considered by different stiffness coefficients to find out what is the proper stiffness coefficient. In the second stage, the simulation of all small biped robot models which have the different ankle joint position, can walk within setting conditions, is implemented. The results are compared to the human ankle joint trajectory in gait performance and frequency and are confirmed by dynamic simulation on Adams (MSC company, USA).

Keywords: biped robot, ankle joint position, walking behavior, big toe, torsion spring

1. INTRODUCTION

The human body has a complicated physical structure and implements difficult movements. During the past several decades, many researchers in the world have concentrated on the field of the biped robot inspired by the human body (Sakagami et al. 2002; Lohmeier et al. 2004; Ogura et al. 2006a; Ishida et al. 2004). The first aim of researches carried out in this field attempts to solve the following problem: "How can the robot walk naturally and stably?". This goal is motivated by several applications of the biped robot development such assistance, entertainment and medical issues. Hence, they have to move in a domestic environment and should have the same ability as humans to carry out stable walking.

In almost every previous studies, the feet of the biped robot have been designed with the rigid flat sole structure which cannot provide the best contact with the ground while in locomotion. Sometimes, it is a point contact at the corner of the sole as depicted in Figure 1, thus, the number of the contact point is reduced. Consequently, the support polygon area and the stability of the robot also decrease.

Furthermore, one of the characteristics of human walk is heel-contact and toe-off motion in steady walking. To implement adaptive walking, a foot is one of the most important regions of the human body in bipedal locomotion because it is the only region that has a direct physical interaction with the environment. The human foot has a complicated structure which consists of toes and several joints. On a human walking cycle, this structure makes the ground reaction force smoothly change in toe-off. Thus, it helps the contact between human foot and ground be smooth, having an important role in walking stability.

To overcome this challenge, from human foot inspiration, there have been some papers mentioned on the flexible foot structure for the biped robot. For instance, Yu Ogura et al. have proposed a new foot mechanism by implementing one passive joint for bending toe motion of Wabian-2R. However, in this study, the number of the robot's Degree of Freedom (DoF) is reduced due to the predetermination is complemented by waist rolling motion (Ogura et al. 2006b). Yamane and Trutoiu (2009) have investigated feet composed of curved surfaces at toe and heel and also a flat section for a simple planar biped robot. Sellaouti et al. (2006) have developed a new model of the humanoid robot HRP-2 with passive tiptoe joints to enhance its walking speed. Lohmeier et al. (2006) have designed the humanoid robot LOLA with an actively driven toe joints. However, the above-mentioned papers mainly focus on the humanoid robot whose parameters are similar to the human's ones. The human-size robots are very convenient for designing structure and integrating an actuator on the feet.

However, for a small bipedal robot, it is difficult to build a foot structure by limited parameters. Nerakae and Hasegawa (2014a) have presented the foot mechanism

with big toe and tiptoe for a 10 DoFs small bipedal robot. The mentioned foot structure equips the robot with a good adaptation. It enables the foot to increase the contact points and improves the stability as described in Figure 2. Nevertheless, in the above-mentioned work, the trajectories of all the joints on both legs are generated by seven isolated gait functions which make a gait pattern generation become complicated. Simultaneously, the robot cannot walk naturally in comparison to the human motion. In addition, the torsion stiffness coefficients are only considered with two values and the ankle joint position is fixed based on the reference of the real robot. It is unreasonable because of the changed robot foot structure.
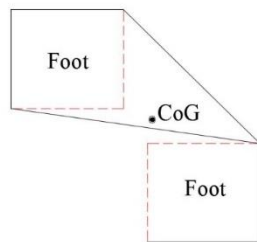
Figure 1: An example presenting the contact points, the support polygon and its center of gravity
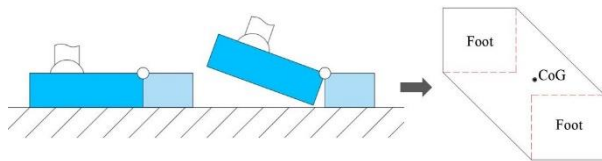
Figure 2: An adaptive foot structure

This study continues to develop a foot structure for a small robot proposed by Nerakae and Hasegawa (2014b). The paper implements to investigate the effect of two characteristics: spring stiffness and ankle joint position on robot walking behaviour. This is to aim to determine the consistent stiffness coefficient for toe joints when the robot performs its locomotion on flat ground. In the second stage, the simulation results of all small biped robot models when changing ankle joint position is compared with human walking behaviour to witness the effect of the changed ankle joint position on the robot walking behaviour and gait functions. It can be said that its walking style, in comparison to those of the other small biped robots, is more similar to that of humans.

This paper is organized in the following manner: A mechanical description of robot is presented in Section 2. The principle of gait pattern generation is in Section 3. Section 4 mentions the simulation procedure. Section 5 shows the results of the development of the robot by dynamic simulation on ADAMS. Finally, Section 6 includes some brief conclusions and future works.

2. EXPERIMENTAL ROBOT MODEL

2.1. Overview of structural design
In this study, the proposed model is built based on the KHR-3HV robot of the Kondo Kagaku Company which is the third generation of humanoid robots developed by this company. The KHR-3HV robot has the weight of 1.5kg, the height of 401.05mm and up to 22 DOFs with 17 actual servos and 5 dummy servos. However, in this work, only the robot's legs are focused on. Thus, the upper body joints are fixed and the lower body has 10 controlled joints for the legs as shown in Figure 3.
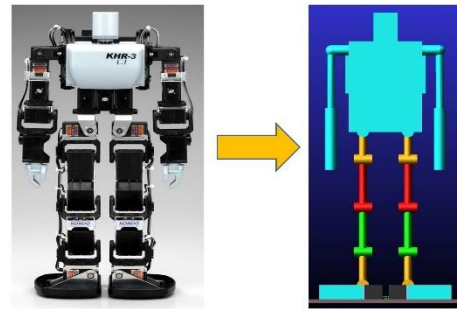
Figure 3: Real robot and experimental model

2.2. Foot mechanism
During locomotion, the human feet support area continuously varies on the sole of each foot as depicted in Figure 4. The black area is the position where supports force areas. Wherewith, LR is heel only in loading response, MSt is foot flat in mid stance, TSt is forefoot and toes is terminal stance, and PSw is medial forefoot in pre-swing. Perry and Burnfield (2010a) found that toe contact with ground is quite variable. The onset of toe involvement followed insolated forefoot support by 10% of the stance period. In this period, toe pressures differ markedly with the greatest pressure of the big toe. It ranged between 30% and 55% of that at the heel. Thus, the big toe has an important role in the human walking, especially during the toe-off period.
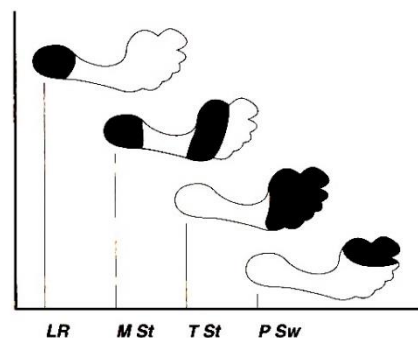
Figure 4. Sequence of foot support areas during stance (Perry and Burnfield 2010b)

By this idea, Nerakae and Hasegawa (2014c) have proposed the foot structure for enhancing the walking behavior of the biped robot as depicted in Figure 5. Their study exhibited that the big toe is a significant part to support and transfer weight from one foot to another foot.

However, in their paper, some parameters are predefined or referred to the real robot such as torsion stiffness coefficient and ankle joint position. Thus, this work is based on the assumption that those mentioned parameters have an effect on the walking behavior, walking distance and gait function. It is considered in a predefined range as described in Table 1 and Table 2.
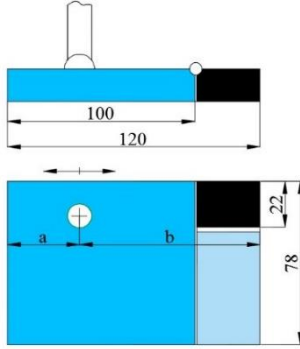
Figure 5. Robot foot structure

Table 1: Torsion spring stiffness coefficient

N_o	d(mm)	T(Kg.mm/210°)
M1	9.30	5.88
M2	9.37	7.60
M3	9.53	10.60
M4	11.71	13.71
M5	11.81	16.82
M6	11.81	17.40
M7	11.96	22.81
M8	14.12	27.19
M9	14.33	38.14
M10	14.45	45.28

Table 2: Ankle joint position

N_o	a(mm)	b(mm)	R=a/(a+b)
F1	80	40	0.67
F2	70	50	0.58
F3	60	60	0.50
F4	50	70	0.42
F5	40	80	0.33
F6	30	90	0.25
F7	20	100	0.17

3. GAIT FUNCTION

The joint angles are defined as described in Figure 6. The range of the angle is based on human motion data as Table 3.
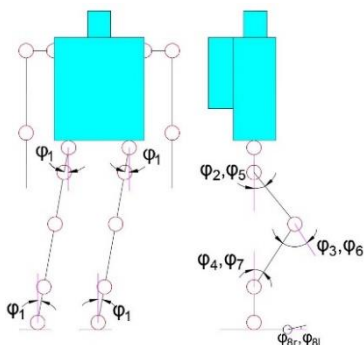
Figure 6. Robot linkage model

Based on the human walking pattern as depicted in Whittle (2007a), the paper supposes that the robot control data was generated by the gait function as a trigonometric function shown in Equation (1).

$$\varphi_i(t) = a_i + b_i.\cos(\omega t) + c_i.\sin(\omega t) + d_i.\cos(2\omega t) \qquad (1)$$

Table 3: The range of joint angle

Angle	View plane	Leg	Joint	Value
φ_1	Frontal	Both	Hip & ankle	-15° to 15°
φ_2	Sagittal	Right	Hip	-50° to 50°
φ_3	Sagittal	Right	Knee	0° to 60°
φ_4	Sagittal	Right	Ankle	-50° to 50°
φ_5	Sagittal	Left	Hip	-50° to 50°
φ_6	Sagittal	Left	Knee	0° to 60°
φ_7	Sagittal	Left	Ankle	-50° to 50°
φ_{8r}	Sagittal	Right	Proximal phalanx	0° to 30°
φ_{8l}	Sagittal	Left	Distal phalanx	0° to 30°

Where φ_i is the angle of i joint, a, b, c, d are coefficients, t is time, and ω is angular velocity. By changing a, b, c, d coefficients, the gait function will be created to allocate to each joint of the robot.

In this study, the bipedal robot is considered the locomotion on flat ground with the total time of 4.8 seconds. The robot is simulated in 3 cycles with a time period of 3.6 seconds, 1.2 left seconds are used for checking robot stability. One cycle is set up to 1.2 seconds. As a results, the angular velocity is calculated by Equation (2). In simulation, one step takes 0.02 second, the total number of step is 240. In the second cycle, the biped robot performs its motion the most natural, hence this cycle will be selected to show the waveform of the gait function as well as the robot walking behavior.

$$\omega = \frac{2\Pi}{1.2} = 5.236 \qquad (2)$$

Gait functions are assigned to all joints as shown in Equation (3-9).

$$\varphi_1 = \begin{cases} 0, & t = 0 \text{ or } t \geq 3.6 \\ \pm1.5, & t = 0.3 \text{ and } t = 3.3 \\ \varphi_1(t), & 0.3 < t < 3.3 \end{cases} \qquad (3)$$

$$\varphi_2 = \begin{cases} 0, & t \leq 0.3 \text{ or } t \geq 3.6 \\ \varphi_2(t + 0.6), & 0.3 < t < 3.3 \\ 15, & t = 3.3 \end{cases} \qquad (4)$$

$$\varphi_3 = \begin{cases} 0, & t \leq 0.3 \text{ or } t \geq 3.6 \\ \varphi_3(t + 0.6), & 0.3 < t < 3.3 \\ 30, & t = 3.3 \end{cases} \qquad (5)$$

$$\varphi_4 = \begin{cases} 0, & t \leq 0.3 \text{ or } t \geq 3.6 \\ \varphi_4(t + 0.6), & 0.3 < t < 3.3 \\ 15, & t = 3.3 \end{cases} \qquad (6)$$

$$\varphi_5 = \begin{cases} 0, & t = 0 \text{ or } t \geq 3.3 \\ 15, & t = 0.3 \\ \varphi_2(t), & 0.3 < t < 3.3 \end{cases} \quad (7)$$

$$\varphi_6 = \begin{cases} 0, & t = 0 \text{ or } t \geq 3.3 \\ 30, & t = 0.3 \\ \varphi_3(t), & 0.3 < t < 3.3 \end{cases} \quad (8)$$

$$\varphi_7 = \begin{cases} 0, & t = 0 \text{ or } t \geq 3.3 \\ 15, & t = 0.3 \\ \varphi_4(t), & 0.3 < t < 3.3 \end{cases} \quad (9)$$

In toe mechanism, due to considering a reduction in energy consumption of the robot, the passive joint is selected as a toe joint. Consequently, φ_{8r} and φ_{8l} have a value in the range from 0° to 30°. Their values depend on the robot's geometric posture as well as an impact force when the robot performs its motion.

4. SIMULATION PROCEDURE

The concept of the optimization process is shown as in Figure 7. Z_f and X_f denote the distance from initial position to final position along z axis and x axis in the robot locomotion, respectively. R_f is the angle of rotation. Definition of optimal design is described as Equation (10 - 17).
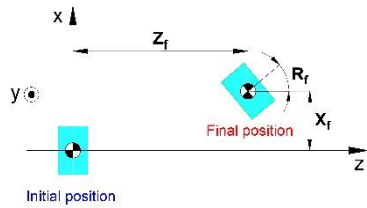
Figure 7. Overview of optimization

Design variables (DVs):

$$x = [a_i, b_i, c_i, d_i], \; i = 1 \div 4 \quad (10)$$

Constraint functions:

$$g_1(x) = 20 - |X_f| \geq 0 \quad (11)$$

$$g_2(x) = 5 - |R_f| \geq 0 \quad (12)$$

$$h_1(x) = 243.53 - Y_f = 0 \quad (13)$$

$$h_2(x) = N - 240 = 0 \quad (14)$$

Where Y_f is distance from Centre of Mass (CoM) to ground. N is a total simulation step.

Objective function:

$$f(x) = -Z_f \rightarrow \min \quad (15)$$

Penalty function:

$$P(x) = \sum_{i=1}^{2} \{\min[g_i(x), 0]\}^2 + \sum_{j=1}^{2} [h_j(x)]^2 \quad (16)$$

Modified objective function:

$$F(x) = -Z_f + \gamma.P \rightarrow \min \quad (17)$$

Where a_i, b_i, c_i, d_i (i=1, 2, 3, 4) are the coefficients of the gait function. There are four constraint functions. In Equation 11-12, X_f distance and R_f angle are constrained under ±20mm and ±5° to ensure that the biped robot can walk straight. In Equation 13, Y_f must be equal to 243.53mm to ensure the robot not to slip and fall down at the final framework. In Equation 14, N is equal to 240 to check the success of the simulation. In Equation 17, γ is a penalty coefficient set to 1000. Equation 11-14 will be also checked again when the simulation finishes.

5. SIMULATION RESULT

5.1. First experiment
The result is shown in Figure 8. As can be seen that all model can walk on flat ground. Side distance and angle of rotation in the simulation have a little difference when comparing with the calculation results on account of the approximating method. In consideration of walking straight and distance, model M7 with the stiffness coefficient of 22.81 Kg.mm/210° has the best performance. When the torsion stiffness increases, the long distance also go up, however, the side distance and angle of rotation have a same way. In contrast, this study plans to decrease the torsion stiffness, the long distance decreases. Thus, model 7 is selected to push the research further. The waveforms of the gait functions assigned to all joints are depicted in Figure 9. In comparison with the gait pattern of the human depicted in Whittle (2007b), as can be seen in Figure 9b and Figure 9c that the hip and the knee joint gait functions are similar to that of human beings. The difference of the ankle joint gait function is expected to occur as consequence of the physical structure dissimilarity with the humans' one.

5.2. Second experiment
The result of the second experiment is shown in Figure 10

The robot ankle joint trajectory of all experiments is shown in Figure 11, its data is collected in the second cycle since the biped robot performs the most natural and stable locomotion in this period. By comparison, human ankle joint trajectory is depicted in Figure 12. The subject in this study was a man. He was aged 47, 167 cm in height, and weighed 67 kg. This kinematic data for lower body while walking was measured using a Mac3D system (Motion Analysis Corporation). Data was recorded at a sampling rate of 200 Hz while the subject was walking.

As it can be seen, in general, the robot ankle joint trajectory has a frequency and a trend similar to the humans' one. From F1 to F7, the height of ankle joint change to adapt to the new ankle joint position. F7 position is near the robot's heel, and same as the human situation. However, the performance of this model is not so good. F4 and F5 position at the middle have the best performances which is the most comparable to human ankle joint trajectory, thus, these ankle joint positions are selected.

Waveform comparison of gait function assigned to all joints are depicted in Figure 13.

When the paper plans to change the ankle joint position from F1 to F7 as described in Table 2, the knee joint angle has gradually declined. The other joint angles are almost constant or change in small amount. Specially, the hip joint angle only has a small change at 0, 0.5, and 1 in a cycle. The ankle pitches joint angle changes at almost time. Figure 10 show that F5 model performs the best result, thus, this model should be selected for the next research in the future.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Zd(mm)	245	240	236	238	229	217	204	189	170	155
|Xd|(mm)	23.10	10.14	19.87	23.81	8.92	6.20	7.43	0.58	3.18	7.24
|Rd|(deg)	15.12	17.69	9.59	7.87	10.84	6.68	5.10	6.28	4.08	0.21
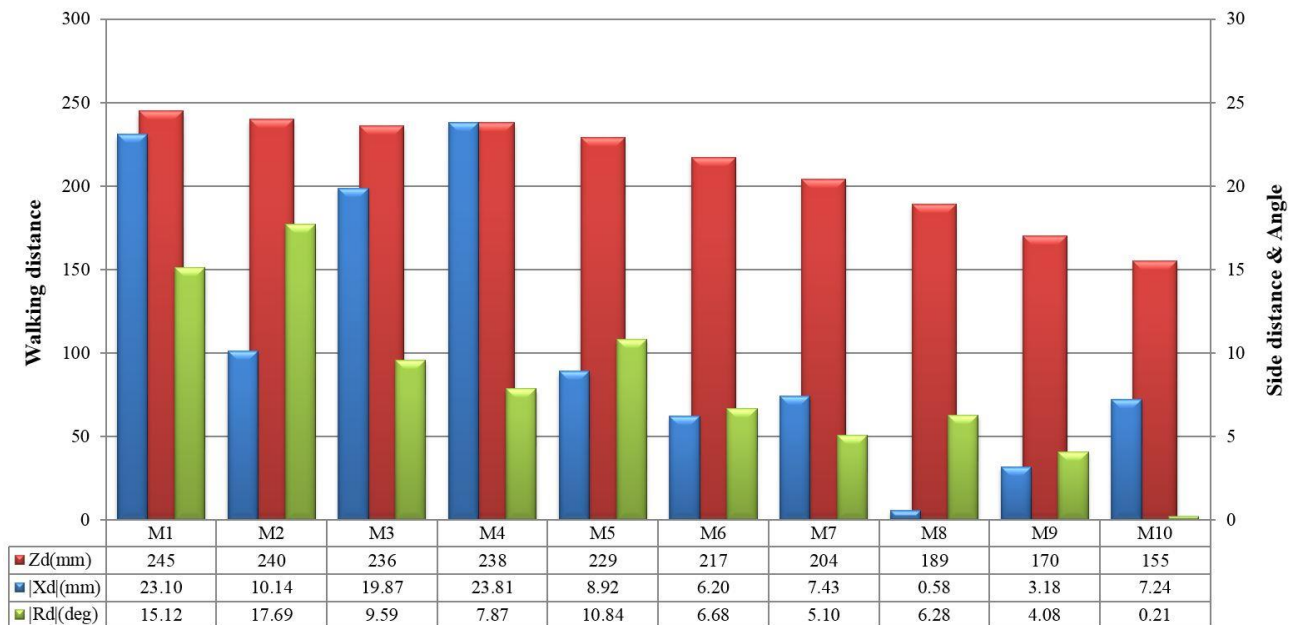
Figure 8. Result of the first experiment
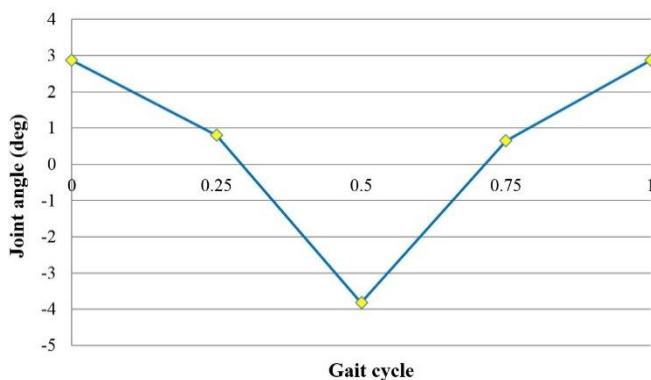
(9a) A cycle of gait function (hip and ankle roll joint angle)
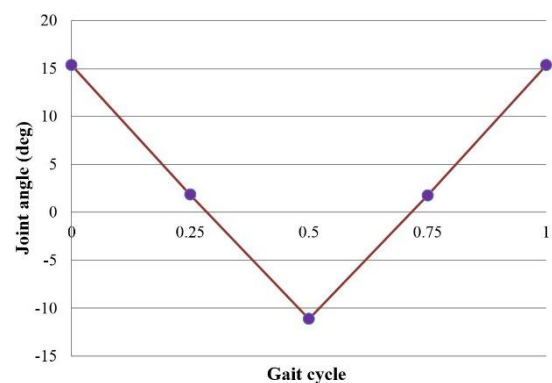
(9b) A cycle of gait function (hip pitch joint angle)
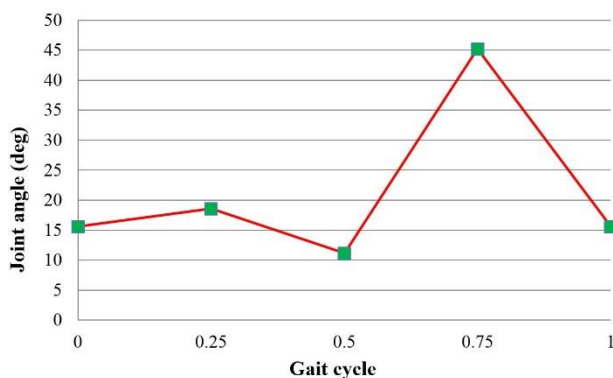
(9a) A cycle of gait function (knee pitch joint angle)
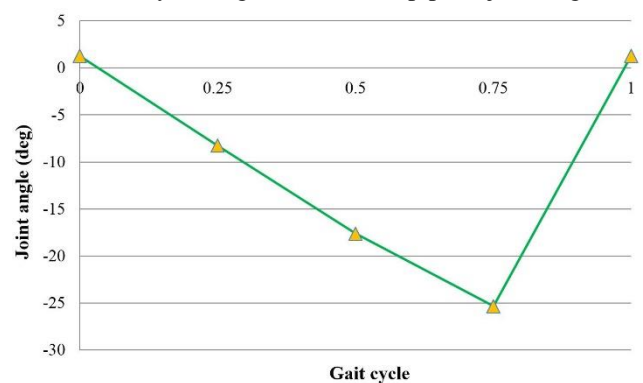
(9a) A cycle of gait function (ankle pitch joint angle)

Figure 9. Waveform of the gait function

Figure 10. Result of the second experiment

	F1	F2	F3	F4	F5	F6	F7
Zd(mm)	196	196	199	197	197	183	169
|Xd|(mm)	4.83	3.84	3.96	6.56	3.08	1.18	1.68
|Rd|(deg)	0.04	3.09	4.18	5.10	0.23	3.32	0.41

Figure 11. Robot ankle trajectory in a cycle

Figure 12. Human ankle trajectory in a cycle

(13a) A cycle of gait function (hip and ankle roll joint angle)
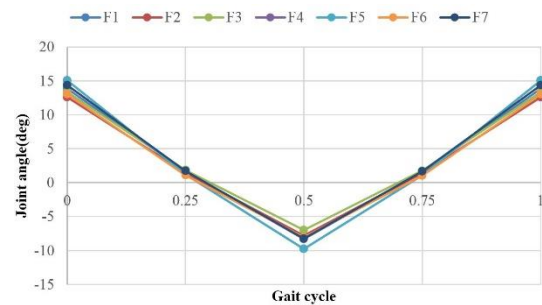
(13b) A cycle of gait function (hip pitch joint angle)
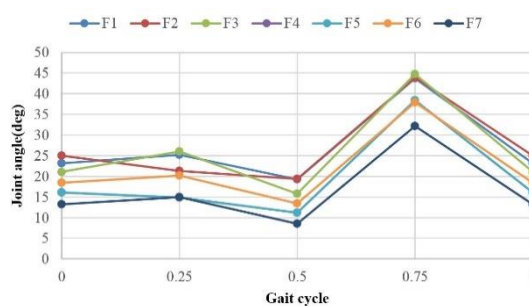
(13c) A cycle of gait function (knee pitch joint angle)
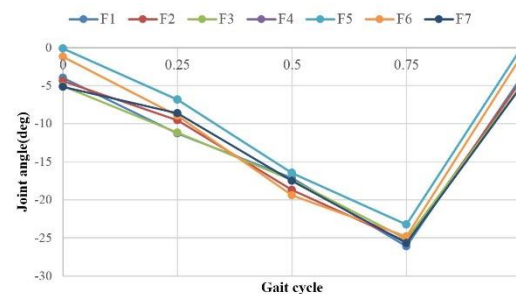
(13d) A cycle of gait function (ankle pitch joint angle)

Figure 13. Waveform of the gait function

6. CONCLUSIONS

The ankle joint position has the crucial effect on the walking behavior as well as the gait pattern of the biped robot. In this paper, the effect of two foot characteristics is conside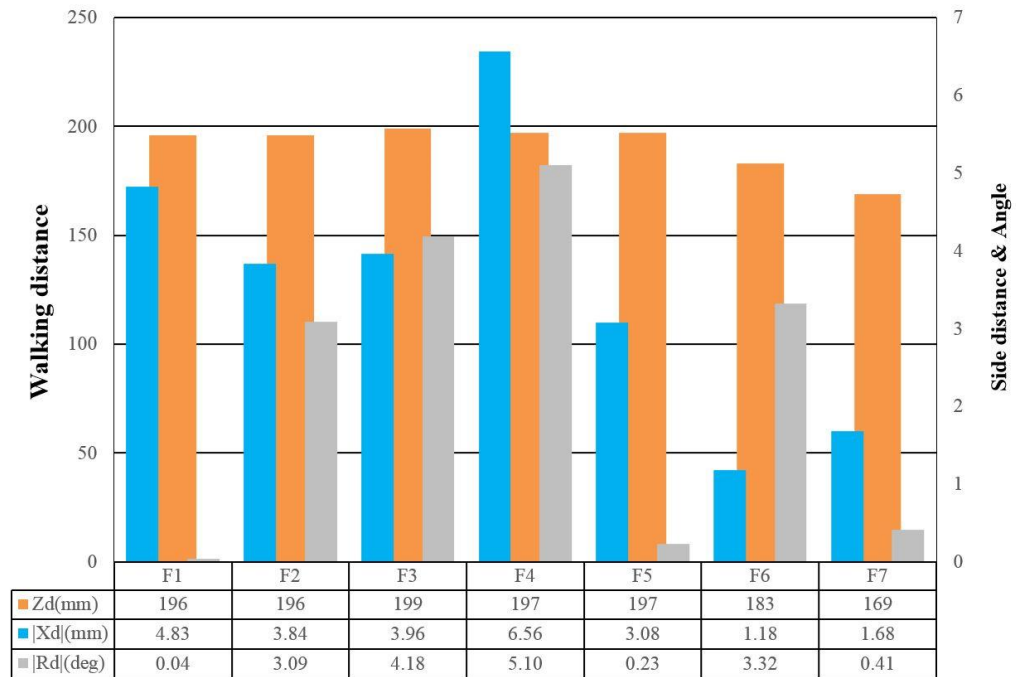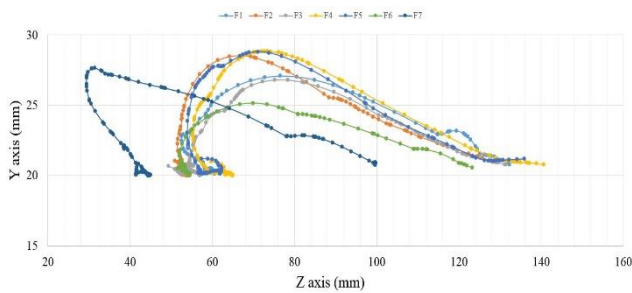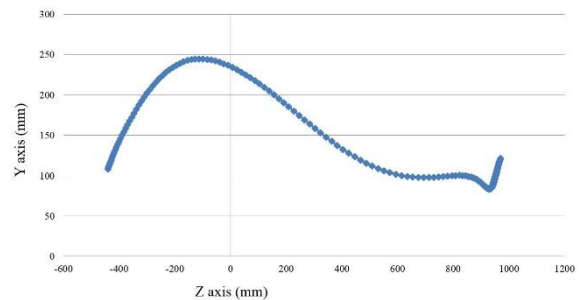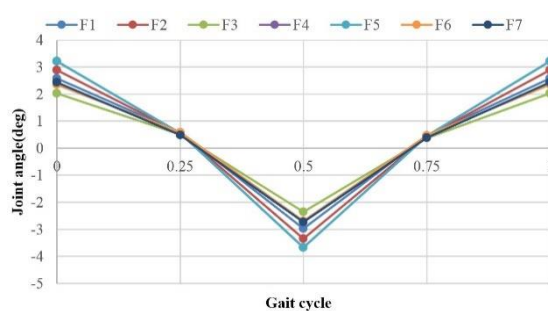red. By the first experiment, the paper found out the most consistent spring stiffness coefficient for the proposed robot which is 22.81 Kg.mm/210º. Through the second experiment, all models with changing ankle joint position can walk straight and within the constraint conditions. The gait functions are successfully generated by the approximated optimization method to each model. The robot ankle joint trajectory is compared with that of the human to find out the best position for the ankle joint.

For a future work, the study will plan to adjust the length of the toes to learn more and consider the locomotion of the robot on a rough ground.

REFERENCES

Sakagami Y., Watanabe R., Aoyama C., Matsunaga S., Higaki N., and Fujimura K., 2002. The intelligent ASIMO, system overview and integration. Proceedings of The IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2478–2483, Switzerland.

Lohmeier S., Loffler K., Gienger M., Ulbrich H., and Pfeiffler F., 2004. Conputer system and control of biped "Johnnie". Proceedings of IEEE International Conference on Robotics and Automation, pp. 4222-4227, New Orleans, LA.

Ogura Y., Aikawa H., Shimomura K., Morishima A., Lim H. O., and Takanishi A., 2006. Development of a new humanoid robot WABIAN-2. Proceedings of IEEE International Conference on Robotics and Automation, pp. 830-835, Orlando, Florida.

Ishida T., 2004. Development of a Small Biped Entertainment Robot QRIO. Proceedings of the Fourth Symposium Micro-Nano mechatronics for Information-Based Society, pp. 23-28.

Yamane K., and Trutoiu L., 2009. Effect of Foot Shape on Locomotion of Active Biped Robots, a Study on Effect of Two-Arch Structure of Foot for Biped Robot. Proceeding of 9th IEEE-RAS International Conference on Humanoid Robot, pp. 2230-2236, Paris, France.

Sellaouti R., Stasse O., Kajita S., Yokoi K., and Kheddar A., 2006. Faster and smoother walking of Humanoid HRP-2 with passive toe joints. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4909-4914, Beijing, China.

Lohmeier S., Buschmann T., Ulbrich H., and Pfeiffer F., 2006. Modular joint design for performance enhanced humanoid robot LOLA. Proceedings of IEEE International Conference on Robotics and Automation, pp. 88-93, Florida, USA.

Nerakae K., and Hasegawa H., 2014. Big toe sizing design of small biped robot by using gait generation method. Applied Mechanics and Materials, 541-542 pp. 1079-86.

Perry J., and Burnfield J. M, 2010. Gait Analysis: Normal and Pathological Function, 2nd ed., Slack Incorporated, p. 82.

Whittle M. W., 2007. An Introduction to Gait Analysis. 4th ed., Oxford, p 59.

AUTHORS BIOGRAPHY

Van-Tinh Nguyen received the B.E. (2012) from Hanoi University of Science and Technology (HUST), Vietnam, and M.E. (2016) from Shibaura Institute of Technology (SIT), Japan. Besides, he works as a lecturer at School of Mechanical Engineering of HUST. Currently, he is pursuing Dr. Eng. degree in Functional Control Systems at SIT. His research interests include optimization system design, multi-body systems, humanoid robot, and evolutionary algorithm.

Ngoc-Linh Tao received his B.E. in 2010 at Hanoi University of Science and Technology, Vietnam and M.E. in 2013 from Taiwan University of Science and Technology, Taiwan. He is pursuing Dr. Eng. degree at the Department of Functional Control Systems, Graduate School of Engineering and Science, Shibaura Institute of Technology, Japan. His research interests include 3D computer vision and intelligent algorithms.

Hiroshi Hasegawa received his B.E. in 1992 and M.E. in 1994 from Shibaura Institute of Technology, Japan. He received Dr. Eng. In 1998 in mechanical engineering from Tokyo Institute of Technology, Japan. He has been working at Shibaura Institute of Technology, and currently is a Professor at the Department of Machinery and Control System, College of Systems Engineering and Science. He is a member of JSEE, JSME, ASME, JSCES, JSST and KES. His research interests include computer-aided exploration, creativity of design and systems engineering.

Fuzzy-Variable Gain PI Control of WECS Based on a Doubly Fed Induction Generator

Tsiory Patrick ANDRIANANTENAINA [(a)], Zazad d'Edissak TSARALAHY [(b)], Jean Nirinarison RAZAFINJAKA [(c)], Hervé MANGEL [(d)]

[(a) (c)] Automatic Laboratory, Department of Electricity, Higher Polytechnic School, University of Antsiranana, Madagascar
[(b)] Electrical machines Laboratory, Department of Electricity, Higher Polytechnic School, University of Antsiranana, Madagascar
[(d)] Research institute Dupuy de Lôme, Department of Electricity, Universities Institute of Technology, University of Brest, France
otantikroysti@gmail.com[(a)], zazad86@gmail.com[(b)], razafinjaka@yahoo.fr[(c)], Herve.Mangel@univ-brest.fr[(d)]

ABSTRACT

This paper presents a study of powers control for a Doubly Fed Induction Generator (DFIG) used in Wind Energy Conversion System (WECS). For this purpose, a new topology using hybrid controller is applied for the powers generated by the DFIG. The hybridization consists to combine a variable gain PI (VGPI) controller with a fuzzy logic one. The results of simulation show that this technique can be realized and leads to good performances as disturbance rejection and robustness with respect of operating variation and parametric variation of the machine.

Keywords—DFIG, vector control, WESC, power control, VGPI, fuzzy logic, hybridization.

1- INTRODUCTION

The development and the exploitation of renewable energies met a great growth these last years. Among these sources of energies, the windmill represents a significant potential not to replace existing energies, but to give solutions for the request, which always increases. The wind power can contribute with a significant part for the new sources of energy not emitting a gas for purpose of greenhouse (Andrianantenaina and *al* 2015). Currently, windmill system with variable speed based on the DFIG is widely used. Indeed, the DFIG presents more advantages. Several controls applied on the DFIG have been already proposed as (Andrianantenaina and *al* 2015; Razafinjaka and Andrianantenaina 2015,2016; Boualouch 2015) which give good performances.

These last years, several researches are about the intelligent controllers such us fuzzy logic and neural network which have enjoyed great success in recent years for their robustness against disturbances which may affect process. Here, a hybridization of fuzzy logic with variable gain PI is proposed. After modeling the wind turbine and DFIG, we have established a vector control to control the active and reactive power control. The aim of this work is to present the performance and robustness of these controllers.

2- WIND POWER CONVERSION SYSTEM

Figure 1 shows a general scheme of the system which is composed by a turbine, a multiplier, the DFIG and two converters.
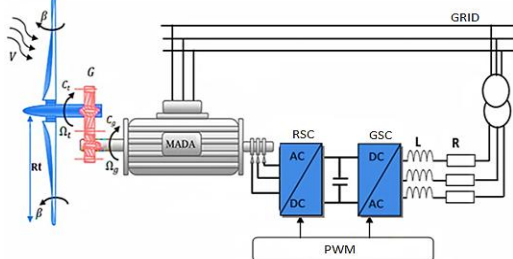
Figure 1: General scheme of wind turbine based on DFIG

The total kinetic is,

$$P = \frac{1}{2}\rho\pi R_T^{\,2} V^3 C_p(\lambda, \beta) \tag{1}$$

With ρ the air density, V, the wind velocity, R_T, the blade length and C_P, the energy extraction coefficient (Doumi *et al 2016*).

For windmills, the energy extraction coefficient C_P, which depends of the wind velocity and the turbine is usually defined in the interval $(0,35 \div 0,59)$. The coefficient C_P is function of the specific velocity λ and the angle of the blade β. Figure2 shows the characteristic of C_P according λ (Razafinjaka and Andrianantenaina 2015, 2016)
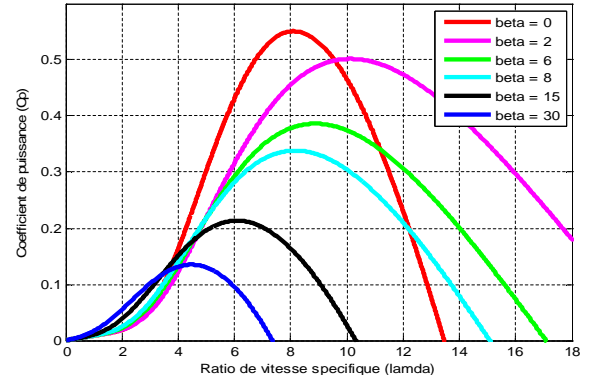
Figure 2: Turbine Power coefficient

The DFIG transforms the mechanical energy to electrical.

3- DFIG MODELING AND ITS VECTOR CONTROL

The DFIG model is described in the referential Park. The different equations below give the global modeling of the machine (Andrianantenaina and *al* 2015); Razafinjaka and Andrianantenaina 2016; Andrianantenaina and *al* 2016; Rouabhi and *al* 2015; Tarfaya and *al* 2015).

3.1. Electrical equations

$$\begin{cases} V_{sd} = R_s i_{sd} + \dfrac{d\phi_{sd}}{dt} - \omega_{coor}\phi_{sq} \\[2mm] V_{sq} = R_s i_{sq} + \dfrac{d\phi_{sq}}{dt} + \omega_{coor}\phi_{sd} \\[2mm] V_{rd} = R_r i_{rd} + \dfrac{d\phi_{rd}}{dt} - (\omega_{coor} - \omega)\phi_{rq} \\[2mm] V_{rq} = R_r i_{rq} + \dfrac{d\phi_{rq}}{dt} + (\omega_{coor} - \omega)\phi_{rd} \end{cases} \tag{2}$$

3.2. Magnetic equations

$$\begin{cases} \phi_{sd} = L_s.i_{sd} + M.i_{rd} \\ \phi_{sq} = L_s.i_{sq} + M.i_{rq} \\ \phi_{rd} = L_r.i_{rd} + M.i_{sd} \\ \phi_{rq} = L_r.i_{rq} + Mi_{sq} \end{cases} \quad (3)$$

3.3. Torque and power expressions

The electromagnetic torque is expressed according to current and fluxes by:

$$C_{em} = -p\frac{M}{L_s}(\phi_{sq}i_{rd} - \phi_{sd}i_{rq}) \quad (4)$$

With p, the number of pair poles

Stator active and reactive powers are expressed by:

$$P_s = V_{sd}I_{sd} + V_{sq}I_{sq} \quad (5)$$
$$Q_s = V_{sq}I_{sd} - V_{sd}I_{sq}$$

The rotor active and reactive powers are given by:

$$P_r = V_{rd}I_{rd} + V_{rq}I_{rq} \quad (6)$$
$$Q_r = V_{rq}I_{rd} - V_{rd}I_{rq}$$

3.4. DFIG vector control

In order to control the electricity production, a method, which not depends of active and reactive powers, is proposed. It consists to establish relations between rotor voltages delivered by the converter with active and reactive powers. Referential d-q related of spinning field and a stator flux aligned is adopted. So (Razafinjaka and Andrianantenaina 2015; Boualouch 2015; Doumi *et al* 2016)

$$\begin{cases} \phi_{sd} = \phi_s \\ \phi_{sq} = 0 \end{cases} \quad (7)$$

Flux equations become:

$$\begin{cases} \phi_{sd} = \phi_s = L_s.i_{sd} + M.i_{rd} \\ 0 = L_s.i_{sq} + M.i_{rq} \\ \phi_{rd} = L_r.i_{rd} + M.i_{sd} \\ \phi_{rq} = L_r.i_{rq} + M.i_{sq} \end{cases} \quad (8)$$

If the network is supposed stable, the stator flux is constant. Moreover, the stator resistor may be neglected; it is a realist hypothesis in the generator used in windmill. Taking into account all these considerations:

$$V_{sd} = 0 \quad (9)$$
$$V_{sq} = V_s = \omega_s\phi_s$$

By the equation (8), a relation between stator and rotor currents can be established:

$$i_{sd} = \frac{\phi_s}{L_s} - \frac{M.i_{rd}}{L_s} \quad (10)$$

$$i_{sq} = -\frac{M}{L_s}i_{rq}$$

Using simplifying hypothesis, the equations of powers give:

$$\begin{cases} P_s = -V_s.\frac{M}{L_s}i_{rq} \\ Q_s = -V_s.\frac{M}{L_s}i_{rd} + \frac{V_s^2}{L_s.\omega_s} \end{cases} \quad (11)$$

In order to control the generator, expressions showing the relation between rotor voltages and rotor currents are:

$$\begin{cases} V_{rd} = R_r i_{rd} + L_r\sigma\frac{di_{rd}}{dt} - g\omega_s L_r\sigma i_{rq} \\ V_{rq} = R_r i_{rq} + L_r\sigma\frac{di_{rq}}{dt} + g\omega_s\left(L_r\sigma i_{rd} + \frac{MV_s}{\omega_s L_s}\right) \end{cases} \quad (12)$$

Where g and σ denote respectively the slip and the leakage coefficient.

Fig.3 built by relations (11), (12), (13) and (14) shows diagram where rotor voltages are the input and active and reactive powers are the output.
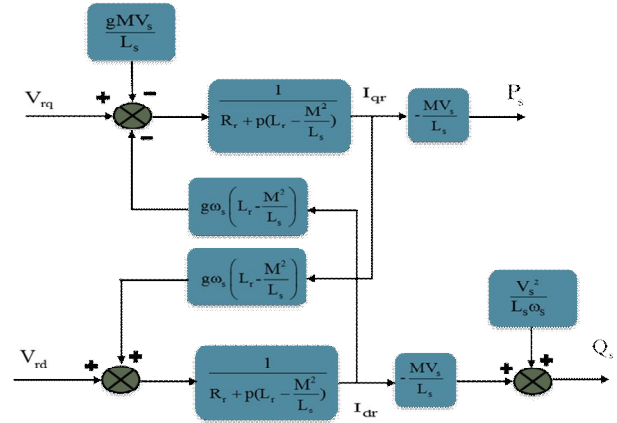
Figure 3: Block Diagram of simplified DFIG model

4- THE HYBRID CONTROLLER

This new topology of hybrid controller named Fuzzy-VGPI is built on combination of the variable gain PI controller and the fuzzy logic one. First, the general characteristics of the VGPI controller is showed and followed by the presentation of the fuzzy logic controller. Based on these two types of controllers, the technique of hybridization will be studied.

4.1- VGPI Controller Structure

The use of PI controllers to command a DFIG is often characterized by an overshoot in tracking mode and a poor load disturbance rejection. This is mainly caused by the fact that the gains of the controller cannot be set to solve the overshoot and load disturbance rejection problems simultaneously. Overshoot elimination setting will cause a poor load disturbance rejection and rapid load disturbance rejection setting will cause important overshoot or even instability for the system (Chikouche 2013; Miloudi 2007)

To overcome this problem, the use of VGPI controllers is proposed. A VGPI controller is a generalization of the

classical PI controller where the proportional and integrator gains vary along a tuning curve. Each gain of the proposed controller has four tuning parameters (Chikouche 2013; Miloudi and Draou 2005; Miloudi 2007; Shreyash Vir and Sarika Kalra 2016):

- Initial gain value or start up setting which permits overshoot elimination.
- Final gain value or steady state mode setting which permits rapid load disturbance rejection.
- Gain transient mode function which is a polynomial curve that joins the gain initial value to the gain final value.
- Saturation time which is the time at which the gain reaches its final value.

The entire number n of the gain transient mode polynomial function is defined as the degree or order of the variable gain PI controller.
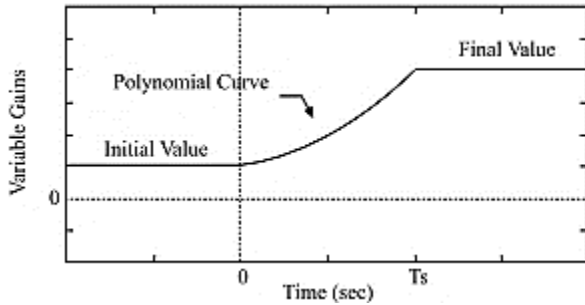
Figure 4: Variable PI Gains Tuning Curve

If e(t) is the signal input to the VGPI controller the output y(t) is given by (Chikouche 2013; Miloudi and Draou 2005; Miloudi 2007; Shreyash Vir and Sarika Kalra 2016):

$$y(t) = K_p e(t) + \int_0^t K_i e(t) dt \qquad (13)$$

With

$$K_p = \begin{cases} (K_{pf} - K_{pi})(\dfrac{t}{t_s})^n + K_{pi} & \text{if } t < t_s \\ K_{pf} & \text{if } t \geq t_s \end{cases} \qquad (14)$$

$$K_i = \begin{cases} K_{if}(\dfrac{t}{t_s})^n & \text{if } t < t_s \\ K_{if} & \text{if } t \geq t_s \end{cases} \qquad (15)$$

Where K_{pi} and K_{pf} are the initial and final values of the proportional gain K_P and K_{if} is the final value of the integrator gain K_i. The initial value of K_i is taken to be zero. It is noted that a classic PI controller is a VGPI controller of degree zero.

The VGPI controller in vector control of DFIG is used as presented in Figure 5.
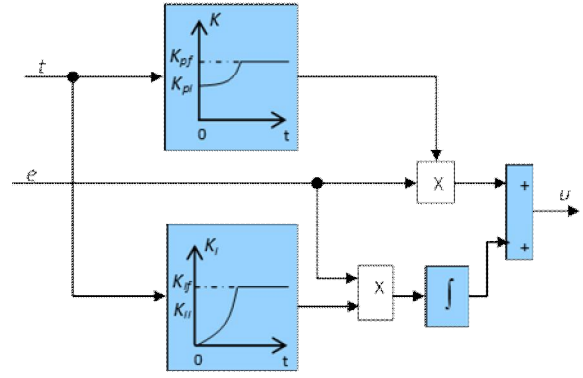
Figure 5: The Structure of VGPI Controller

For the VGPI synthesis, a generalized method using by (Andrianantenaina and *al* 2015, 2016) is chosen to determine the parameters of the classical PI controller. The gains of classical PI are taken to be the terminal values of the VGPI controller. These conditions are adopted: n =1, T_S=0.1[s]

4.2- Fuzzy logic controller

The method built around fuzzy logic avoids modeling the system but it is clear that having knowledge of its behavior is always useful. The reasoning is close to human perception. Nowadays, the fuzzy logic controller begins to take an important place in electrical applications. It can be used for optimization, and command (Andrianantenaina and *al* 2015; Razafinjaka and Andrianantenaina 2015, 2016). Figure 6 gives the common scheme for fuzzy logic controller.
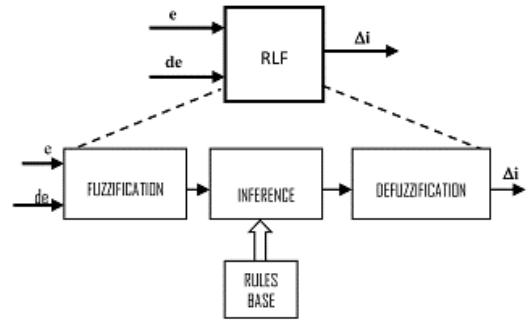
Figure 6: Structure of a fuzzy controller

With e, de and Δi denote respectively the error, the error variation and the output. The fuzzification consists in projecting a real physical variable distributed on the domains variable characterizing this variable: linguistic variable is so obtained and the fuzzification makes it possible to have a precise measurement by the membership degree of the real variable to each fuzzy subset. Generally, the inference method is a logical operation by which one admits a proposal under the terms of its relation with other proposals held for true. At this stage, rules are established by the knowledge of the desired behavior of the system.

They are often as:

(If x_1 is A) AND (x_2 is B) THEN $S_k = C_k$ (16)

Here x_1 and x_2 are the inputs and S_k the output which is also a linguistic variable. Membership functions may be defined for the output variable and there are several

inference methods, which may be applied. The results of aggregation of the inference rules give still fuzzy variables. To be used in a real control, these fuzzy variables must be translated into real or numerical variables: it is the function of the defuzzification block.

In this paper, the Sugeno's methods are chosen: a singleton is used as the membership function of the rule consequent combined by max-min method for the rule evaluation. Thus, in relation (16), C_k is a constant. The Sugeno defuzzification is then weighted average method.

$$s = \sum \frac{\mu(s_k).s_k}{\mu(s_k)} \qquad (17)$$

For the two inputs (e, de), the triangular and trapezoidal forms are used (Andrianantenaina and *al* 2015; Razafinjaka and Andrianantenaina 2015, 2016). The number of membership functions may be N=3, 5 or 7. Here N=3 is adopted. The output uses the singletons as membership function.
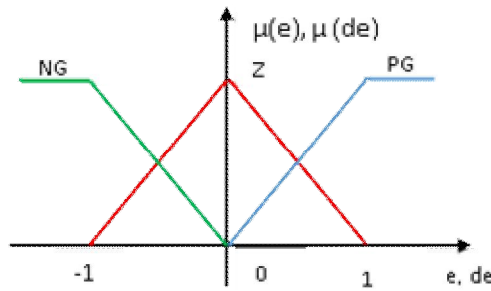
Figure 7: inputs Membership functions
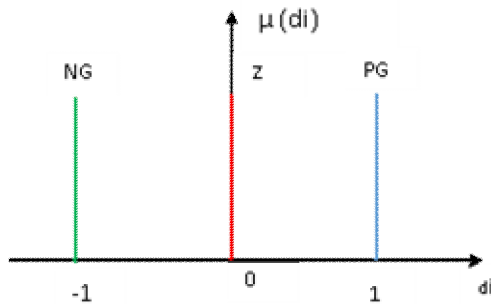
Figure 8: Output Membership functions

Table 1 gives the inference matrix.

Table 1: Rules base for N = 3

		E		
		NG	Z	PG
de	NG	NG	NG	0
	Z	NG	0	PG
	PG	0	PG	PG

The table 1 gives 9 rules. For example,

R$_1$: (IF e = NG) AND et (de = NG) THEN Δi = NG

4.3- Hybrid controller: FUZZY-VGPI

The hybrid controller is built on combination of these two topologies. Figure 9 shows the block diagram of the fuzzy-VGPI controller system.
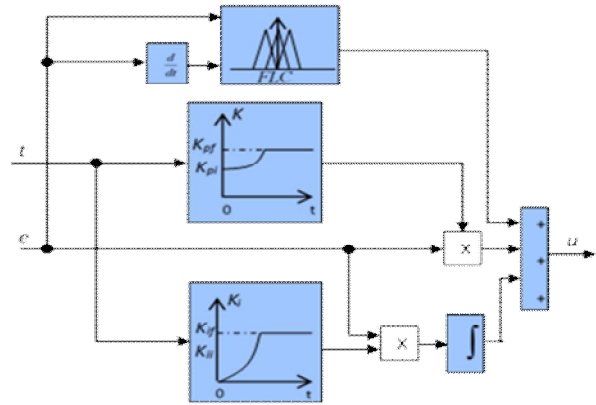
Figure 9: Basic scheme of proposed hybrid controller: FUZZY-VGPI

There are generally two used methods for controlling the independent regulation of active and reactive powers of the DFIG (Andrianantenaina and *al* 2015): the direct method, which consists to neglect the coupling terms and to put a controller on each axe to control active and reactive powers. In this case, the controllers command directly the rotor voltages of the machine. The second method takes into account the coupling terms and compensates them by using two loops that permit to control the powers and the rotor currents: it is called the indirect command and is based on the relations (11) and (12).

The block diagram with VGPI and FUZZY-VGPI is given by figure 10.
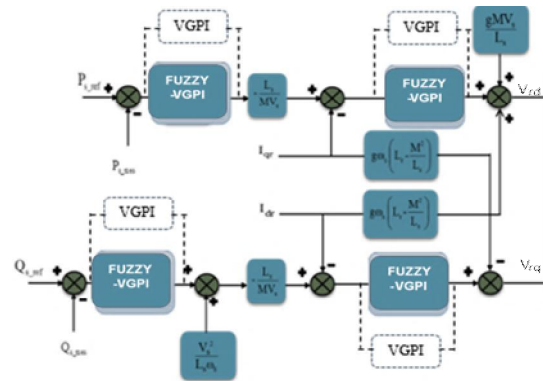
Figure 10: Block diagram of the vector control of the DFIG with VGPI and FUZZY-VGPI

5- SIMULATION AND RESULTS

The proposed method has been tested by tracking, disturbance rejection and robustness following speed and parameters variations specially the rotor resistance. These conditions are so adopted:

▪ For the reference Tracking:
Variation of the active power reference P$_{ref}$ and the reactive power reference Q$_{ref}$.

▪ For the Robustness Test:
We varied the speed,1450 rpm to 1600 rpm at t = 4,5 [s], and. the rotor resistance Rr to 2*Rr at t = 5 [s]
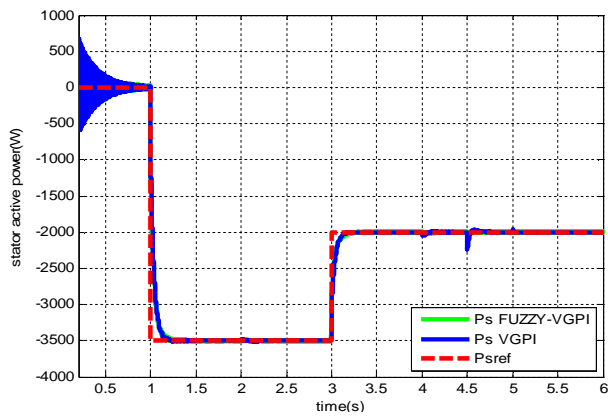
Figure 11: Active power curves
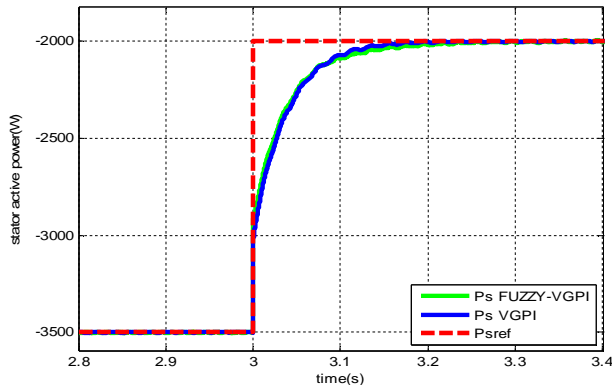
Figure 12: zoom on active power tracking
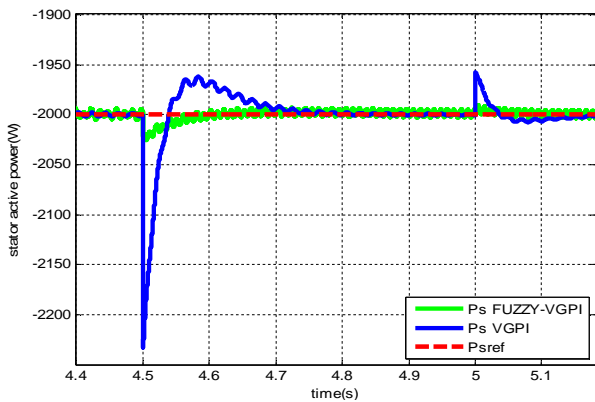
Figure 13 : Zoom on active power curve following speed and rotor resistance variations
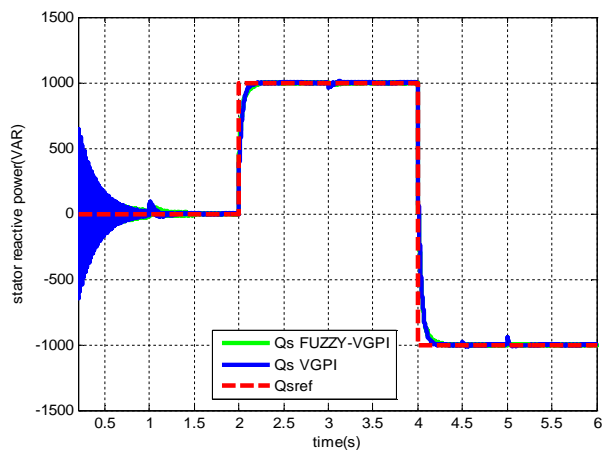
Figure 14: reactive power curves
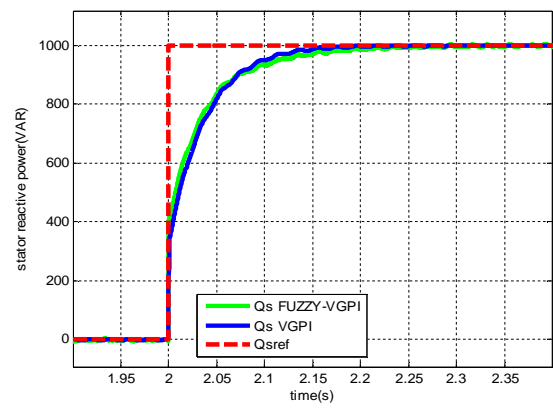
Figure 15: zoom on reactive power tracking
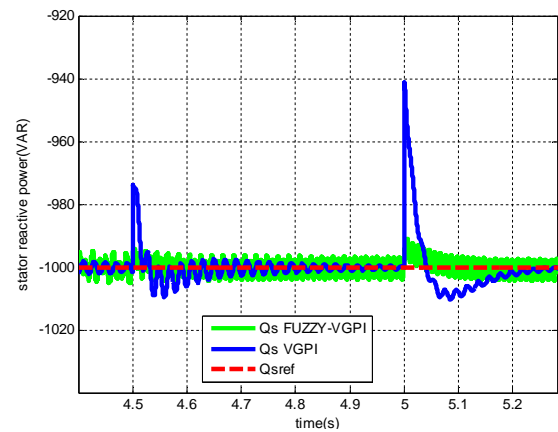
Figure 16: Zoom on reactive power curve following speed and rotor resistance variation
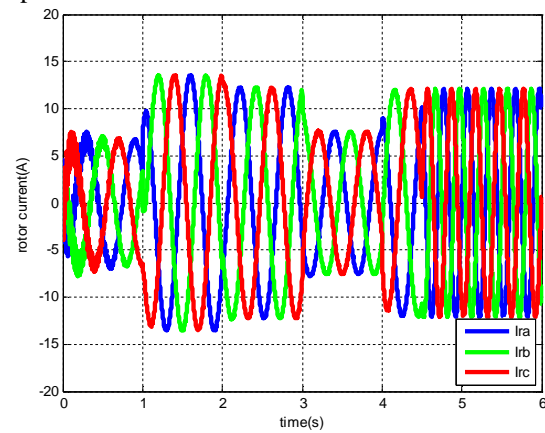
Figure 17 : Rotor currents
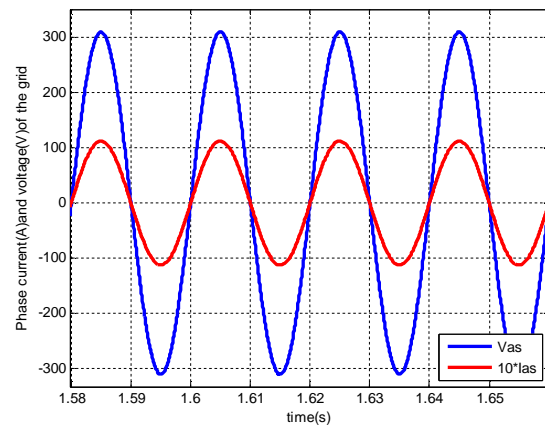
Figure 18: Stator voltage and current when Qs = 0

Table 2: Simulation results

Tests	VGPI n = 1, Ts=0,1(s)	FUZZY-VGPI n = 1, Ts=0,1(s)
Transient Behavior	$D_1 = 0\%$ $t_M =0,2$ [s]	$D_1 = 0\%$ $t_M =0,2$ [s]
t = 4,5 [s] $\Omega \uparrow$	$\Delta P = 230$ [W] $\Delta t = 0,25$ [s]	$\Delta P = 25$ [W] $\Delta t = 0,1$ [s]
t = 5 [s] $R_r \uparrow$	$\Delta P = 42$ [W] $\Delta t = 0,2$ [s]	$\Delta P = 10$ [W] $\Delta t = 0,05$ [s]

Here t_M denotes the settling time or the duration, which the output reaches the steady state when the set point change is applied.

By Figure 11 and Figure 14, we can notice that the power references are well followed by the generator.

The negative sign of the active power shows that the generator injects the energy into the grid and the negative sign of the reactive power functions in capacitive mode, for inductive mode the power becomes automatically positive.

By Figure 12 and Figure 15, we can conclude that the both controllers give the same performance on tracking test with a quicker response and without overshoots.

For the robustness test, Figure 13 and Figure 16 show the effect of varying the parameters of the generator. Based on these results, we found that the fuzzy-VGPI is more robust and has better performance than VGPI one.

When Qs = 0, statorique voltage and statorique current are purely sinusoidal and in phase (Figure 18). It ensures a good quality of the energy injected into the grid.

6- CONCLUSION

In this paper, a new topology to obtain a hybrid controller is proposed to be applied on DFIG used in WECS. The hybridization consists to combine a variable gain PI controller and fuzzy logic controller. The comparative study is made between VGPI controller and Fuzzy-VGPI one. The simulation results highlight that Fuzzy-VGPI gives better performances on disturbance rejection and robustness in respect of the speed and parameters variation especially with rotorique resistance variation.

REFERENCES

Tsiory P. Andrianantenaina, Jean N. Razafinjaka Hervé Mangel, 2015. Controls of powers generated by a Doubly-Fed Induction Generator used in a chain of wind power conversion. The 5th International Conference of Energy, Environment and Climate Change (ICEECC 2015). July 8-9, Mauritius.

Jean N. Razafinjaka, Tsiory Patrick Andrianantenaina, 2015. Hybrid controller for wind turbine based on doubly fed induction generator. **The** 7th International conference on Fuzzy Logic Theory and Application (FCTA). November 12-14, Lisbon (Portugal)

Jean N. Razafinjaka, Tsiory Patrick Andrianantenaina, 2016. Model Reference Adaptive Control based on Fuzzy Logic for Doubly Fed Induction Generator used in a Chain of Wind Power Conversion. International Journal of Engineering Research and Management (IJERM). February , Rajasthan , India

A. Boualouch, A. Essadki, T. Nasser, A. Boukhriss and A. Frigui, 2015. Power Control of DFIG in WECS Using Backstepping and Sliding Mode Controller. International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. Vol:9, N°6.

M'hamed Doumi *et al*, **2016.** Nonlinear Integral Backstepping Control of Wind Energy Conversion System Based on a Double-Fed Induction Generator PRZEGLĄD ELEKTRO TECHNICZNY, ISSN 0033-2097, R. 92 NR 3

Tsiory P. Andrianantenaina, Jean N. Razafinjaka and Andrianirina Charles B., 2016. Commande par PI à gain variable d'un Système Eolien à Base d'une Génératrice Asynchrone à Double Alimentation**.** Conférence internationale **sur le plasma et énergie (CIPE), Décembre 24-26, Antsiranana, Madagascar.**

Rouabhi Riyadh, Abdessemed Rachid, Chouder Aissa and Djerioui Ali, 2015. Hybrid Backstepping control of a doubly fed wind energy induction generator**.** Mediterranean journal of measurement and control, September.

Anis Tarfaya, Djalel Dib and Mehdi Ouada, 2015. Study Contribution to Control Optimization of a Wind Turbine based on a DFIG**.** International Conference on Mechanical and Industrial Engineering (ICMAIE), June 10-11, Antalya, Turkey

T. Mohammed Chikouche, A. Mezouar, T. Terras and S. Hadjeri, 2013. **Variable Gain PI Controller Design for Speed Control of** a **Doubly Fed Induction Motor.** Engineering, Technology & Applied Science Research (ETASR), Vol. 3, No. 3, 433-439

A. Miloudi, A. Draou, **2005.** A **novel Variable Gain PI Controller used for Speed Control of** a **direct torque neuro fuzzy controlled Induction Motor** ACTA ELECTROTEHNICA, Vol. 46, No. 1

A. Miloudi, Eid A. Al-Radadi and A. D. Draou, 2007. A **novel Variable Gain PI Controller used for Speed Control of** a **direct torque neuro fuzzy controlled Induction Machine Drive.** Turk J Elec Engin, Vol. 15, No. 1

Shreyash Vir, Sarika Kalra, 2016. **Simulation and Modeling of Sinusoidal Pulse Width Modulated Inverter Fed Induction Motor Using PI Controller.** IJESC, Vol. 6, No. 5

A RISK MANAGEMENT MODEL FOR MEASURING PROJECT COMPLEXITY

De Felice, F.[a], Petrillo, A.[b], Zomparelli, F.[c], Esposito M.[d]

[a],[c]Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio,
Via G. di Biasio 43, 030434 – Cassino (FR) – Italy
[b]Department of Engineering, University of Naples "Parthenope", Centro Direzionale,
Isola C, 480143 – Napoli (NA) – Italy
[d]ABC Via Argine, 929, 80147 – Napoli - Italy

[a]defelice@unicas.it, [b]antonella.petrillo@uniparthenope.it, [c]f.zomparelli@unicas.it, [d]marcovik@libero.it

ABSTRACT

Any project is subjected of high risk, which can lead to a project failure and economical loss. Thus, risk management analysis is a key role in project management. Several research models have been developed to analyze and manage project risks. The aim of this work is to develop a multi-criteria decision support system to define and evaluate risks in a project development. More specifically, the purpose of the research is to define a risk management model that combines AHP methodology and project management approach to measure complexity and riskiness of a project. The model analyzes a real case study concerning the construction of a new industrial plant.

Keywords: risk management, project management, AHP, decision support system

1. INTRODUCTION

Today, project management is strongly linked to the risk analysis. It is necessary to identify and assess the major risks that could cause the failure of the project (McComb and Smith, 1991). Risk is any event that has a probability of realization: if the effect risk is negative, it is a threat, while it is an opportunity (Remenyi, 1999). Risk may affect the achievement of objectives. During the last years, academic literature has highlighted the importance of risk management during a project development (Nguyen *et al.*, 2017). Ward and Chapman (2003), outline how project risk management processes might be modified to facilitate an uncertainty management perspective. Raz and Michael (2001) find which tools are more likely to be used in the project management and those to be used for the contribution of risk management processes. International Standard Organization (ISO) creates a working group to guide the development of risk management. This working group develops ISO 31000:2009 "Risk Management – Principles and Guidelines" which represents the views and the experience of hundred of knowledgeable people involved in risk management (Purdy, 2010). Project management tools related to the decision-making models allow to identify and manage project risk factors (Marques *et al.*, 2011). In this context the present paper tries to develop a new integrated model based on project management and AHP methodology to evaluate the project riskiness according to ISO31000:2009 standard . The model is implemented in a real case concerning the construction of a new industrial plant. Firstly, the project is decomposed into elementary work packages through the use of work breakdown structure. Secondly, project risks are identified using the "Ishikawa" diagram (cause and effect diagram) that defines causes (risks) that produce an effect (project failure). Risks are assessed through the implementation of Analytic Hierarchy Process (AHP) that allows to define a project risk ranking. Finally, the model proposes two performance indices to measure risk and complexity of the project. These two indices allow to the project manager to develop optimal planning strategies to complete project within certain constraints. The rest of the paper is organized as follows. In section 2 a literature review on risk project management and multi criteria approach is presented. Section 3 describes the proposed model and a case study is analyzed. Section 4 presents discussion. Finally, in section 5, conclusion are analyzed.

2. LITERATURE REVIEW

Risk management is the methodology that improves the project success probability (Olechowski *et al.*, 2016). Several studies show that project risks can affect industrial performance. For example Mishra et al., (2016), have collected 82 federal technology projects across 519 quarterly time periods. The research highlights that each of the three types of risks has a significant negative effect on project performance. While Wallace and Keil (2004), have explored how different types of risk influence both process and product outcomes in software development projects by analyzing input from more than 500 software project managers representing multiple industries. It is evident that risk management is a topic that covers many fields of application. It is necessary to manage critically the risk management practices, because they often allow the project realization. In fact, Oehmen *et al.*, (2014) examined 291 projects, showing that more than 70% of management practices are not relative to product or

process, but they depend by risk management. Risk management is tackled through the most advanced project management tools. A standardized model for project management may increase project success (Milosevic and Patanakul, 2005). The growing importance of risk management issues led to the creation of a standardized norm for risk management: "ISO 31000:2009". The norm aims to standardize the effects of risk management to better address the effects of uncertainty in project management. The standard emphasizes the development of risk management in organizations. The framework ensures that the process is supported in an iterative and effective way. Thus, risk management becomes a key component in the company's governance (Gjerdrum and Peter, 2011). ISO31000:2009 requires risk identification and assessment. Risk assessment is a very complex task, and to manage it, often it is necessary to make a choice. For this reason we rely on multi-criteria decision-making systems that allow to make the best possible choice (De Felice and Petrillo, 2013). Among the multi criteria model, one of the most popular is the AHP, used to solve complex decision problems and introduced by Saaty (1977). AHP model breaks down a problem into several levels forming a hierarchy with a unidirectional relationship between levels. AHP model is very flexible and it can be used in different fields. De Felice et al., (2016) use the AHP model to define a key performance indicators for safety management. AHP model is widely used for the risk management analysis. Da Silva and Camanho, (2015) use the AHP model to define an IT project priorization in a Brazilian multinational company in the oil & gas segment. The research identifies the influence of AHP methods in each of several variables which represent the IT project.

Yassine and Brahim (2016) propose a supply chain risk management framework that identifies and monitors risks. AHP model tries to assess factors which influence risk management. Zhang *et al.* (2015), define a hierarchy level to analyze risk management in a construction project. The model also analyzes the economic losses from any project failures. Mustafa and Al-Bahar (1991), analyze construction project considering: time, budget and quality. They use the AHP model to assess the riskiness constructing of a bridge in Bangladesh. Grekul, *et al.* (2015) try to decrease the subjectivity of decisions made during the risk management process. AHP is used for risk assessment. Dey (2012) uses AHP and decision tree analysis for risk assessment in an Indian oil refinery.

3. METHODOLOGICAL APPROACH

According to several studies, the risk analysis plays a key role in the project of a new system. The risk analysis allows the project manager to allocate efficiently resources to minimize future problems. The research develops an integrated model of project-risk analysis through the AHP model application. The study follows the steps of ISO31000 for risk management. The aim of this paper is to demonstrate the application

of AHP to assess the risk factors of a project and it define an overall index that quantifies risk. Figure 1 shows the research framework.
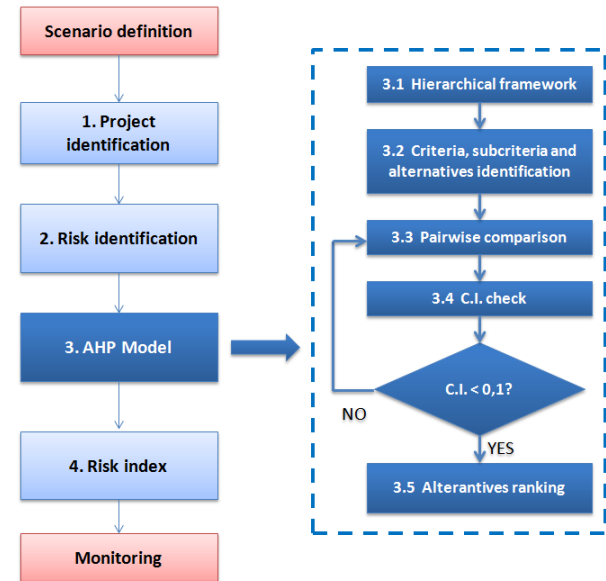
Figure 1:Research framework

3.1. Scenario definition
The analyzed case study defines a project management for new industrial plant construction for a medium enterprise in a central area of Italy. The study follows the steps of ISO31000:2009 about risk management. The considered model is subject to high variability depending on many factors, among them, some of the most important are:

- time;
- climate;
- forecast uncertainty.

3.2. Project identification: Phase#1
Project involves construction of three main elements:

- infrastructures;
- mechanical plants;
- electrical plants.

Customer wants to reduce time construction as much as possible. Project description includes work breakdown structure (WBS) to identify: activity, sub-activity and work packages easier to manage (Tausworthe, 1979). WBS of project (Figure 2) is divided into three hierarchies. Obviously, for each element represented in Figure 2, there will be another WBS more specific, to describe even the most basic activities.

3.3. Risk identification: Phase#2
ISO31000:2009 defines risk identification as the most important process of risk management. Risk identification requires a brainstorming work by a selected expert group (Winch, 2002). In this case the

expert group includes: project manager, architect, mechanical engineer and electrical engineer.

Experts are chosen based on their knowledge and based on previous experience in the construction industry.

Risk identification can be supported by a valuable tool: Ishikawa diagram (Hishikawa, 1986). This tool was created for qualitative analysis, but it is used to identify general problems. Considering work packages of WBS, the expert group has identified the associated risks. Figure 3 shows the Ishikawa diagram, which represents the main risks associated with the project failure.

3.4. AHP model: Phase#3

AHP model was developed through a three-level hierarchy. The top level of the hierarchy is the main goal of the decision problem. The lower levels are criteria and sub-criteria that contribute to the goal. Finally, there are alternatives of the model. In this case criteria and subcriteria are risk factors identified in phase #2, while alternatives are the work packages

identified in phase#1. The design of hierarchy required experience and knowledge of the specific problem. It is necessary to define a team of experts which consists in 1 project management, 1 mechanical engineer, 1 electrical engineer and 1 architect. Figure 4 shows the hierarchy of the problem under study. The model prioritizes alternatives, based on their failure probability according to the criteria and sub-criteria analysis.

The model is divided into three different steps:

- pairwise comparison and relative weight estimation considering: criteria, sub criteria and alternatives;
- priority weight vector calculation;
- consistency index estimation to verify the accuracy of the judgments.
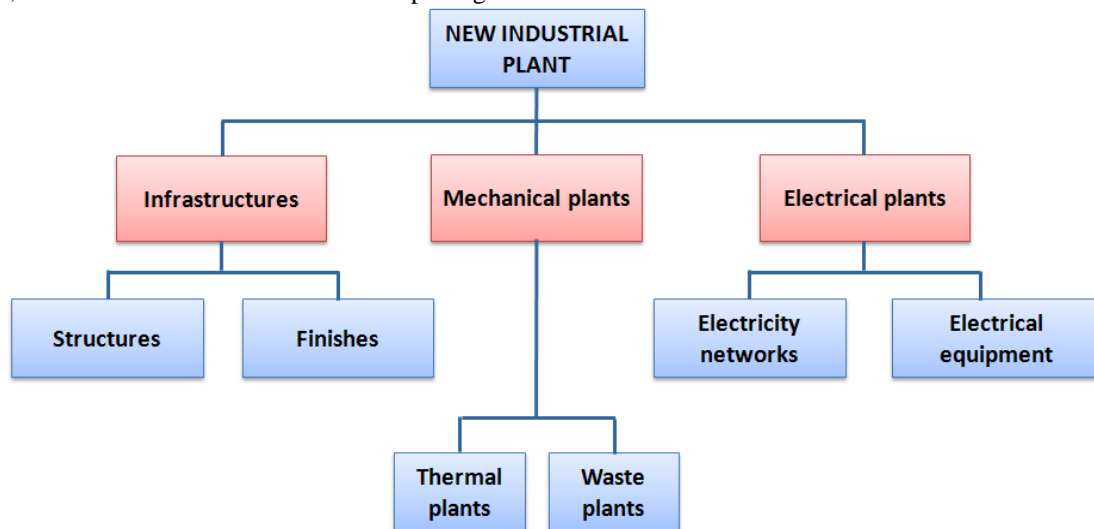
Figure 2: Work breakdown structure of the project
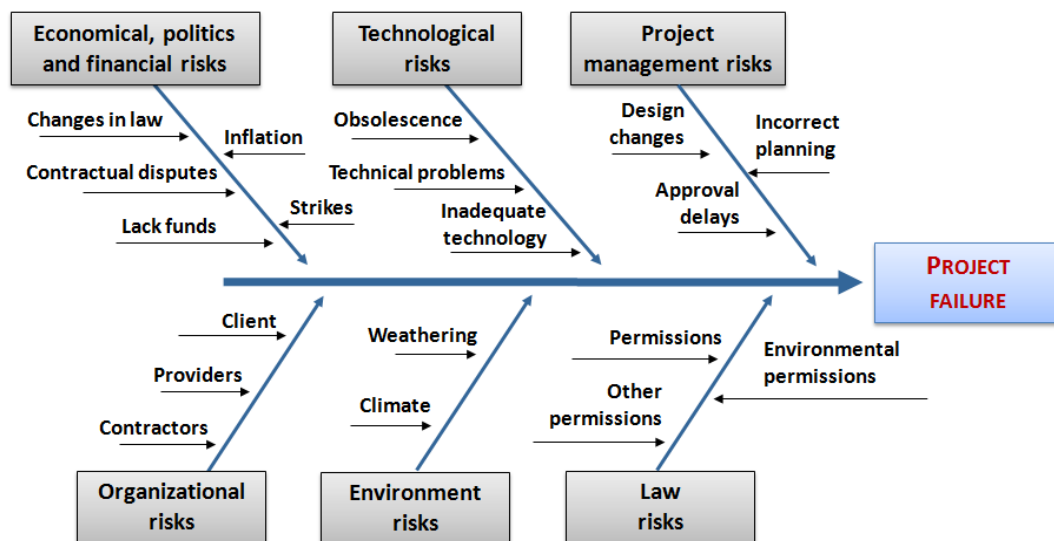
Figure 3: Ishikawa diagram
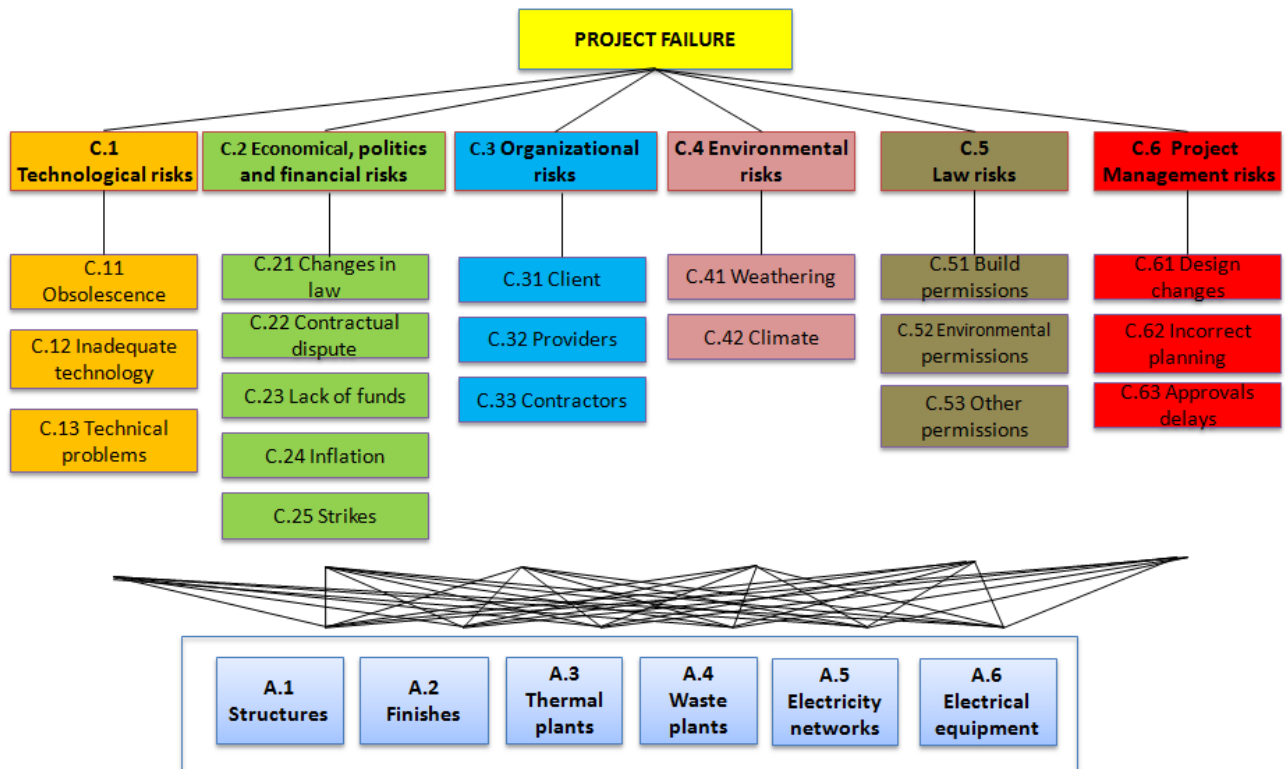
Figure 4: AHP risk management Model

After hierarchy definition, the pairwise comparison matrices were developed to determine criteria, sub criteria and alternatives weights. Table 1 shows a parwise comparison of technological risks sub-criteria. Values in the matrix are the arithmetic average of the ratings according four experts.

Table 1: Pairwise comparison–technological sub criteria

Sub criteria comparison				
	C.11	C.12	C.13	Priority Vector
C.11	1	1.429	0.669	0.327
C.12	0.700	1	1.000	0.295
C.13	1.495	1.000	1.000	0.378

In this case vector identifies "Technical problems" as the most important sub-criteria in technological risks with a score of 37.8%. Figure 5 shows a summary of priorities for criteria and subcriteria. The most important criteria are C.1 (Technological risks) with a score of 22.6%; C.6 (Project management risks) with 19% etc.

The most important subcriteria are C.31 (client) with a score of 0.564 and C.61 (design changes) with 0.512. Table 2 shows the consistency index obtained from all comparison matrices of criteria and subcriteria. Consistency index is acceptable if is less than 0.10. Table 2 shows that all judgments are consistent.

Table 2: Consistency index – Criteria and sub criteria

Consistency index	
Comparison	C.I
Criteria	0.023
Technological subcriteria	0.033
Economical subcriteria	0.082
Organizational subcriteria	0.048
Environmental subcriteria	0
Law subcriteria	0.015
Project management subcriteria	0.039
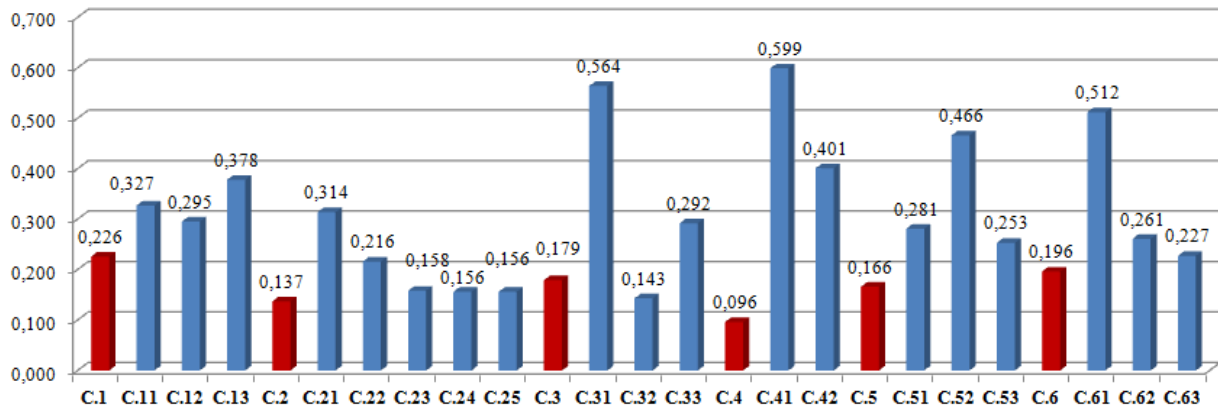
Figure 5: Priorities for criteria and sub criteria

Following, alternatives are compared with sub-criteria. Table 3 shows the comparison of alternatives with the sub-criterion C.11. Values in the matrix are the arithmetic average of the ratings according four experts.

Table 3: Pairwise comparison – alternatives with C.11

Alternatives comparison							
C.11	A.1	A.2	A.3	A.4	A.5	A.6	Priority
A.1	1	3.364	2.632	0.904	0.816	0.760	0.224
A.2	0.297	1	1.732	1.000	0.904	0.841	0.142
A.3	0.380	0.577	1	1.189	1.000	1.414	0.146
A.4	1.107	1.000	0.841	1	1.316	1.000	0.164
A.5	1.225	1.107	1.000	0.760	1	0.707	0.150
A.6	1.316	1.189	0.707	1.000	1.414	1	0.174

Finally, alternatives are compared with the goal (Table 4). Values in the matrix are the arithmetic average of the ratings according the four experts.

Table 4: Pairwise comparison – alternatives with goal

Alternatives comparison							
GOAL	A.1	A.2	A.3	A.4	A.5	A.6	Priority
A.1	0.156	0.217	0.160	0.268	0.171	0.155	0.188
A.2	0.096	0.116	0.123	0.161	0.171	0.107	0.129
A.3	0.104	0.097	0.099	0.080	0.079	0.128	0.098
A.4	0.202	0.261	0.296	0.239	0.300	0.273	0.262
A.5	0.208	0.116	0.222	0.139	0.171	0.209	0.178
A.6	0.234	0.193	0.099	0.113	0.107	0.128	0.146

The comparison of alternatives, defines some CI index, which are less to 0.1. So judgments are consistent. Finally, the principle of hierarchical composition, identifies the global priority of alternatives (Table 5).

Table 5: Global priority of alternatives

Priorities	
Alternatives	Global priority
A.1	27.22%
A.2	18.45%
A.4	16.44%
A.5	13.04%
A.3	12.43%
A.6	12.42%

Considering goal, criteria and sub-criteria the alternative A.1 (structures) is the most critical element in the construction of new industrial plant, with a score of 27.22%. Figure 6 shows the ranking of alternatives.
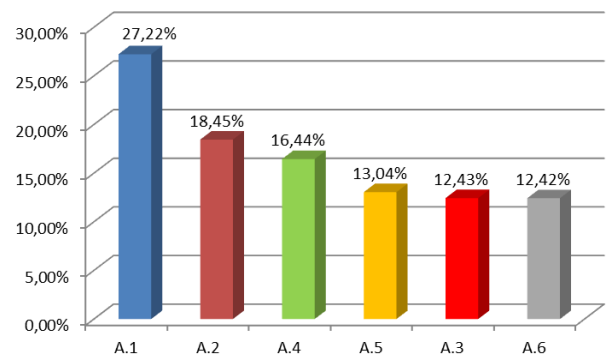
Figure 6: Priorities for alternatives

3.5. Risk index: Phase#4
According to the above results, two risk indices were identified in order to evaluate two fundamental elements in a project:

- complexity level;
- risk level.

Project complexity is its distinctive feature. According to Vidal et al., (2010), project complexity can be measured through an index defined as follows:

$$CI_i = S(i) / \max [S(i)] \qquad (1)$$

where:

- CI is the complexity index;
- i is the number of alternatives;
- S(i) is the priority of alternative "i";

$$0 \leq CI_i \leq 1 \qquad (2)$$

The complexity index allows to identify the critical aspects of the project to optimally manage them. Table 6 shows the critical indices for the analyzed project. While Figure 7 presents data with a color complexity scale.

Table 6: Critical index

Critical index		
Alternative	Priority	CI_i
A.1	27.22%	1.000
A.2	18.45%	0.678
A.4	16.44%	0.604
A.5	13.04%	0.479
A.3	12.43%	0.456
A.6	12.42%	0.456
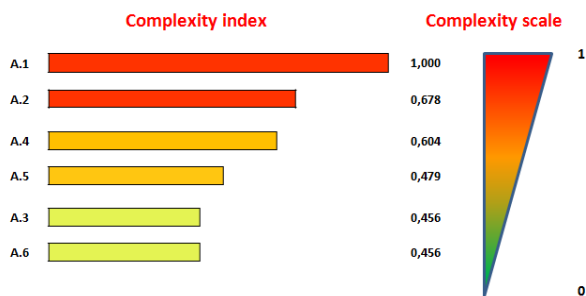
Figure 7: Complexity index

According Zayes *et al.*, (2007), it is possible to calculate a risk rating of a project, using a synthetic number as follows:

$$R = \sum W_i * V_i (x_i) \qquad (3)$$

Where:

- R is risk index;
- Wi is risk priority;
- xi is the risk;
- Vi is the impact of each risk on the project.

$$0 \leq R \leq 1 \qquad (4)$$

Wi are the global priority of subcriteria obtained with AHP model. Vi(xi) is calculated as an average of the opinions expressed by four decision makers (Table 7).

Table 7: Riskiness index

Riskiness index			
	W_i	$V_i(x_i)$	Wi * Vi (xi)
C.11	0.074	0.36	0.02664
C.12	0.067	0.38	0.02546
C.13	0.086	0.28	0.02408
C.21	0.043	0.21	0.00903
C.22	0.030	0.30	0.009
C.23	0.022	0.20	0.0044
C.24	0.021	0.10	0.0021
C.25	0.021	0.18	0.00378
C.31	0.101	0.41	0.04141
C.33	0.052	0.33	0.01716
C.32	0.026	0.36	0.00936
C.41	0.057	0.24	0.01368
C.42	0.038	0.23	0.00874
C.52	0.077	0.29	0.02233
C.51	0.047	0.38	0.01786
C.53	0.042	0.19	0.00798
C.63	0.044	0.29	0.01276
C.61	0.100	0.41	0.041
C.62	0.051	0.36	0.01836
		TOT	0.31

Figure 8 shows riskiness value for each sub-criterion.
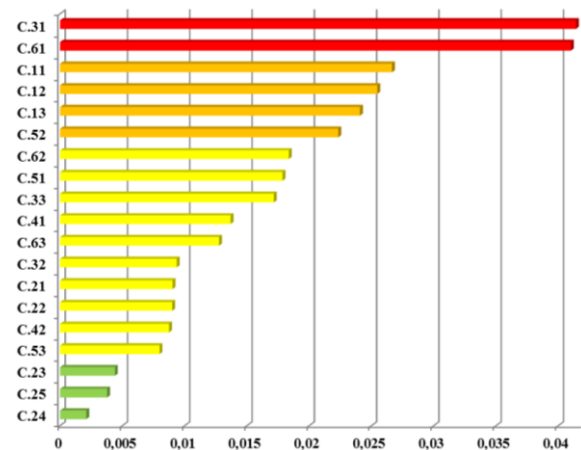
Figure 8: Risk index

The total project risk index is 0.31. Zayes et al., (2007) propose a scale for the risk analysis:

- $0 \leq R \leq 0.3$ low risk;
- $0{,}3 < R \leq 0.6$ medium risk;
- $0{,}6 < R \leq 0.8$ high risk;
- $R > 0.8$ very high risk.

In this case the analyzed project is subject to a low level of overall risk. Two most hazardous sub-criteria are C.31 (Customer) with a score of 0.04 and C.61 (Design changes) with a score of 0.041

3.6. Monitoring

The analysis defines a set of risk indicators. Through this ranking is possible to continuously monitor several critical processes of the project to avoid inefficiencies. If the system changes, it is possible to repeat the analysis to identify new critical issues of the project. The continuous control is a fundamental for the project successful.

4. DISCUSSION

The application of this analytical model allows to analyze the project of industrial plant construction and to calculate the complexity and the riskiness index of the project. The AHP model has identified an important ranking on criteria, sub-criteria and finally alternatives. AHP analysis highlights that "structures" (A.1) is the most critical system related to the project failure with a score of 27.22%, while the less critical element is "electrical equipment" (A.6) with a score of 12.22%. The analysis of consistency index (CI) showed uniformity of the judgments of the four experts. The consistency index is always less than 0,10. The complexity index has traced the results obtained with AHP model defining alternatives A.1 and A.2 the most critical. It will therefore be necessary to invest more resources in these activities because they have a higher index of complexity than the other. Finally the model calculates an overall risk ratio with a score of 0.31. Comparing it with scale proposed by Zayes et., (2007), the project is evaluated as "low risk".

5. CONCLUSION

The necessity to control and monitor risks during the development of a project, pushes research towards increasingly sophisticated and accurate analysis. The presented research, allowed to identify and quantify project risks, using an integrated model that uses different tools of project management, such as: work breakdown structure, Ishikawa diagram, AHP model to define two indices that summarize the attribute of the project. It is important to emphasize the model validity. Its application allows to analyze and identify risks during the implementation of a project, defining a set of indicators to identify criticality of the project and to develop appropriate strategies for improvement. The presented model is a decision support system useful to the decision maker in the design phase of the project. The analyzed case study is the project of the development of a new production plant. The results presented show that the model is a valuable diagnostic tool that allows to identify the highest risks in a project and to evaluate them, supported through an AHP model run by a group of four experts. The two proposed indices allow us to identify complexity and riskiness of the project. The advantages of the model are:

- It provides a structured model for risk analysis;
- It allows to define optimum strategies for resource allocation;
- It reduces project delays due to unforeseen events;
- It is a flexible model applicable in many fields.

Future work will focus on the possibility to define a financial and economical profile, to assess the loss of costs related to the risk management during the project planning.

REFERENCES

Da Silva Neves, A. J., Camanho, R., 2015. The use of AHP for IT project priorization–a case study for oil & gas company. Procedia Computer Science, 55, 1097-1105.

De Felice, F., Petrillo, A., 2013. Absolute measurement with analytic hierarchy process: a case study for Italian racecourse. International Journal of Applied Decision Sciences, 16(3), 209-227.

De Felice, F., Petrillo, A., Di Salvo, B., Zomparelli, F., 2016. Prioritising the safety management elements through AHP model and key performance indicators. In Proceedings of 15th International Conference on Modeling and Applied Simulation, MAS 2016 Larnaca; Cyprus; 26 September - 28 September 201, pp. 49-56.

Dey, P. K., 2012. Project risk management using multiple criteria decision-making technique and decision tree analysis: a case study of Indian oil refinery. Production Planning & Control, 23(12), 903-921.

Gjerdrum, D., Peter, M., 2011. The new international standard on the practice of risk management–A comparison of ISO 31000: 2009 and the COSO ERM framework. Risk management, 31, 8-13.

Grekul, V., Korovkina, N., Korneva, K., 2015. Decision making in ITSM processes risk assessment. Computer modelling & new technologies, 19(5C), 12-16.

Ishikawa, K. (1986). Guide to quality control. 2nd ed. Asian productivity organization.

International organization for standardization, 2009. ISO31000:2009.

Marques, G., Gourc, D., Lauras, M., 2011. Multi-criteria performance analysis for decision making in project management. International Journal of Project Management, 29(8), 1057-1069.

McComb, D., Smith, J.Y., 1991. System project failure: the heuristics of risk. Information System Management, 8(1), 25-34.

Nguyen, L.D., Tran, D.Q., Nguyen, A.T., Le-Hoai, L., 2017. Computational model for measuring project complexity in construction. Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS 7851609.

Milosevic, D., Patanakul, P., 2005. Standardized project management may increase development projects

success. International Journal of Project Management, 23(3), 181-192.

Mishra, A., Das, S. R., Murray, J.J., 2016. Risk, process maturity, and project performance: An empirical analysis of US federal government technology projects. Production and Operations Management, 25(2), 210-232.

Mustafa, M. A., Al-Bahar, J.F., 1991. Project risk assessment using the analytic hierarchy process. IEEE transactions on engineering management, 38(1), 46-52.

Oehmen, J., Olechowski, A., Kenley, C.R., Ben-Daya, M., 2014. Analysis of the effect of risk management practices on the performance of new product development programs. Technovation, 34(8), 441-453.

Olechowski, A., Oehmen, J., Seering, W., Ben-Daya, M., 2016. The professionalization of risk management: What role can the ISO 31000 risk management principles play? International Journal of Project Management, 34(8), 1568-1578.

Purdy, G., 2010. ISO 31000: 2009—setting a new standard for risk management. Risk analysis, 30(6), 881-886.

Raz, T., Michael, E., 2001. Use and benefits of tools for project risk management. International journal of project management, 19(1), 9-17.

Remenyi, D., 1999. Stop IT project failure through risk management. Routledge.

Saaty, T.L., 1977. A scaling method for priorities in hierarchical structures. Journal Math. Psycol., 15(3), 234-281.

Tausworthe, R.C., 1979. The work breakdown structure in software project management. Journal of Systems and Software, 1, 181-186.

Vidal, L.A., Marle, F., Bocquet, J.C., 2011. Measuring project complexity using the Analytic Hierarchy Process. International Journal of Project Management, 29(6), 718-727.

Wallace, L., Keil, M., 2004. Software project risks and their effect on outcomes. Communications of the ACM, 47(4), 68-73.

Ward, S., Chapman, C., 2003. Transforming project risk management into project uncertainty management. International journal of project management, 21(2), 97-105.

Winch G., 2002. Managing construction projects, an information processing approach. Oxford: Blackwell Publishing.

Yassine, E.K., Brahim, H., 2016. Analytical hierarchy process based frame-work for supply chain risk management. International journal of scientific engineering research, 7(7), 1167-1172.

Zhang, G. H., Tang, W. J., Liao, M.J., 2015. Project Risk Management Analysis Model Based on AHP: A Case of Hangqian Freeway Project. In Intelligent Computation Technology and Automation Conference, pp. 663-666.

Zayed, T., Amer, M., Pan, J., 2008. Assessing risk and uncertainty inherent in Chinese highway projects using AHP. International Journal of Project Management, 26(4), 408-419.

AUTHORS BIOGRAPHY

Fabio De Felice, Professor at the University of Cassino and Southern Lazio, board member of several international organizations. The scientific activity developed trough studies and researches on problems concerning industrial plant engineering. Such activity ranges over all fields from improvement of quality in productive processes to the simulation of industrial plants, from support multi-criteria techniques to decisions (Analytic Hierarchy Process, Analytic Network Process), to RAMS Analysis and Human Reliability Analysis.

Antonella Petrillo, degree in Mechanical Engineering, Phd at the University of Cassino and Southern Lazio. Now Researcher at the University of Naples "Parthenope" (Department of Engineering) where she conducts research activities on Multi-criteria decision analysis (MCDA), Industrial Plant, Logistic and Safety.

Federico Zomparelli, degree in Management Engineering at University of Cassino and Southern Lazio. Now, he is a PhD student in Mechanical Engineering at the University of Cassino and Southern Lazio. His research activity is focused on MCDA, lean management, risk analysis and industrial plant optimization.

Marco Esposito, degree in Economics at University of Napoli "Federico II", he is chartered accountant and auditor. After being General Manager, now he is the CFO of ABC Napoli AS (Public Utilities). He also collaborates with the Faculty of Engineering of University of Napoli "Parthenope" like honorary fellow of Industrial Services Management.

SIMULATION-BASED DESIGN WITH UQ FOR CREATING
NEW COMBINATION OF FUNCTIONS CONSIDERING UNCERTAINTY

Takahiro Ichimaru(a), Hiroshi Hasegawa(b), Yuji Kado(c)

(a)(b) Graduated School of Engineering and Science Shibaura Institute of Technology, Japan
(c) Japan Aerospace Exploration Agency ,JAXA, Japan

(a)mf17009@shibaura-it.ac.jp, (b)h-hase@shibaura-it.ac.jp, (c)kado.yuji@jaxa.jp

ABSTRACT
The situation of product development takes a high cost, because designers have to remake the design plan for the design specification uncertainty of the upstream design. 75% of product development costs is the most serious problem for the product development. In this study, SBD with UQ (Simulation-Based Design with UQ) as a new design process is proposed to improve the problem of product development. SBD with UQ creates design plans with an optimized combination of functions and parameters by classifying uncertain specifications created in basic design and quantifying the uncertainties. In this paper, we consider the proposed method using CanSat which is a simple artificial satellite, and show the effectiveness of SBD with UQ.

Keywords: Simulation, Taguchi Method, Uncertainty Quantification,

1. INTRODUCTION

Since the 1990s, the globalization of world markets has progressed rapidly, and many companies have been faced with heating up product competition. Along with this, difficulties to develop products are an important issue in the industry because of trade-offs, like between improving product qualities and reducing production costs. When a designer plans a design process in a complex and rapidly changing design environment, it is difficult for the designer to grasp mutual relationships among tasks and to manage quality of design and product comprehensively throughout the design process if the complexity of the product increases (Karl and Steven 1995). Furthermore, there are a lot of uncertain information about a design object in upstream designs. As a result, it is difficult for the designer to make a decision because detailed specifications are not decided due to uncertainties. According to Bruno Lotter, 75% of product development costs are caused by mistakes of the making decision at the design stage (Bruno 2013). Figure 1 shows the breakdown of responsible stage of product development costs. The design stage consists of concept design, basic design and detail design. If a fault is discovered after the design phase like a manufacturing and a shipping, redesigns take much cost to deal with the fault. In other words, it can be said that the responsibility of the designer is heavy.
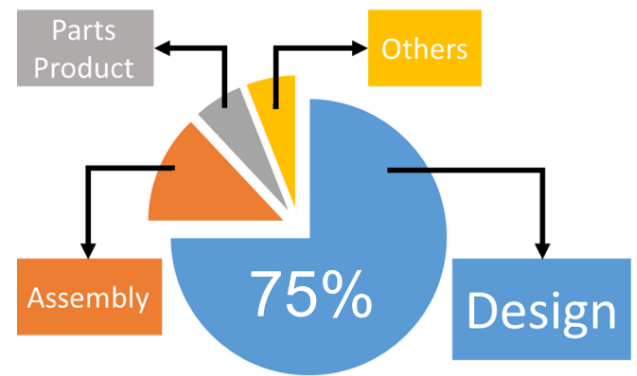
Figure 1: Responsible stage of product development costs

On the other hand, Ohtomi described the importance of the design stage in product development. According to Ohtomi, 80% of the total life cycle cost of the product is decided by the design stage (Ohtomi 2005). Focusing on the relationship between easiness of design change and cost of occurrence, it can be seen that the upstream design, which has not determined the details, has much choice for the design plan. These relationships are shown in Figure 2.
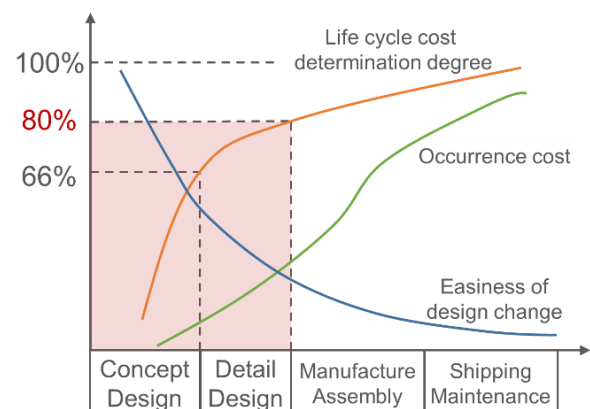
Figure 2: Importance of design in product developments

Therefore, it is important to quantify uncertainties to help making decisions and to simulate product development with trials and errors using by CAD, CAE, etc. In addition, it is the key to improve the quality of design whether or not it is possible to obtain a design plan that

maximizes the function while satisfying many constraints.

The purpose of this research is to propose a new design process which introduces simulation which combines various functions and parameters in concept design and creates many design plans taken into consideration of uncertainties of specification. This design process eliminates the difficulty of decision making due to uncertainties in upstream design and maximizes functions while satisfying various tradeoffs and constraints such as design quality improvement and cost reduction.

2. SIMULATION-BASED DESIGN WITH UQ

In this chapter, in order to solve the problem of the current design process, firstly we analyze the current design process and reliability of simulation and identify problems. From the result, we describe what elements are missed and needed to solve, and model new design method.

2.1. Present state and problems of design process
2.1.1. Current design process and 1DCAE

According to Ohtomi, it is pointed out that redesigns have a big influence on the development cost as the design process progresses if the upstream decision making contents are mistaken (Ohtomi 2005). With the current design process, the widespread adoption of Computer Aided Designing (CAD) and Computer Aided Engineering (CAE) to detailed design greatly changed the way of development. Meanwhile, in improved designs, developments based on structural design by CAD and CAE is suitable, but in new designs, developments based on the values and functions required for products. A system that reflects them in the structure is necessary. However, it is difficult to develop the system because of a gap between the upstream design and the downstream design. This gap is shown in Figure 3.
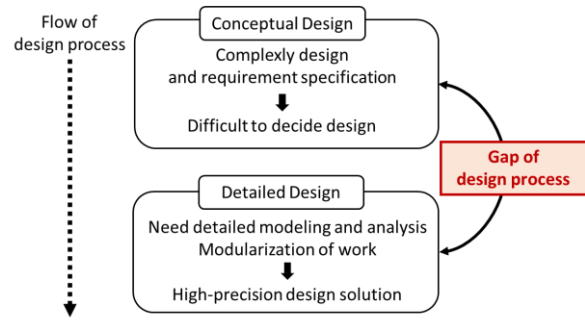
Figure 3: Gap between the upstream design and the downstream design

In the design based on value or function, the current CAD/CAE cannot be applied. In upstream design, decision making for product requirements and constraints is ambiguous. Also, the designer needs to consider all uncertainties of the design object. As a result, the designer needs to divide the design problem into partial problems that can be solved, however it is necessary to have a deep understanding and knowledge about the design object, the design process, the use environment, the market, the life cycle, etc. Therefore, there is a gap that the downstream design requires detailed design solution while there is a lot of uncertain information in the upstream design. Because of this reason, it is difficult to manage comprehensive design quality and work flow with the conventional design method.

In order to solve these problems, Ohtomi and colleagues have proposed One Dimension Computer Aided Engineering (1DCAE) which introduces and supports concept design simulation of new products (Ohtomi and Hato 2012). 1DCAE is a design method that supports various computer aided tools and models for decision making in conceptual design while constantly overlooking the entire design problem by expressing the essence of the design problem in a simple form. However,
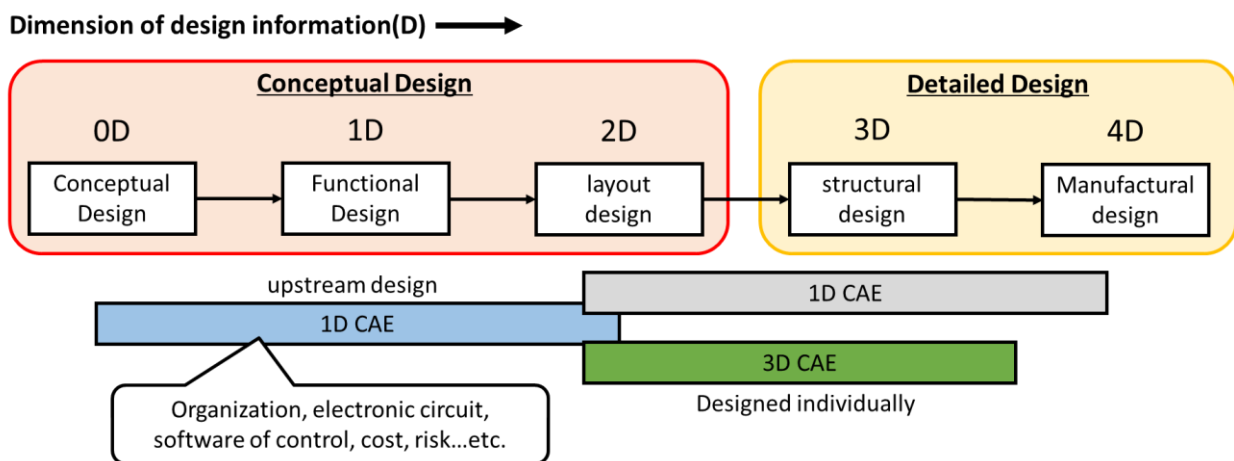
Figure 4: Position of 1DCAE in the design process

it is stated that "1D" means not expressing one dimension in particular, but expressing functions simply. 1DCAE makes it possible to evaluate from the upstream to the downstream stage of design by CAE. The position of 1DCAE in the design process is shown in Figure 4.

Before designing the structure, 1DCAE expresses the entire system to be designed by function, so that it can be evaluated and analyzed. Then, it is possible to create an optimum design plan in consideration of various functions of the entire design object in the upstream design. Further, using the result as an input, the divided design task is solved. Finally, return the result of the individual design to the whole design, and perform system verification.

By using 1DCAE, designers can gain two benefits. One is the creation of an optimal design solution in the upstream design and the other is the visualization of the entire system. By creating an optimal design solution in the upstream design, it is possible to obtain a robust design solution overlooking the downstream design. This is because it is applicable from the upstream design, so it can cover a wide design area and lead to the creation of new value. In addition, design problems can be found at an early stage of design, and design quality and development efficiency can be improved. Although the CAE method to the conventional structure made partial optimization possible, it was difficult to optimize the whole system by overlooking the entire system. However, 1DCAE can maximize value, minimize cost, and minimize risk by overall optimization. In addition, visualization of the whole system makes it possible to formulate specifications crossing fields such as machines, electrical machinery fields, and software that were designed separately. This makes partial optimization of field alone possible, enables overall optimization across fields, eliminating unreasonable and wasteful. Also, it is possible to prevent missing in the whole system, leading to quality improvement and safety and security. In addition to the visualization of the results, it is possible to visualize information such as parameters constituting each function, function, mechanisms, mechanisms, electric fields, and which fields are targeted. Based on these benefits, 1DCAE can be said to be a method of searching for global design solutions based on constraint conditions and design areas.

2.1.2. Reliability of simulation and uncertainty

Simulations such as CAD and CAE are the key to improving product quality in current product development. In other words, design quality can be complemented by improving the accuracy of simulation. However, from the 1990s to the present, the diffusion of CAD and CAE contributed greatly to the field of engineering, but simulation tools packed with advanced knowledge such as dynamics and computational science became black boxes. CAD and CAE can be designed very efficiently if the designer properly uses them. However, if the designer makes an incorrect usage, the safety of the design object developed based on that

mistake can not be guaranteed. Securing safety is the most important in product development, and this is the same in the simulation for product development. In other words, it is extremely important to create a system to guarantee the quality of simulation. The importance of this problem was pointed out in the National Agency for Finite Element Methods and Standards (NAFEMS) and The American Society of Mechanical Engineers (ASME), strategic efforts to international standardization are being developed. ASME recommends Verification and Validation (V&V) (ASME Standard V&V 10-2006 2006). Figure 5 shows the V-model process.

The purpose of V&V is to ensure the reliability of numerical simulation. "Verification" is a confirmation as to whether the object meets the specifications, design, planning, and other requirements. Also, "Validation" is an evaluation on whether the target function or performance meets the purpose or purpose and whether it has practical effectiveness.

The precision of the simulation also depends on the amount and quality of the information to be designed. The variation of these information is expressed as uncertainty. Uncertainty Quantification (UQ) is proposed in ASME V&V 10-2006 so that uncertainty is not reflected largely in the simulation model. The lower the quality of simulation, the lower the design quality. That is, consideration of uncertainty greatly affects design quality.
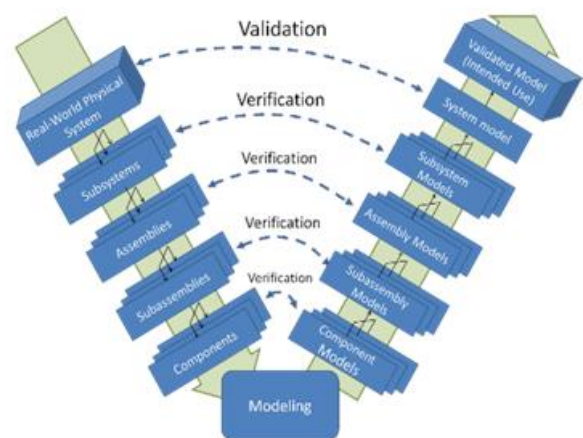
Figure 5: V-model process

2.2. Proposal of Simulation-Based Design with UQ

Based on the contents of chapter 2.1, in this research, we propose a new design process introduced a new uncertainty classification systematically considering uncertainties and a new quantification that enables design of various functions like 1DCAE for the basic design. The authors named this design process Simulation-Based Design with Uncertainty Quantification (SBD with UQ). Figure 6 shows the flow of SBD with UQ.

SBD with UQ creates a design plan with an optimized combination of functions and parameters by classifying uncertain specifications created in concept design and quantifying the uncertainties based on the classification in basic design. There are four advantages of SBD with
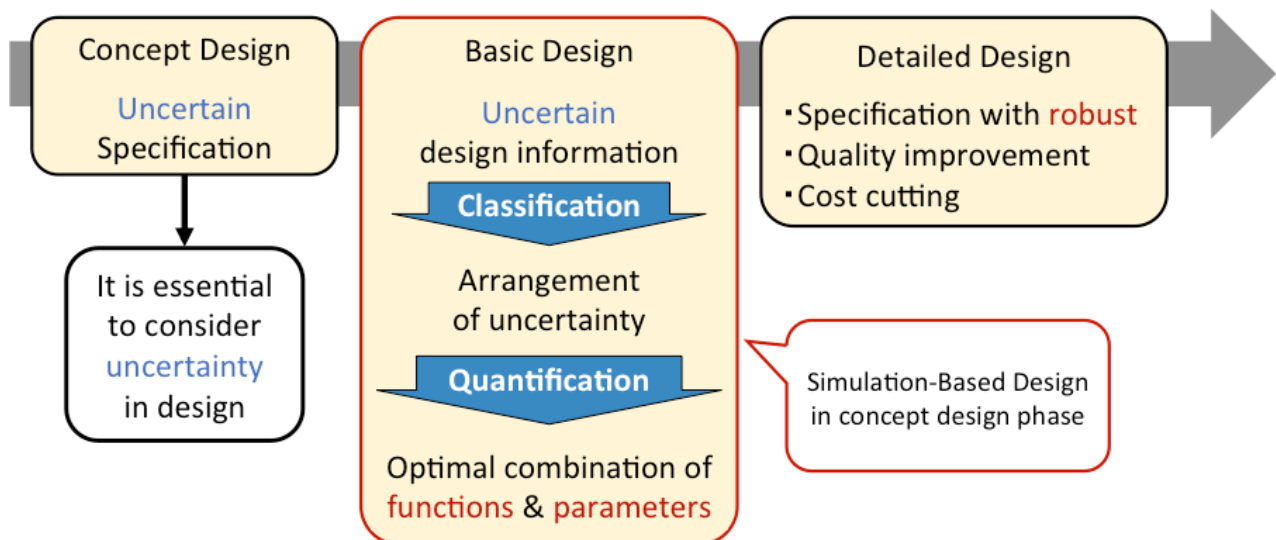
Figure 6: Flow of SBD with UQ

UQ. The first is to be able to obtain a combination of optimum functions and parameters from a large number of design plans. And the design plan takes the trade-off into account. Furthermore, in the upstream design, it is possible to optimize the whole function by the characteristics of 1DCAE, and it is possible to perform model creation using CAD in detailed design and verification simulation using CAE as robust design. As a result of these advantages, it is possible to secure reliability by early functional verification and to obtain a design solution considering V&V. In addition, detailed modeling of SBD with UQ is shown below.

2.2.1. Uncertainty classification and UQ

There are various uncertainties such as those due to factors that can not be modeled in simulation when designing and manufacturing products. In the preceding research various classifications are done. Among them, Kuroda classifies the uncertainty of the building field into six categories (Kuroda 1983). Since this classification by Kuroda covered the uncertainty in the building field, the authors applied this classification to the design field. The new classification of uncertainty in this study is shown in Figure 7.
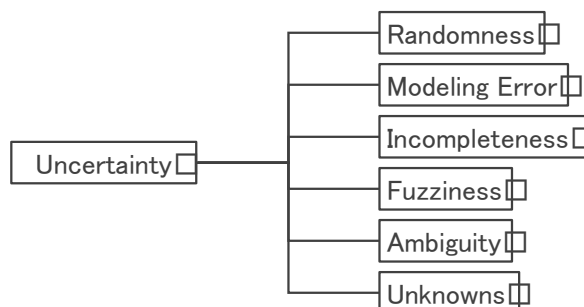
Figure 7: New classification of uncertainty

In this classification, variations in natural phenomena and physical properties are classified as "Randomness". Randomness is relatively quantifiable from experiments

and observations. On the other hand, there are gaps between calculation results given by simplified mathematical models and analytical models via various assumptions and real phenomena. Such uncertainty is called "Modeling Error". Although it is difficult to quantify the uncertainty of the Modeling Error, it can be reduced by giving a parameter range to the error. Furthermore, in Figure 6, "Incompleteness" is caused by lack of information at the time of design plan creation. In addition, there is "Ambiguity" which is difficult to determine in one analysis field by parameters belonging to a plurality of analysis areas. And there is "Unknown" which makes it difficult to predict after product shipment, such as change of social values .

The quality of design depends on how the uncertainty can be quantified in the upstream design. Current researches have developed various methods and support tools as Uncertainty Quantification (UQ). A design that considers uncertainty is called a robust design, and it is a robust design that is not easily affected by variations in product manufacturing and the use environment. In the current research, there are many researches on the method used for UQ and its robustness evaluation (ASME Standard V&V 10-2006 2006, Madsen, Krenk and Lind 1986, Hoshiya and Ishii 1993). In this research, we classified UQ into four levels according to the parameters used for UQ and its robustness evaluation. Level 1 is a definite value for which the parameter to be used is determined to be one. Examples include criteria such as upper limit and safety factor. In level 2, although the parameter to be used can not be decided as one value, its formation range is determined. Examples include physical property values such as temperature and humidity. Level 3 is a probabilistically confirmed parameter because even the established range of the parameter to be used is not fixed yet. It is often used for safety assessment of design objects, and there are many applications to space satellites and nuclear facilities. Examples include random numbers and genetic algorithms. Finally, in addition to the uncertain parameters related to the

robustness of the product, the concept of cost and benefit is introduced. Considering the cost of manufacturing, maintenance and repair, interest rate etc. for evaluation, design in total sense. "Design" which vaguely considered everything is considered to belong to level 4.

As the level increases from 1 to 4, the difficulty of quantification increases. There are many parameters with high difficulty of quantification in the specifications of the upstream design, and the reliability is low. As a result, the accuracy and stability of the design solution are lowered. Figure 8 shows the classification of UQ based on the used parameters and its robustness.

Lv.	Type	Examples	Volume of information	Reliability
1	Fixed value	size, weight etc.	Low	High
2	Values for which the range is fixed	physical property value etc.		
3	Confirmed value probabilistically	random number etc.		
4	Unknown	Fluctuations in Social values etc.	High	Low

Figure 8: Classification of UQ based on the used parameters and its robustness

In this research, we introduce the Taguchi method to stabilize the design quality. Application of the Taguchi method evaluates the uncertainty of Level 1 to 3 with the technical characteristics of each element by early functional verification and makes design development more efficient. In addition, early functional verification prevents level 4 market complaint and achieves cost reduction.

2.2.2. New UQ applying Taguchi method

Taguchi method is a concept created by Taguchi to efficiently realize product development and process management (Inoue 2008, Tatebayashi 2005). The quality in the Taguchi method is the reproducibility of the function. Reproducibility is how much precision you can make a thing in the same way and stabilizes quality by improving reproducibility. Characteristics of the Taguchi method are the evaluation of functionality and its improvement method. Taguchi method is to perform functional evaluation, and the big feature is use of orthogonal array, Signal Noise Ratio (SN Ratio) and sensitivity. By applying two parameters of control factor and error factor to an orthogonal array that can reduce the number of trials by performing statistical processing later, it is possible to know the trend of the overall functionality of the phenomenon. These trends can be known from indices of SN ratio and sensitivity. The SN ratio shows how stable the result can be obtained, and the sensitivity can be known to what extent each factor influences the output. However, since the Taguchi method is a method for quality checking and it is used in detailed design, it is rarely used in upstream design. This is because the parameters of the control factor and the error factor need to be determined. However, the stabilization of quality,

which is a characteristic of the Taguchi method, may solve the problem of the current upstream design. Therefore, this research proposes UQ using Taguchi method.

Kado proposes Product & Operation Sensitivity Analysis Method (P&O method) as new UQ (Kado 2014). The P&O method creates experiments and simulations using the orthogonal array by the Design of Experiments (DoE), and the calculation method of SN ratio and sensitivity is the same algorithm as Taguchi method. The difference from the Taguchi method is the following two points.

A) Utilizing the data of experimental capsules of "product information" and "operation information" created by DoE at the PLAN (experiment creation) stage of the P&O design method.
B) Definition of "noise factor" required for calculation of sensitivity and SN ratio, "operation information" is noise factor as seen from "product information" and "Product Information" is treated as error factor as seen from "operation information".

First, in the Taguchi method, the definition of "control factor" "noise factor" is common. This applies to Item A, control factor is what can be controlled by the designer, noise factor is what cannot be controlled by the designer. Using the orthogonal array applying a noise factor, its SN ratio and sensitivity are obtained. In other words, the intersection of the orthogonal array is synonymous with considering the error of the design object itself and the error in use as the design object. Therefore, by assigning the information product information about the design object itself and the usage situation operation information of the design object to the two axes of the orthogonal array, it can be regarded as experimenting the design object in the actually assumed environment. Furthermore, in Item B, it can be considered that it can be established as a DoE by considering the product information assigned to the two axes of the orthogonal array and the uncertainty which is a noise factor in the operation information.

Therefore, Taguchi method became applicable in the upstream design, and it was able to prepare to play the role like 1DCAE. This makes it possible to obtain the robust design solution in upstream designs. Moreover, in the P&O sensitivity analysis method, it is possible to have multiple levels for one factor. This makes it possible to obtain a noise-resistant design, although it is a discrete parameter, using the orthogonal array in DoE.

First, in this method, the design object is hierarchized as shown in Figure 9. By performing this hierarchy, it is possible to clarify the system hierarchy of the system, function, and parameter to be designed, so it is possible to relatively easily set the parameters considering the uncertainty to the level of the orthogonal array. The SN ratio and the sensitivity of the design object are calculated for each hierarchy based on the set lowest layer parameters, and the overall SN ratio and sensitivity are calculated by the bottom up scheme.
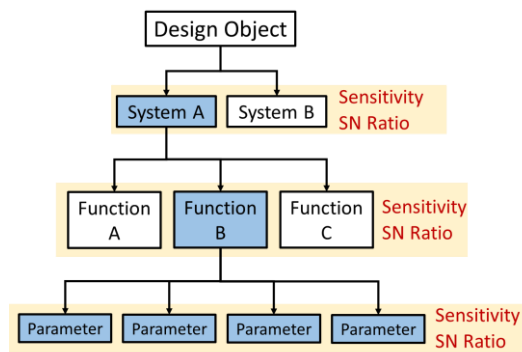
Figure 9: Hierarchized design targets

Based on the six types of uncertainty classification in design, SBD with UQ reclassifies the information into two types of uncertainty information to be assigned to the P&O method. In order to take account of uncertainty, these information are assigned to the orthogonal array as an established range. For example, it is assumed that the "Parameter A" exists and the established range is 0 to 100. Also, when an orthogonal array with three levels is used in this time, it is assigned to 0, 50, 100 and so on. This assignment is performed as same for other parameters. The example is shown in Figure 10.

Example : 【Parameter A】 Range 0 to 100

Parameter	Level 1	Level 2	Level 3
A	0	50	100

Figure 10: Setting "Parameter A"

2.2.3. Selecting the optimum combination

The parameters of the function selected as shown in Figure 11 are similarly calculated for other functions. Parameters efficiently selected by the DoE belong to the each function because of the hierarchicalization. Therefore, selecting various parameters according to the DoE is synonymous with deriving many design plans combining various functions. Moreover, it is possible to contribute to decision making in the upstream design because it is possible to set whether or not to use the function in these obtained design plans. For deriving the optimum design solution from among them, we choose the one with the highest average of sensitivity and SN ratio of each calculated factor. In this way, it is possible to select an optimum design solution in consideration of sensitivity and SN ratio.

The validation of this optimum design solution is confirmed by comparison with the requirements specification.

3. METHOD VALIDATION WITH CANSAT

We used Can Satellite (CanSat) for the validation of the proposed method in this research. CanSat is a simple artificial satellite that is as large as a hole can size.

3.1 Outline of verification

This validation is a comparison between CanSat which was designed using SBD with UQ and CanSat which the author made in the past. Figure 12 shows the overall view of CanSat.

Figure 12: The overall view of CanSat.

The feature of CanSat which we made in the past is that this CanSat was made by amateur designers with little consideration of uncertainty and was made by many trials and errors, but the mission was successful. These features are consistent with the problem which should be solved by SBD with UQ, so it is considered to be suitable for verification. Therefore, we compare and verify by using

Figure 11: Example of parameter selection

SBD with UQ in the same mission as CanSat which we made in the past.

The outline of the mission is shown in Figure 13. CanSat is raised from the ground to about 50 meter with the aircraft, and CanSat is released from there. And CanSat unfolds the parachute and is landed. After that, it is guided by a sensor such as GPS and moved to the destination using Rovers.
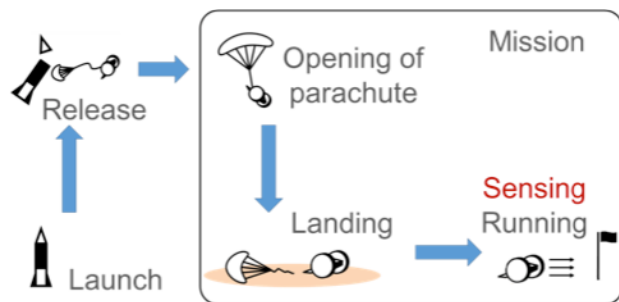
Figure 13: Outline of the mission

Moreover, the comparison verification items are two points. The first is the entire functionality of CanSat. And the other is the optimum combination of sub functions for position identification. However, it is assumed that they are selected from many candidates of sub function, and the more measurement items are selected, the higher the measurement accuracy will be.

3.2. Simulation setting

In accordance with the flow of SBD with UQ, we first create CanSat specifications and success criteria (Kawamo and Suga 2003). Based on these specifications, parameters of each function to be considered in this trial are allocated to product information and operation information. Parameters of each function are weight, power consumption, cost, driving force and so on. Moreover, in this trial, L54 orthogonal array is used for product information, and L121 is used for operation information.

3.3. Result and discussion

As a result of trial based on the simulation setting in Section 3.2, 6534 design plans were obtained. Table 1 shows the comparison between trial results of the simulation and the past CanSat and improvement values. It can be seen that SBD with UQ realized about 7% to 10% weight reduction, energy saving, cost reduction and performance improvement overall. Furthermore, in selecting the sub function for position identification, while realizing weight saving, energy saving, and cost reduction, it was possible to increase the number of selections and improve the system. Also, a part of the obtained sensitivity and SN ratio is shown in Figure 14. The design direction can be obtained from the indicators of the influence on the design plan such as the sensitivity and the SN ratio shown in Figure 14. In particular, it is understood from this trial that the presence or absence of
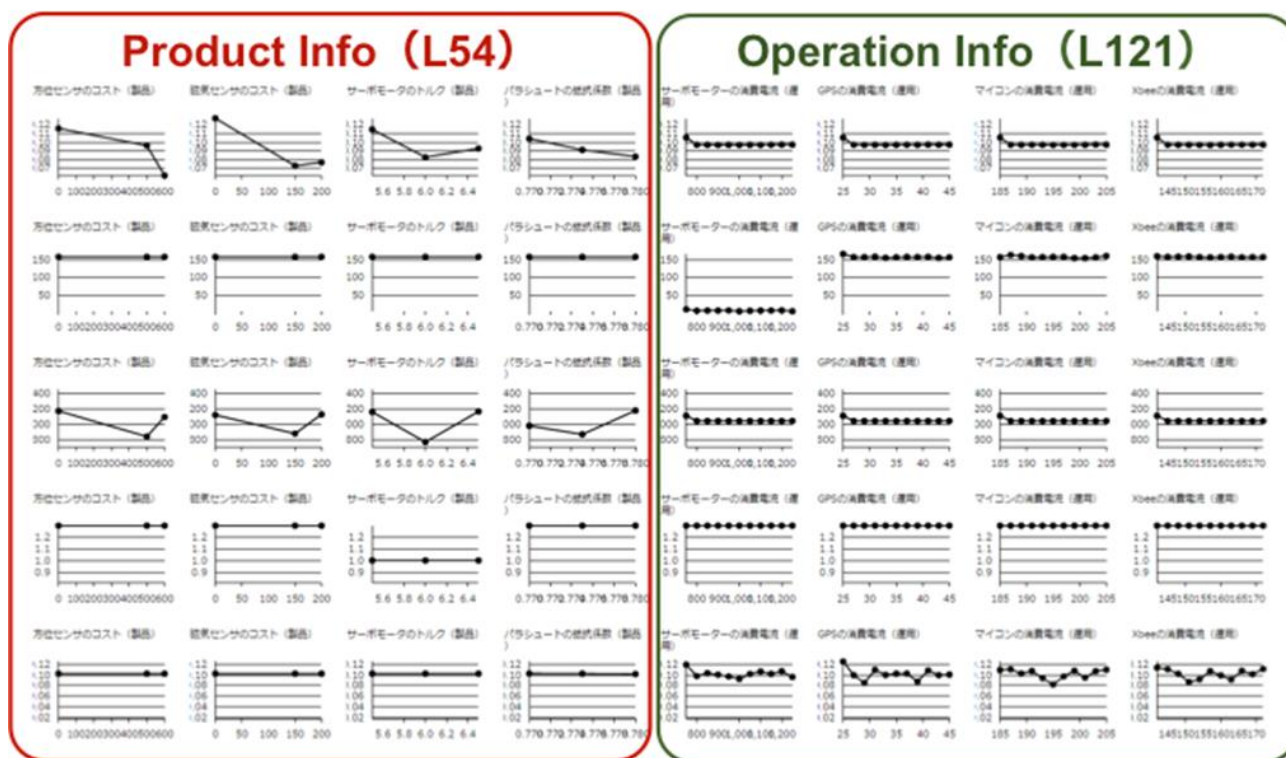
Figure 14: Part of the obtained sensitivity and SN ratio

selection of function is important. Figure 15 shows changes in SN ratio and sensitivity depending on whether or not the function is selected.

As can be seen from the change in the SN ratio and the sensitivity, the selection of the function has a large influence on the design plan, and it is possible to lead to the cost reduction by performing the early function verification in the upstream design and selecting the correct function.

Table 1: Comparison and Improvement Values

	Past	Optimal	Improvement[%]
Weight of structure [g]	380	350	7.89
Power consumption [mAh]	1240	1100	11.29
Cost [yen]	19000	17150	9.74
Thrust[kg・cm]	9.2	10.42	11.71
Parachute area[㎡]	0.62	0.67	7.46
Support function for position identification	1	2	1[個]
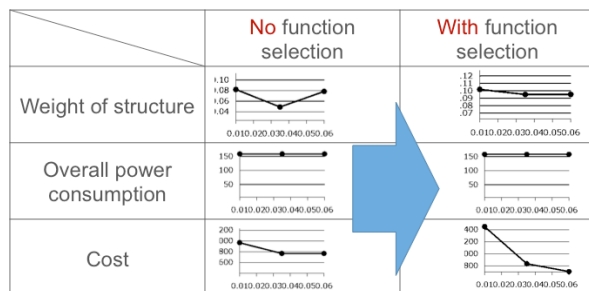
Figure 15: Changes in sensitivity and SN ratio depending on presence or absence of function selection

4. CONCLUSION

In this research, we proposed Simulation-Based design with Uncertainty Quantification, introduced the new UQ applied Taguchi method in upstream design

There are three points which we checked

Firstly, SBD with UQ is able to obtain a design solution with the optimal combination of functions and parameters. Second, we can choose and change easily design plan by early functional verification using sensitivity and SN ratio. Finally, function selection give a big impact on the pecification, therefore early functional verification at the upstream improves reliability of the design as V&V.

However, the level of uncertainty was low only in this trial. More design plans can be obtained when considering the parameter of Level 3 or higher, hence it is expected that a more complete design solution can be obtained.

REFERENCES

Karl T. Ulrich and Steven D., 1995. Eppinger, Product Design and Development, McGraw-Hill Inc.

Bruno L., 2013. Manufacturing Assembly handbook, Butterworth-Heinemann.

Ohtomi K. and Hato T., 2012, Design Innovation Applying 1DCAE, Toshiba Review Vol.67 No.7 (in Japanese).

Ohtomi K., 2005. Importance of Upstream Design in Product and Its Methodologies, Toshiba Review Vol.60.

Madsen H.O., Krenk S. and Lind N.C., 1986. Methods of Structural Safety, Chap.4, Prentice-Hall, Inc.,

Masaru Hoshiya, Kiyoshi Ishii, 1986. Reliability Based Design for Structures , p140 , KAJIMA INSTITUTE PUBLISHING CO., LTD.

Tachibayashi K., 2004. Introductory Taguchi Method, Japan Science and Technology Publishing Company (in Japanese).

Inoue K., 2008. Introductory parameter design, Japan Science and Technology Publishing Company (in Japanese).

Kado Y., Maru Y., Miyoshi H., Kakechi S., Hamada K., 2014. Tsuyoshi Koga and Kazuhiro Aoyama , Concept Design Method by Combination of Products and operations (Report 2: Parameter Sensitivity Analysis Method for Concept Design), The Japan Society of Mechanical Engineers, Design and System Division, 2014.

Kawamo K., SugaM., 2003. Creative Engineering Design Method, New Product Creation Principle, Yokendo LTD., pp.32-43.

AUTHORS BIOGRAPHY

Takahiro Ichimaru received his B.E. (2017) from Shibaura Institute of Technology (SIT), Japan. He is a master student at Division of Systems Engineering and Science, Graduate School of Engineering and Science, SIT. His research interests are design optimization and simulation-based design.

Hiroshi Hasegawa received his B.E. (1992) and M.E. (1994) from Shibaura Institute of Technology (SIT), Japan. He received Ph.D. (1998) in Mechanical Engineering from Tokyo Institute of Technology, Japan. He has been working at SIT, and currently is a Professor. He is member of JSME, ASME, JSST, JSCES and JSDE. His research interests include computer-aided exploration, especially multi-peak optimization, robust design and multi-disciplinary optimization

SIMULATION OF A READINESS-BASED SPARING OPTIMIZATION

John Wray[a], Arnold Buss[b], Javier Salmeron[c]

[a],[b] Modeling, Virtual Environments, and Simulation, Naval Postgraduate School, Monterey, CA 93940, U.S.A.
[c] Operations Research Department, Naval Postgraduate School, Monterey, CA 93940, U.S.A.

[a] jdwray1@nps.edu, [b]abuss@nps.edu, [c]jsalmero@nps.edu

ABSTRACT
We develop a discrete event simulation to complement a new optimization tool that establishes inventory levels for aviation weapon systems (WS) in the U.S. Navy. The optimization seeks cost minimization while achieving required readiness rates for hundreds of WS, each comprising thousands of indentured parts. Based on work in similar realms, the optimization employs the Vari-Metric model to estimate overall WS readiness and a variant of a greedy heuristic algorithm to set stock levels for all parts. Our simulation tests the assumptions and provides additional metrics for decision makers. We find that the estimates for readiness yielded by the optimization tool (a) have no systemic bias, and (b) remain within 5% in 53 of 64 WS (with an 8% worst-case). We also test two legacy optimization tools currently used by the Navy and find they have larger errors in expected readiness. We also identify factors correlated to these differences.

Keywords: Readiness based sparing; discrete event simulation; optimization; multi-indenture

1. INTRODUCTION
The United States Chief of Naval Operations (2011) requires the use of "readiness-based sparing (RBS) methodology to spares and repair parts allowance determination to ensure that prescribed readiness thresholds and objectives are achieved at the lowest possible cost." RBS uses advanced analytics to set inventory levels for most U.S. Navy parts and sub-parts at different locations.

To guarantee required combat power for the combatant commanders of the U.S. Navy (USN), all naval aviation Weapon Systems (WS) must maintain specified readiness (i.e. availability) rates. The term WS here identifies platforms such as the F/A-18 (Hornet) attack aircraft, or the MH60 (Seahawk) helicopter, among others. While reliability and maintainability are primarily set in the design phase of a WS, supportability is a crucial aspect of readiness that can be adjusted throughout the lifecycle of the system to achieve desired readiness rates. Supportability is affected by several factors; one of the key controllable elements is stock levels for spare parts at different echelons of supply.

Selecting the right mixtures of parts to stock at any given site in the USN is a very challenging task in a budget-constrained environment. A naval site contains numerous WS of different types and each WS may contain thousands of parts each failing at different rates. While it may not be possible to identify a provable optimal inventory for every site, our goal is to design and implement optimization and simulation tools that approximate such solutions and provide inventory policies that result in significant cost savings and improved fleet readiness over alternative solutions.

Although fill rate is a popular choice for evaluating inventory policies, it is problematic in a military setting where the ultimate goal is sufficient availability of WS. Although improving fill rates or reducing backorders will in fact improve readiness, policies developed with these metrics alone will be inefficient (Moulder et al. 2011). Looking solely at fill rates will inadvertently punish more complex WS. With all other factors such as failure rates and mean time to repair (MTTR) being equal, a WS with more parts will be requesting more parts from supply. If 95% of the parts are available upon request, a WS with more parts will be unavailable more often while awaiting parts than a WS with fewer parts.

In order to assist Naval Supply Systems Command (NAVSUP) with RBS planning, we develop an RBS Simulation (RSIM) to verify the recently developed Navy Aviation RBS Model (NAVARM) estimates and also compare its performance to the legacy Service Planning Optimization (SPO) and Aviation Readiness Requirements Oriented to Weapons Replaceable Assemblies (ARROWS) tools.

2. NAVARM AND RSIM

2.1. NAVARM Overview
NAVARM (Salmeron 2016) embeds a heuristic algorithm that approximates the optimal inventory quantities for a single-site, multi-indenture problem. Specifically, NAVARM recommends reorder quantities that minimize the cost of inventory held while maintaining pre-specified target availability rates for all WS.

NAVARM uses an (S-1, S) inventory model for all parts and sites. That is, S is the (maximum) stock level at a site determined by NAVARM and an order is placed as soon as that level decreases by one (i.e., the reorder point is S-1). This means that each time a part fails, it is turned into the system for repair. If the part cannot be repaired, a new part is ordered to replace it.

Assuming every part i is given a stock level S_i, every WS has an estimated availability that is calculated as a function of the expected backorders (EBOs) of the highest indenture parts in the WS. Naturally, backorders for any part in the system are a random variable which depends on: (a) the part's stock level; (b) its (possibly different) failure probability distributions for all common parts in the same or different WS; and (c) the backorder distribution for sub-indentured parts to all common parts. The underlying theory to calculate EBOs for a given set of inventory levels S_i is known as the Vari-Metric model, see Sherbrooke (1986, 2004, pp. 101-125).

The Vari-Metric model estimates EBOs under the assumption that, even though the number of failures for a given part can be approximated using a Poisson distribution, the actual number of failures after accounting for sub-indentured parts' failures is distributed as a Negative Binomial.

The multi-indenture structure used to describe WS repair with more fidelity complicates the problem significantly. Figure 1 illustrates this idea for three hypothetical WS at a site. For simplicity each WS has only one first-indenture part. If we are interested, for example, in improving the availability of WS 3, we can look at ways to decrease backorders of sub-parts "R" and "L". But, noting that part "L" is common to WS 2, its backorders are impacted by parts "M" and "N", and therefore by "G" in WS 2. Moreover, since this is common to WS 1, stocks of parts "H" and "I" in WS 1 will affect backorders of "L" in WS 3. The fact that WS 3 can be influenced by WS 1's parts (which have no direct commonality with parts in WS 3) is a challenging aspect of RBS optimization.
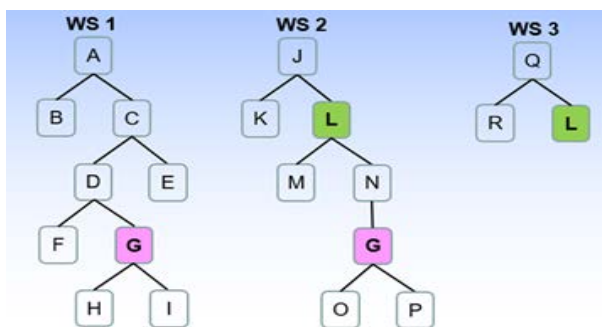
Figure 1: The Chain of Influence in a Multi-Indenture Part Structure

Sherbrooke (2004) points out that while the multi-indenture structure and the likelihood of common parts across WS types "does complicate the computer programs substantially … the basic logic is the same."

The use of heuristics to approximate the problem of satisfying a certain availability at minimum cost is justified due to the lack of a closed-form expression for expected readiness rates for a given reorder policy.

The Vari-Metric model suggests using a greedy heuristic based on an "effectiveness ratio" that measures improvement in EBOs with respect to cost. Parts with higher ratios are chosen until the desired availability is met. Even though this greedy heuristic is not provably optimal in a discrete setting (where we cannot order a fraction of a part), and counterexamples can be easily built, the method appears to work well in practice.

The matter becomes more complex when there are multiple WS with common parts. This is because if we follow the greedy algorithm for one WS at a time, we will achieve the desired availability at (approximately) minimum cost for, say, WS 1. But then, we will need other parts when working on WS 2. If some of those parts are common to WS 1, we will increase its availability unnecessarily above its target. Thus, refinements are needed, and NAVARM implements some of those, which basically consist of revisiting all of the WS above target in order to remove parts and reduce cost.

2.2. RSIM Introduction and Scope

RSIM (Wray 2017) simulates failures at the individual part level and then aggregates up to the individual WS level and WS type to help assess the accuracy of expected backorders and WS availability. To simulate the system of interest, three major classes of entities are created: parts, WS and part positions. Each part has attributes that include:

- Status (i.e., functioning or down for maintenance),
- Planned failure time (detailed below), and
- Position (specifying where the part is installed if currently in use)

Each WS has attributes that include:
- Type (e.g., CH-53 helicopter),
- Availability status (i.e., up or down), and
- A list of part positions that comprise the WS (e.g., hydraulic pump).

A part position has attributes that include:
- The WS (if currently in use),
- Parameters describing expected failure times, and
- Parameters describing the time for a working part to return to inventory after breaking.

Modeling failures in a manner that closely mirrors reality is crucial to attaining realistic outputs. Expected failure rates can be derived from existing databases and are broken down into failures that can be repaired at the site and failures that cannot. Some parts have only one type of failure or the other while some have both.

The type of failure is tracked in RSIM to later develop an expected time the part will return to inventory in a working status. To handle the difference in types of failures, RSIM first combines the failure rates and assigns a failure time based on the rate. When the failure occurs, a random number draw is compared to the ratio of repairable and non-repairable failures to assign the type.

While multi-parameter distributions such as the Weibull that allow specification of mean and variance are generally preferred for detailed modeling of failure rates, the databases used for RBS currently provide only the mean failure rates. As a result, the exponential distribution is employed by RSIM to generate the next failure. This is also consistent with the assumptions established for NAVARM, SPO and ARROWS.

Failure rates in the database and RSIM are specific to a part position on a WS type; a hydraulic pump used on a CH-53E utility system may have a different failure rate than the same type of hydraulic pump used on an SH-60S utility system. In fact, the same pump may be installed on different WS types between failures and thus have different failure rates assigned based on where it is installed. When a part fails, RSIM removes the part from the usable pool for a specified period of time until repaired or replaced.

Although RSIM is best described with an event graph (see Figure 2), the basic steps are as follows:

- Read data in from database and instantiate all entities specified in the data.
- Assign parts to fill each WS and assign a first failure time stochastically for each part based on its specified distribution.
- When a failure occurs:
  − Assign a time the part will return to service.
  − If a part of the correct type is available in inventory, place WS in down status for the specified MTTR. If a part is not available, add the WS to a first-in, first-out (FIFO) queue for that part type.
- When a part returns to a ready-for-issue status at the site, use it to repair the first WS in the FIFO queue awaiting that part type. If no WS are awaiting that part type, return the part to inventory.

To manage complexity, simplify the verification and validation process, and ensure acceptable simulation run-time, RSIM tightly scopes the factors considered in the simulation while maintaining flexibility to add new factors as desired to closer mirror reality or support study objectives.

In its current state, RSIM ingests summary level data on flight hours, failure rates, repair times and shipping times and most of these factors are treated deterministically. While RSIM could simulate actual flight sorties and assign failures based on WS flight times, the effect of this added fidelity would likely be nominal when considering inventory policies and thus is not included. Likewise, scheduling the repair process at intermediate and depot level and including manpower and part availability consideration here would also have minimal effect on the metrics currently of interest; instead, expected values are substituted in lieu of this detailed analysis.

Finally, RSIM could consider the phasing of required repairs and how they may coincide with required periodic inspections and planned maintenance to minimize downtime. Part failures that render the WS partial mission capable could remain on the WS until an optimal time to complete the repair. Again, the effect of including this would not shed light on the objectives at hand though it may be a worthwhile future enhancement to provide decision makers with a fuller sense of what to expect if the given NAVARM solution is implemented.

2.3. RSIM Assumptions

A number of assumptions are made in the RSIM implementation, some of which could significantly impact the results. These assumptions are made for a variety of reasons to include limited data availability, code simplicity, and reduced run-time. The inherent flexibility in RSIM implementation makes these assumptions fairly easy to modify or eliminate through code manipulation. The following are significant assumptions currently made in RSIM:

- Failure rates are accurately represented by an exponential distribution: As stated early, failure rates would likely be better represented with a Gamma or Weibull distribution, but the limited failure data provided does not allow for such implementation. The exponential distribution is not very well suited to represent wear out failures that occur at fairly predictable intervals as compared to random failures.
- Failures are independent: Because failure times are scheduled into the future on a continuous timeline and there are no dependencies in our program, simultaneous failures will not occur despite real world experiences that suggest otherwise.
- Failures in the simulation should continue to happen when the WS is down: While failure rates in the database are given per flight hour, this data along with average flight hours is used to develop expected mean time between failures. Although parts are much less likely to fail when the WS is out of service, scheduled failures continue to occur in the simulation to ensure the expected failure rate is maintained. This may result in overlap of delay times for backordered parts. With a higher fidelity data set, this could be improved by developing conditional probabilities for failures that better reflect the empirical data.

- Expected sub-indentured part failure times are not reset when a parent part is changed: This assumes that all parts are repaired and that when they undergo repair, it does not affect reliability of the separate sub components. This assumption will fail if the part is replaced and sub components are not salvaged, but the available data does not delineate how often parts are repaired when they go off-site and what happens to sub-indentured parts when a replacement is necessary for the parent part. Of note, this assumption will lead to a conservative estimate of availability, though the extent of the impact is unknown with the data currently available.
- Demands are FIFO: This assumes that no priority will be given to WS of types that are below their availability goal or some other prioritization scheme.
- No lateral resupply: There is no cross-leveling between sites that have high and low inventories (or backorders) for a particular part.
- Cannibalization: While moving working parts from a down WS to one that can be returned to an up status is practiced in the real world, RSIM is conservative in not assuming so.
- Repair times are independent: RSIM does not attempt to simulate backlogs in the repair pipeline that would likely occur if multiple parts of the same type were in the repair pipeline simultaneously.

2.4. RSIM Event Graph and Implementation

Figure 2 depicts an event graph describing the overall model of part failures and subsequent repairs in RSIM. This version of the event graph is simplified and intended to provide a broad understanding of the flow of parts through the system.
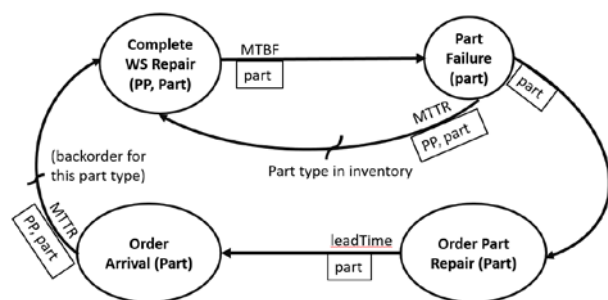
Figure 2: Simplified RSIM Event Graph

The RSIM event graph depicted in Figure 2 has been implemented in the Java programming language using the Simkit library (Buss 2002, 2004). Simkit provides the necessary support for easily converting the event graph into working code. As an open source library, Simkit is free of the encumbrances of commercial licensing, while providing excellent support for the model's features. An additional open source library,

UCanAccess (2017), has been used to interact with the MS Access database inputs.

3. RESULTS

3.1. Introduction

In its current configuration, RSIM outputs several metrics by WS and part type to allow comparison to other RBS optimization software (i.e., NAVARM, SPO, and ARROWS). Additionally, it provides decision makers with a more comprehensive understanding of what to expect if a certain inventory policy is implemented. For each WS type, RSIM provides the mean number of WS available, the corresponding readiness rate, and the percent of simulated time the WS type was at or above its given readiness goal. For each part type, RSIM outputs the mean inventory level, mean number of backorders, and the fill rate.

The primary metric of interest is the readiness rate by WS type. Given the crucial nature of having required force levels available at any given time, NAVSUP must ensure the inventory quantities selected will enable this objective. RSIM, NAVARM, SPO, and ARROWS each have assumptions built in that may not be accurate, but a comparison of the outputs can be helpful in assessing the validity of the readiness estimates.

The data used for the analysis below is generated using Dell Inspiron I5378 laptop running Windows 10 with an Intel Core i7-7500U 2.7 GHz CPU and 8 GB of RAM. RSIM is implemented in JDK 1.8 and utilizes the 64-bit Simkit version 1.4.6 and UCanAccess version 4.0.1. Based on steady state analysis conducted for several sites, we use a warmup period of 3,000 simulated days before collecting 7,000 simulated days of data for 30 replications at each site analyzed. These settings result in less than 1% margin of error for readiness levels of all WS tested. Run times range between 2.5 and 59 minutes for the seven sites analyzed. NAVARM runs are conducted on the same laptop described above using the 32-bit version of Microsoft Excel 2016. While there are several NAVARM setting that can significantly affect the run time, standard settings result in run times ranging between 30 seconds and 18 minutes.

3.2. RSIM Compared to NAVARM

We have run RSIM on seven representative USN sites to compare expected WS availability rates and expected backorder rates for a given allowancing to those anticipated by NAVARM. The number of WS types at these Naval sites ranges from 3 to 23 with a mean of just over 9 WS types per site and a mean of 111 individual WS per site.

Figure 3 shows the summary histogram of the differences in expected availability for the 64 WS types tested. Out of the 64 WS types analyzed, 53 have expected readiness levels within 5% and the mean difference for all WS types in this sample is .2% with no systemic bias to over or under estimate readiness noted.
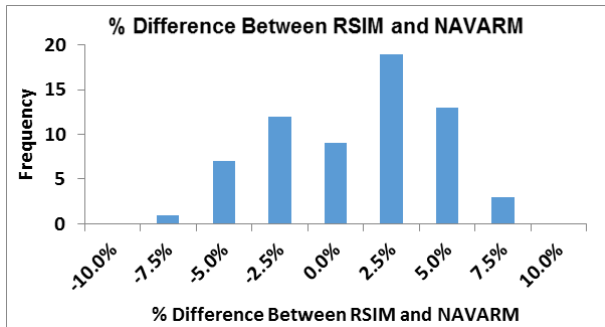
Figure 3: Differences in Expected Availability between NAVARM and RSIM

3.3. Comparison to Legacy RBS Tools

As NAVSUP considers whether to switch RBS optimization tools, it is crucial for the decision makers to assess the accuracy of the NAVARM expected readiness calculations and compare its accuracy to the SPO and ARROWS tools currently in use. Because RSIM models the system at the part and WS level, its method of observing readiness rates through the course of a simulation provides an independent observation to compare against the optimization tool estimates available. SPO, ARROWS, and NAVARM have been run at a representative site with seven different WS types and a total of 62 WS. Their recommended inventory policies have been simulated in RSIM in order to compare the expected readiness rates for each WS type. Figure 4 shows a summary of the resulting differences in estimates. In this case, it becomes clear that NAVARM's estimated readiness rates are much closer to RSIM than SPO or ARROWS estimates are.
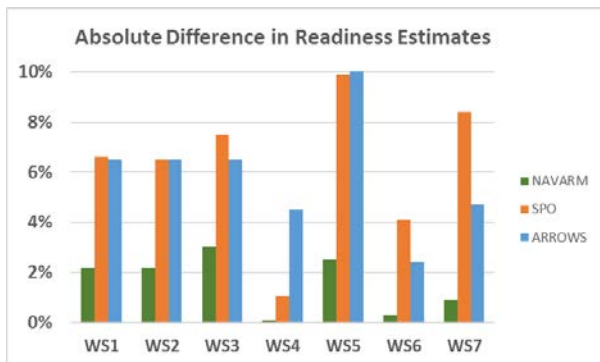
Figure 4: Absolute % Difference between Available Optimization Tools and RSIM Readiness Estimates by WS Type at a Single Site

3.4. Analysis of Attributes Affecting Performance

While the difference between expected readiness given by RSIM and NAVARM is likely acceptable in the current versions, we have tried to identify key drivers of any differences found in the hope of further explaining the differences and ideally reducing the errors. First, we consider attributes of the WS type that may complicate calculations for availability in the models. The factors of interest by WS type are:

- Quantity per application (QPA): Some part types are found on a WS multiple times at the same indenture level. In these cases, the part type is represented with a single line in the database and the number of parts used is given as QPA. While QPA implementation is straightforward in simulation, its use in the readiness estimation for NAVARM has been debated.
- Commonality: This is a measure of how often the same part type is used throughout the site. This adds a layer of complexity in the optimization model as changes in inventory can affect numerous WS types.
- Number of parts: The number of parts tracked on a given WS type in our data sets ranged from 80 to over 8,000.
- Indenture depth level: Assumptions are made in both RSIM and NAVARM regarding the impact of indenture depth level on part failure rates and readiness.

Each of the above factors has been examined compared to the difference in readiness estimates produced by NAVARM and RSIM. None of the correlations are significantly strong, with total number of parts being the highest (0.49), indenture depth level and mean number of common parts slightly lower (0.43 and 0.40, respectively), and QPA being clearly non-significant (0.02).

3.5. Comparison of EBO Outputs

Because EBOs play an integral part in the NAVARM calculations of expected readiness, we configure NAVARM and RSIM to output EBOs for every part position to compare expected EBO levels. At the sample site described above, there are over 11,000 part positions tracked. Of these, approximately 6,400 part positions are expected to have some backorders based on their failure rate and stock level. The magnitude of the difference is generally negligible with only a 2% difference in the sum of EBOs for NAVARM and RSIM and an average difference of less than .0003 per part. Figure 5 shows a histogram comparison of EBOs. Parts with extremely low EBO levels will not significantly impact overall WS readiness. Of note, the counts are nearly identical for EBOs above .001.

We consider the same abovementioned factors as potential drivers in the difference in EBO levels between NAVARM and RSIM, but again neither of these factors demonstrates strong correlation with the differences in EBO levels. Specifically, the fact that indenture depth does not strongly influence availability or EBO differences suggests that the Vari-Metric assumption of negative binomial distribution for modeling EBOs of sub-indentured parts is reasonable.
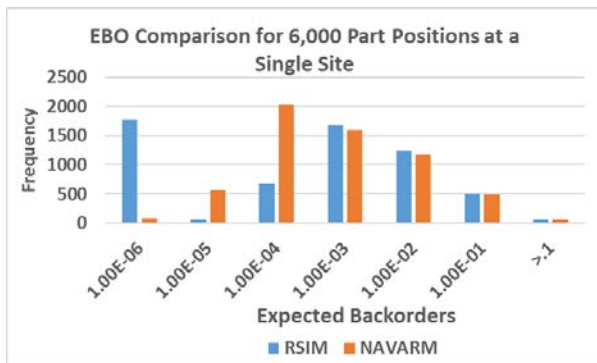
Figure 5: Observed RSIM EBO Levels Compared to NAVARM Expected EBO Levels at a Single Site

3.6. Additional RSIM Insights

In addition to verifying NAVARM outputs and providing an independent comparison of the three available RBS optimization tools, RSIM can provide additional insights not available with the optimization output. One example of this is the expected readiness levels. While the optimization tools only provide the expected readiness levels overall, RSIM provides metrics that include percentage of time above the stated readiness goal and the readiness levels observed at the beginning of each simulated day.

For example, Figure 6 shows a histogram of observed daily readiness levels over a period of 7,000 simulated days for a single WS type with 22 WS at a single site. Even though the most important output of RSIM is the expected readiness achieved (in this case 60.7%, slightly below the 63% goal), additional valuable information can be gleaned: In this simulation run, 48.2% of the simulated time had readiness rates above the goal. A decision maker may be more interested in worst-case scenarios to ensure that assumptions made for contingency planning are realistic. The fact that we expect less than 50% readiness during 11% of the time may be of interest.
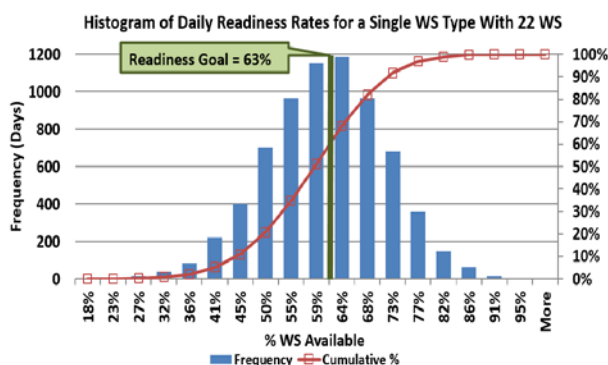
Figure 6: Histogram of Daily Readiness Rates in RSIM for 7,000 Simulated Days of a Single WS Type

In addition to providing useful metrics for decision makers, RSIM provides metrics that may be used to improve recommended inventory levels calculated by NAVARM. Table 1 shows a small sample of output

from RSIM for six part types at a single site. The mean backorder level, mean inventory level, number of failures over the period of the simulation and the fill rate are output for each part type. The full output will contain thousands of entries, but an analyst could sort this list to help identify areas where inventory levels could be manually adjusted to incorporate other factors not accounted for in the NAVARM optimization, such as those due to modeling assumptions described in Section 2.3. This process could be automated and take advantage of both NAVARM and RSIM to evaluate the changes. Moreover, RSIM could be extended to implement its own adjustments and become a complement to NAVAIR's optimization.

Table 1: Sample RSIM Output by Part Type

part type	mean backorders	mean inventory	# failures	fill rate
Part 1	0.01	0.89	11	0.91
Part 2	0.37	1.23	1592	0.62
Part 3	0	0.93	7	0.86
Part 4	0.11	0.55	65	0.63
Part 5	0	4.96	6	1
Part 6	0	0.93	6	1

Simulation lends itself well to the collection of numerous metrics. Here we have presented several metrics that may be useful to the decision makers or analysts. As our understanding of the problem continues to develop, we expect to modify the RSIM assumptions and metrics accordingly.

ACKNOWLEDGMENTS

REFERENCES

Buss, A., 2002. Component Based Simulation Modeling with Simkit. Proceedings of the Winter Simulation Conference, pp. 244-249. December 8-11, San Diego, California.

Buss, A., 2004. Simkit Analysis Workbench for Rapid Construction of Modeling and Simulation Components. Paper Presented at the Fall Simulation Interoperability Workshop. April 18-23, Arlington, Virginia.

Moulder, R.D., O'Malley, T.J. and Peterson, D. K., 2011. Sizing Inventory with Readiness-Based Sparing. Air Force Journal of Logistics 35: 18-29.

Salmeron, J., 2016. "Readiness-Based Sparing Inventory Optimization Model." Presentation to NAVSUP. Naval Postgraduate School, Monterey, CA. (Available from author.)

Sherbrooke, C. C., 1986. VARI-METRIC: Improved Approximations for Multi-indenture, Multi-

echelon Availability Models. Operations Research 34 (2): 311-319.

Sherbrooke, C. C., 2004. Optimal Inventory Modeling of Systems: Multi-Echelon Techniques Second Edition. New York: Kluwer Academic Publishers.

UCanAccess, 2017. A pure java JDBC Driver for MS Access. Available from: https://sourceforge.net/projects/ucanaccess/ [accessed 29 June 2017]

United States Chief of Naval Operations, 2011. Readiness Based Sparing. OPNAV Instruction 4442.5A.

Wray, J.D., 2017. A Simulation of Readiness Based Sparing Policies. M.S. Thesis in Modeling, Virtual Environments and Simulation. Naval Postgraduate School, Monterey, CA.

AUTHORS BIOGRAPHY

Major John Wray is a helicopter pilot in the United States Marine Corps. He earned a B.S. degree in Systems Engineering from the United States Naval Academy, and two M.S. degrees in (a) Operations Research and (b) Modeling, Virtual Environments, and Simulation, from the Naval Postgraduate School in Monterey, CA.

Arnold Buss received his B.A. from Rutgers University, M.S. from University of Arizona, and Ph.D. from Cornell University. He is currently a Research Associate Professor in the Modeling, Virtual Environments, and Simulation Institute at the Naval Postgraduate School. His research interests are in simulation methodology, applied simulation modeling, and agent-based modeling. His previous positions include Assistant Professor at Washington University and Visiting Assistant Professor in Operations Research at the Naval Postgraduate School. His simulation software, Simkit, has been used to teach discrete event simulation for the past 20 years. http://faculty.nps.edu/vitae/cgi-bin/vita.cgi?p=display_vita&id=1023568056

Javier Salmeron received the M.S. and Ph.D. degrees in mathematics from Complutense University and Polytechnic University of Madrid, Spain, respectively. He is currently an Associate Professor in the Operations Research Department at the Naval Postgraduate School, Monterey, CA, USA, where he teaches courses in optimization and its applications. His research focuses is in the area of applied modeling and optimization, which he has actively used in multiple civilian and military projects. http://faculty.nps.edu/jsalmero/

HEURISTIC MODULES MULTI-LIFT PLANNING TOOL FOR INDUSTRIAL SITE

Mahmoud Farajmandi[(a)], Ulrich (Rick) Hermann[(b)], Hosein Taghaddos[(c)], Simaan M. AbouRizk[(d)]

[(a),(d)]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada T6G 2W2
[(b)]PCL Industrial Management Inc., 5404 99 St., Edmonton, AB, Canada T6E 3P4
[(c)] Assistant Professor, School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran

[(a)]farajmandi@ualberta.ca, [(b)]rhhermann@pcl.com, [(c)]htaghaddos@ut.ac.ir [(d)]abourizk@ualberta.ca

ABSTRACT

Planning and sequencing module installation on industrial sites is critical for delivering projects safely, on time, and within budget. Depending on module size and weight, as well as crane availability, location, and configuration, various sizes of heavy-duty mobile cranes are used to safely pick, swing and place modules for installation. High crane operating costs, vast number of installation options, and multiple crane-module technological constraints require schedulers to spend weeks using a trial-and-error based approach to prepare and improve module installation plans. A formalized framework for generating a feasible optimum installation plan is essential to minimize crane operation costs. This paper presents a novel heuristic-based methodology for planning and sequencing module installation on industrial construction sites. The algorithm is incorporated into a developed software toolkit. Case studies are presented to demonstrate the ease and effectiveness of the developed methodology.

Keywords: module installation planning, mobile crane, heuristic, decision support tool

1. INTRODUCTION

In industrial modular construction, various types of modules, such as pipe rack and equipment modules, are prefabricated in remote module yards. Fabricated modules are then shipped to construction sites where they are sequentially assembled according to module-specified patterns indicated in design documents. Using mobile cranes, modules are picked from trailers and are lifted into place for installation on site. Depending on module size and weight, as well as crane availability, location, and configuration, various sizes of heavy-duty mobile cranes are used to safely lift, swing, and place modules into their designated locations.

Mobile crane operation expenses, which include mobilization, demobilization, foundation preparation, reconfiguration, relocation, and rental fees, represent a substantial proportion of overall module installation costs (Taghaddos et al. 2010). By minimizing the frequency of crane reconfiguration, relocation, and the number of rigging changes during project execution, crane operation-associated expenses can be substantially reduced. Optimization of module installation plans, therefore, has the potential to considerably reduce overall project costs.

However, generating and optimizing on site module installation plans is a complex process. Sequencing and scheduling of module installation must consider several crane-module technological constraints, such as bottom-top module precedence relationships, neighbor module precedence relationships, and module blocking precedence relationships. Furthermore, planners must balance the use of larger cranes, which minimize the number of foundation preparation, movement, and setup, with the use of smaller cranes, which minimize hourly crane rental and supporting task costs (Lin and Haas 1996). Typically, mobilization and demobilization costs are determined upon crane selection, while reconfiguration and relocation costs depend on module installation plans.

Currently, schedulers manually sequence module installation and constraint satisfaction using a lengthy trial-and-error-based approach. Notably, any attempts to reduce crane operation costs must, in turn, be evaluated for feasibility and for satisfaction of technological constraints. Development of a tool capable of automatically generating and optimizing module installation plans would considerably reduce the time-intensiveness and laboriousness associated this process.

The installation planning problem for industrial modules is a resource-constrained project scheduling problem (RCPSP). Alternative schedules are available for sequencing module installation using limited crane resources. The RCPSP is a non-deterministic polynomial hard (NP-hard) problem: in essence, there is no computationally feasible algorithm that guarantees that the problem can be solved to global optimum for a project with practical size and complexity. For example, formulation of an installation plan for 60 modules with 1500 feasible crane locations for lifting each module amounts to over 10^{210} possible solutions.

In the existing body of knowledge, there is no formalized approach for determining the sequence of module installation. As such, this paper proposes a novel planning methodology to produce feasible installation

plans for modular construction. The proposed approach uses heuristic rules to formulate module installation plans that both minimize cost and ensure satisfaction of several technological constraints. Due to the complicated nature of iterating the module installation sequence, the development of a decision-support tool for automatically generating the module installation plan is explored, and a prototyped software system that automates the iterative process is presented.

2. LITERATURE REVIEW

Multiple design and planning activities, such as the selection of rigging design and crane model, ground bearing pressure calculations, crane location and configuration assignments for each module or vessel, path planning, access planning, and formulating an overall lifting plan (Haas and Lin 1995, Lei et al. 2015), are required to ensure lift safety and feasibility on industrial construction projects. Researchers have developed computer-aided planning tools to facilitate many of these decision-making processes. Previous research has focused on the automation of mobile crane design and planning activities on industrial sites. For instance, Hornaday et al. (1993) and Al-Hussein et al. (2001, 2005) have developed computer-aided systems to automatically identify potential crane locations based on crane capacity, lifting range, and crane utilization percentage. Haas and Lin (1995), Reddy and Varghese (2002), Olearczyk (2014), Lei et al. (2013), and Lei et al. (2015) analyzed the lifting, swinging, and placement of a single object, and automated clash detection based on site constraints and crane configuration. Lei et al. (2014) and Han et al. (2014) analyzed crane walking paths for instances where a crane picks up and travels with an object (e.g. modules, equipment, or vessels) before placing the object in its final position. Hermann et al. (2010) and Olearczyk et al. (2015) proposed incorporating the above analyses in an integrated software platform for preparing engineered lift drawings and detecting potential on site conflicts in consideration of crane capacity, object weight, rigging requirements, and site constraints. Industry has also developed in-house planning tools. For example, the Automated Lift Planning System (ALPS) developed by Bechtel can be used to provide a visualization environment for each lift (William and Bennett 1996).

In addition, some researchers have attempted to automate the planning and scheduling of multiple lift sequences. Lin and Haas (1996) proposed an interactive platform that allows the selection of an optimum schedule for a single crane using linear programming. Lin and Haas (1996) have also proposed a semi-automated approach for the formulation of a lifting schedule for one crane. However, they did not consider crane reconfiguration or site constraints (e.g. top-bottom module relationships) that can impact project cost and duration. Reddy et al. (2007) presented a multiple lift planning tool that visualizes the simulation of an installation schedule of heavy vessels in accordance with particular crane types and site locations. Taghaddos et al. (2011) optimized

crane lift schedules using an ascending auction protocol. When computer-aided planning tools are not utilized, planning multiple heavy lifts in a congested industrial site is complicated, error-prone, and time consuming (Olearczyk et al. 2015).

In current practice, practitioners plan the lifting sequence in a semi-automatic manner using a heuristic rules approach (e.g., minimizing the number of crane relocations). Solutions are manually determined using a trial-and-error method based on the experience and expertise of the planners (Hermann et al. 2010). In an iterative process, the subject matter expert (SME) chooses the most critical modules (in terms of weight and size) to be processed and determines a crane location for the selected modules that could be used for future module installation. If any previously established location can be used to lift the current module, it is selected over a new crane location. This process is repeated until a feasible crane schedule for the project is determined.

The module planning problem is a combinatorial type problem with multiple technological constraints, where the objective is to devise a feasible installation plan for which the cost of crane operation is minimized. The size of the project, however, limits the ability to approach this type of problem by carrying out an exhaustive search. A medium-sized project with 60 modules and approximately 1,000 potential crane locations for each module has as many as 8.3×10^{261} ($1000^{60} \times (60\text{-}1)!$) search options.

Heuristic algorithms have been commonly proposed and used in literature to solve complex, large NP-hard combinatorial type problems, such as RCPSP and VRP. Kolisch and Hartmann (1999) have acknowledged that heuristic procedures are essential for solving large, practical, NP-hard problems. The heuristic procedure has been proposed to solve RCPSP (Boctor 1990, Kolisch and Hartmann 2006) and VRP (Laport et al. 2000) in situations where the exact method fails to provide an optimal solution in a reasonable timeframe.

The use of heuristic, knowledge-based approaches to plan construction tasks are also common. Knowledge-based schedule generation tools, which incorporate automated mechanisms to ensure constraint satisfaction, have been developed and recommended to facilitate installation planning and eliminate errors. Applications of these tools can be found in building construction (Koo et al. 2007; Chen et al. 2013), offshore platform installation (Hendrickson et al. 1987), and bridge construction (Wu et al. 2010).

Given the scope and difficulty of the problem, the goal of this paper is to propose an algorithm and framework that automates the search and generation of a near-optimum module installation plan for medium- to large-sized problems and has several project-associated technological constraints embedded into the provided solution. This research builds on the previous work of Al-Hussein et al. (2001, 2005) and Hermann et al. (2010), which identify potential lifting options for individual objects, and that of Lei et al. (2013) and Lei et al. (2015), which analyze clash detection automation of lifting

single objects, to propose a framework for site-wide installation planning that considers multiple installation options and constraints.

3. PROPOSED METHODOLOGY

The proposed approach generates a module installation sequence for modular construction that is based on (i) a list of modules, (ii) rigging requirements for module installation, (iii) crane availability, and (iv) available crane configurations (e.g., boom length, superlift type, and superlift weight). The proposed methodology also considers three crane/module technological constraints:

1. The bottom-top module finish-to-start precedence relationship, which exists between the lower and upper modules during installation,
2. The neighbor module precedence relationship, which dictates that any module belonging to a defined module group cannot be installed between two previously installed modules from the same module group. This constraint is illustrated in Figure 1.
3. The module-blocking precedence relationship, which dictates that previously installed or earlier-sequenced module installation eliminates certain installation options for to-be-sequenced modules. Figure 2 demonstrates this precedence relationship.
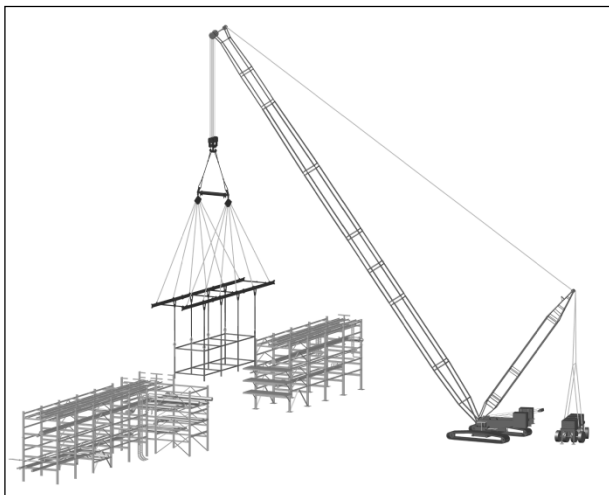
Figure 1: Neighbor Module Precedence Relationship

To generate the solution, heuristic rules are proposed to determine possible installation sequences factoring in feasible installation options. Based on the formalized approach, the module installation plan minimizes the number of crane locations, crane relocations, crane reconfigurations, and crane moving distances. The following subsections discuss the inputs, processes, and outputs of the proposed approach.
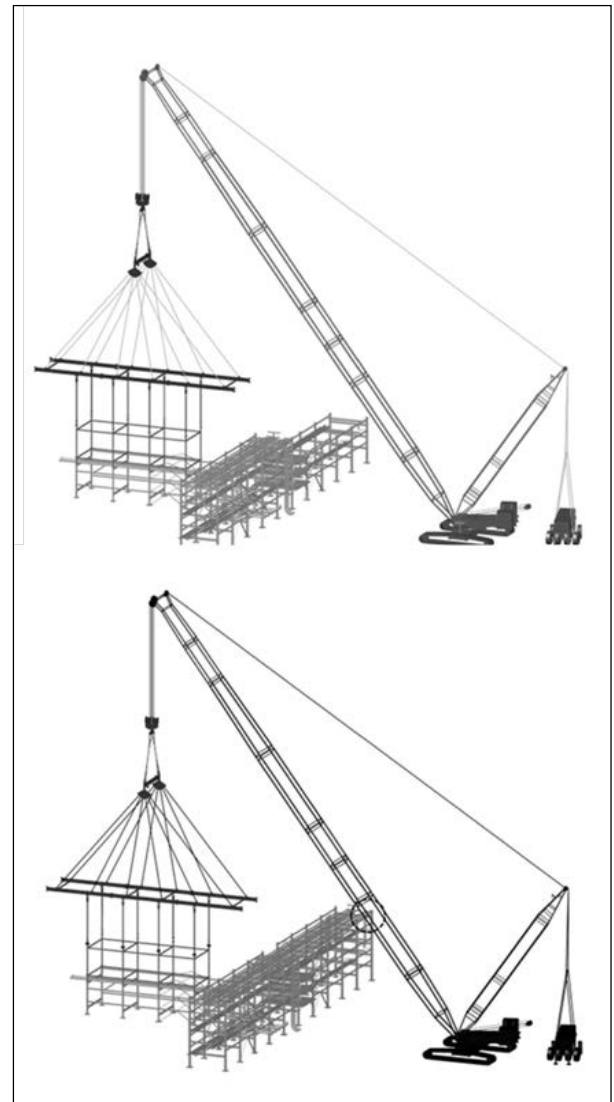
Figure 2: Module Blocking Precedence Relationship

3.1. Input

Various project details and constraints must be considered to generate a practical module installation plan. The following inputs are minimum requirements: (i) feasible crane configurations, (ii) feasible crane location coordinates associated with individual modules, and (iii) module installation precedence relationships, as discussed in Section 3. As mentioned in Section 2, feasible crane configurations and crane locations for each module are determined using previously developed tools. Input information is assumed to be available in a database or in a format that can be easily converted to a database.

3.2. Process

Given the inputs, a feasible plan for module installation is formulated based on the proposed iterative procedure. The process consists of (i) feasible solution generation, (ii) solution ranking, and (iii) solution selection using the proposed heuristic rules. Figure 3 summarizes the methodology for the overall process. The proposed procedure consists of the following steps:
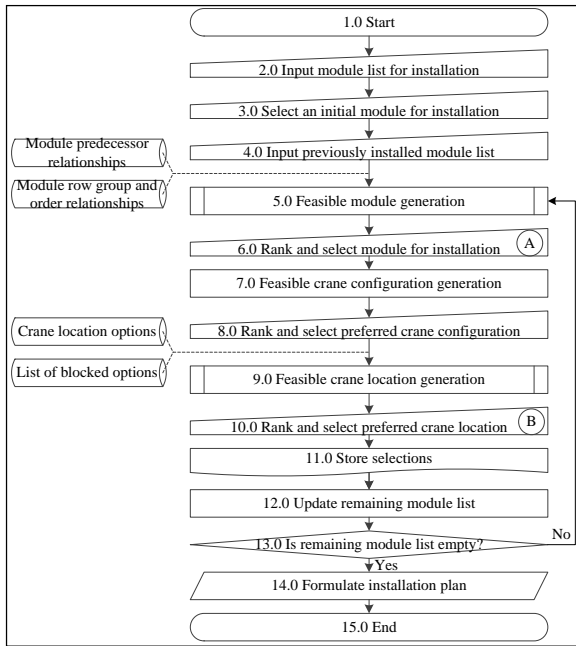
Figure 3: Overall Methodology Flowchart

1. *Start:* Begin the planning session.
2. *Input module list for installation:* User selects and lists the modules for installation planning from project module list.
3. *Select an initial module for installation:* User specifies the first module for installation in Installation Iteration 1.
4. *Input the previously installed module list:* User selects the installed module(s) from the module list for the project.
5. *Feasible module generation:* Algorithm generates a list of the to-be-installed modules in consideration of module precedence relationships and the list of previously installed modules.
6. *Rank and select the module for installation:* Algorithm ranks the modules in the to-be-installed module list prepared in Step 5 based on multiple criteria. Algorithm or user can select the module with the lowest ranking for installation. The flowchart in Figure 4 demonstrates the heuristic ranking rules.
7. *Feasible crane configuration generation:* Algorithm lists the feasible crane configurations for installing the selected module.
8. *Rank and select crane configuration:* Algorithm ranks the feasible crane configurations based on two criteria: (1) check if the previous crane configuration can be reused and (2) check how many modules the crane configuration can be used to install the to-be-sequenced modules if the previous crane configuration cannot be reused.
9. *Feasible crane location generation:* Algorithm prepares the location list for the selected module and crane configurations in consideration of feasible crane lifting options and module-blocking precedence relationships.
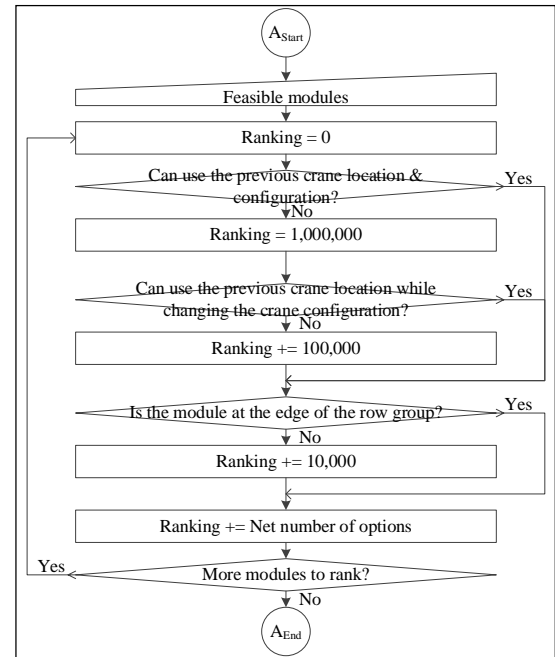
Figure 4: Flowchart of Step 6, Ranking Modules

TNM = Total Number of Modules" in the project, FIM = "Future Installable Module" from the same crane location, TIM = "Total Installable Module" from the same crane location, and MD = "Moving Distance" of the crane from the last assigned location.
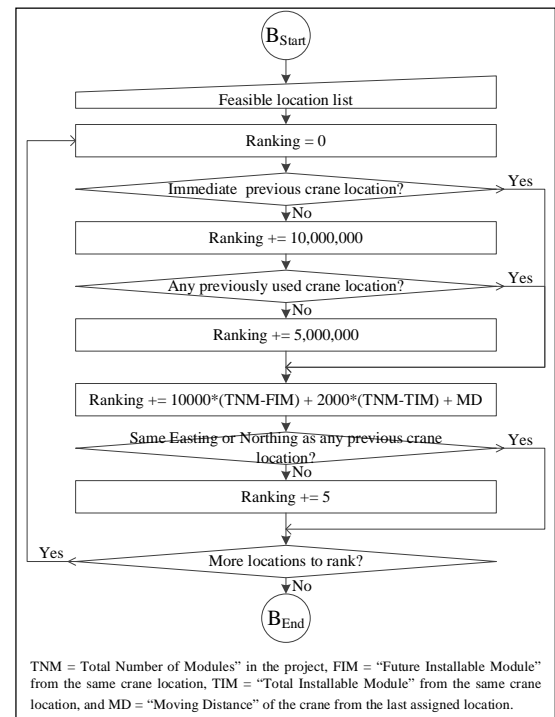
Figure 5: Flowchart of Step 10, Ranking Crane Location for Heuristic-Based Methodology

10. *Rank and select the crane locations:* Algorithm ranks crane locations from Step 9 based on the ranking process flowchart shown in Figure 5.
11. *Store selections:* Algorithm stores the selected module, crane configuration, and crane locations as part of the installation plan for this installation iteration.
12. *Update remaining module list:* The selected module is moved from the to-be-sequenced module list to the sequenced module list.

13. *Is the remaining module list empty?* Algorithm checks if all the modules have been sequenced. If not, it repeats Steps 5 through 12.
14. *Present the installation plan:* Once all module installations are sequenced, the final installation plan is presented.

3.3. Output
The output of the proposed methodology is a module installation plan. The installation plan specifies the installation sequence, the crane configuration, and crane location for installing each module. As a result of the provided module installation plan, the precedence relationships for installing the modules are satisfied, while the costs of crane foundation preparation, crane relocation, and crane reconfiguration are minimized.

4. SAMPLE CASE STUDY
In this section, a sample case study is used to explain the calculation procedures of the proposed methodology. Figure 6 illustrates the postulated site layout for module installation. In this problem, there are 8 modules to be sequenced using one of 9 available crane locations. The crane assigned to this project can occupy any of the two configurations described in the following section.
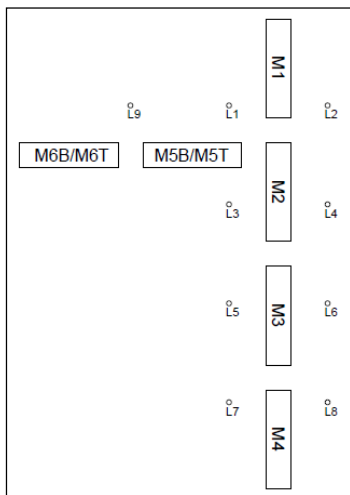
Figure 6: Site Layout for Sample Case Study

4.1. Input
Table 1 demonstrates that bottom-top module precedence relationships exist between M5B and M5T and between M6B and M6T. Table 2 details the existing neighbor module precedence relationships. M1, M2, M3, and M4 are classified as a group of modules with neighbor precedence relationships (G1); M2, M5B, and M6B are classified as another group (G2) with neighbor precedence relationships. The Module Order in each group represents the module location with respect to other modules of the same group. Modules with subsequent order numbers are immediate neighbors. For example, M2, with order 2, is the immediate neighbor of M1, with order 1, as well as M3, with order 3. Table 3 demonstrates three inputs: (i) all feasible crane configurations for lifting the module at each location, (ii)

all feasible crane location options for lifting each module, and (iii) the module blocking precedence relationships associated with each lifting option. For example, L1 location is disallowed for lifting M2 with C1 configuration if M1 or M5 have been previously installed.

Table 1: Bottom-Top Precedence Relationships in the Sample Case Study

Module	Predecessor
M1	–
M2	–
M3	–
M4	–
M5B	–
M5T	M5B
M6B	–
M6T	M6B

Table 2: Neighbor Module Precedence Relationships in the Sample Case Study

Module	Group	Order	Group	Order
M1	1	1	–	–
M2	1	2	2	3
M3	1	3	–	–
M4	1	4	–	–
M5B	–	–	2	2
M6B	–	–	2	1

4.2. Process
Using the above input data, the methodology process outlined in Section 3.2 is applied to formulate the module installation plan. The following should be noted regarding the process steps:

1. For *Step 3: Select an initial module*, M1 is selected. Notably, this selection may affect the optimality of the final plan. In practice, the first module to be installed is determined by project planners based on module delivery schedules, criticality of the modules, or SME experience. It may be beneficial to generate the module installation plan by setting various starting modules to achieve solutions with global optimality.
2. For *Step 5: Feasible module generation*, in the first iteration, M5T and M6T are eliminated since bottom-top module constraints are not satisfied.
3. After the first module is sequenced, Steps 4 through 12 are repeated 7 times to plan and sequence the installation of all remaining modules.

4.3. Output
Table 4 summarizes the installation plan obtained after completing the above process for 8 installation iterations. The plan provides the installation sequence for the modules and specifies the crane location and configuration for lifting each module. The plan minimizes the number of crane foundations, relocations,

and configurations used, as well as the crane travel distance.

Table 3: Module Lifting Option and Blocking Precedence Relationships for the Sample Case Study

Module	Crane Configuration	Crane locations	Blocking modules
M1	C1	L1	–
M1	C2	L1	–
M1	C1	L2	–
M1	C2	L2	–
M2	C1	L1	M1, M5T
M2	C1	L2	M1
M2	C1	L3	–
M2	C2	L3	–
M2	C1	L4	–
M2	C2	L4	–
M3	C1	L3	M2
M3	C1	L4	M2
M3	C1	L5	–
M3	C2	L5	–
M3	C1	L6	–
M3	C2	L6	–
M3	C1	L7	M4
M3	C1	L8	M4
M4	C1	L7	–
M4	C2	L7	–
M4	C1	L8	–
M4	C2	L8	–
M5B	C1	L1	–
M5B	C2	L1	–
M5B	C1	L2	M1, M2
M5B	C1	L3	–
M5B	C2	L3	–
M5B	C1	L9	–
M5T	C1	L1	–
M5T	C2	L1	–
M5T	C1	L2	M1, M2
M5T	C1	L3	–
M5T	C2	L3	–
M5T	C1	L9	–
M6B	C1	L1	M5T
M6B	C1	L9	–
M6T	C1	L1	M5T
M6T	C1	L9	–

Table 4: Final Installation Plan for the Sample Case Study

Installation Iteration #	Module	Crane Configuration	Crane Location
1	M1	C1	L1
2	M6B	C1	L1
3	M6T	C1	L1
4	M5B	C1	L1
5	M5T	C1	L1
6	M2	C1	L3
7	M3	C1	L7
8	M4	C1	L7

4.4. Method Validation

To validate the plan generated in Section 4.3, three validation techniques presented by Sargent (2005) are utilized. First, individual behavior, priority, and ranking of modules were traced to ensure method logic was correct. Secondly, an animation for installing the modules in accordance with the formulated installation plan was created, reviewed, and scrutinized to ensure crane-module technological constraints were satisfied. Figure 7 demonstrates enforcement of bottom-top precedence relationships. As shown in Figure 7, bottom module M6B is installed during Installation Iteration 2 prior to installation of top module M6T during Installation Iteration 3. Figure 8 demonstrates enforcement of neighbor module precedence relationships. After installing M6B during Installation Iteration 2, M5B is chosen for sequencing during Installation Iteration 4 (rather than M2). Finally, Figure 9 demonstrates enforcement of the module blocking precedence relationship. Given that M2 could be installed from L1 (Table 3), the sequenced M5T blocks the path for installing M2 from L1. As such, the crane must be moved to a new location before installing M2 during Installation Iteration 6. Finally, face validation, where two knowledgeable individuals are asked to validate both the method behavior and result, was completed. Since the size of the problem was small, optimality of the solution was manually reviewed; given the project input, a more optimal solution could not be found.
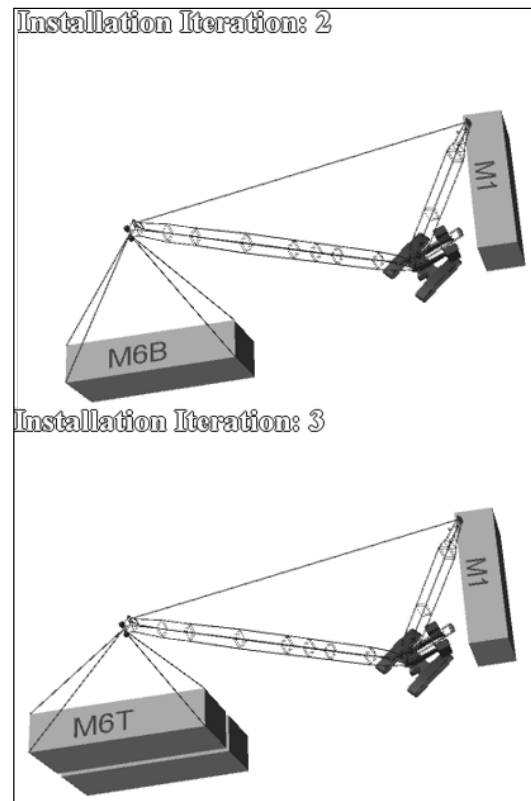
Figure 7: Validation of the Bottom-Top Module Precedence Relationship
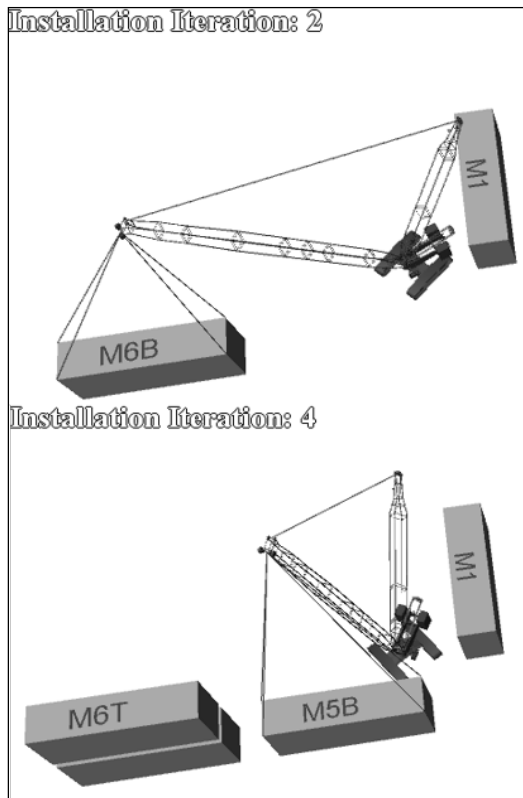
Figure 8: Validation of the Neighbor Module Precedence Relationship
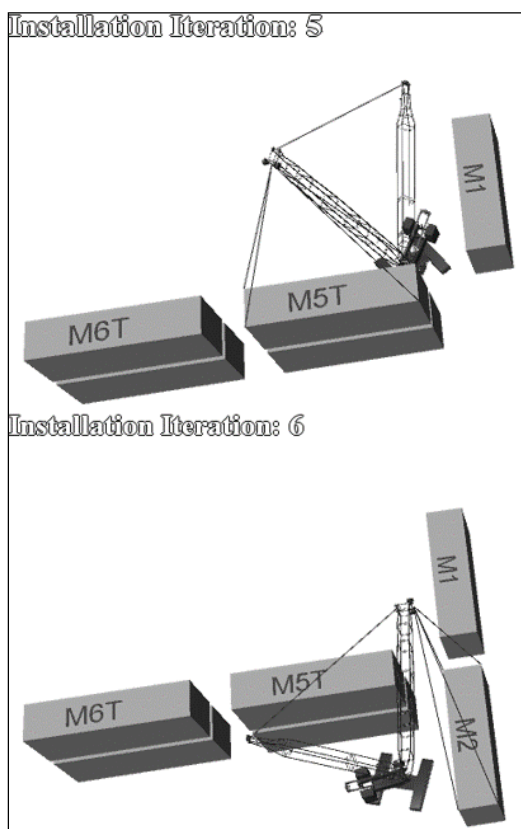
Figure 9: Validation of the Module Blocking Precedence Relationship

5. PRACTICAL CASE STUDY

In this section, a practical case study is presented to demonstrate the ease of obtaining an automated solution in practice. The plan is then compared to the plan generated by industry practitioners using an experience-based approach.

Figure 10 illustrates the designated module layout. The project consisted of 68 modules. The module types included pipe rack, electrical, building, and equipment modules. Module weights ranged from 20,000 to 200,000 pounds, and module lengths from 18 to 36 meters. Three groups of straight run modules, where neighbor module precedence relationships existed, were identified. Multiple areas that contained two or more modules stacked on top of each other, where bottom-top module precedence relationships existed, were detected.
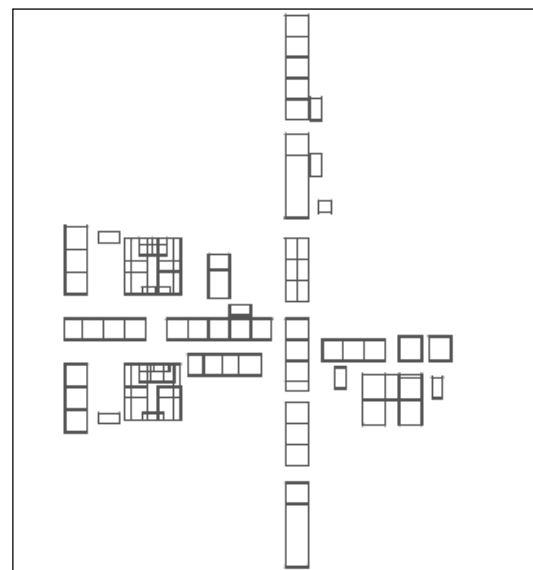
Figure 10: Site Layout for Installation of 68 Modules of Practical Case Study

On average, there were approximately 3,000 crane locations available for installing each module, with a total number of 200,000 options for installing all 68 modules. These installation options, as well as the module blocking precedence relationships, were generated using the previously developed program ACPO (Hermann et al. 2010). For example, Figure 11 illustrates possible crane locations, represented as points 3 feet apart, for installing one module. These locations are shown regardless of the module blocking precedence relationship. It is assumed that one crane was used to install all modules. The installation plan was evaluated based on the number of distinct crane locations, crane relocations, crane reconfigurations, and the total crane movement distance.
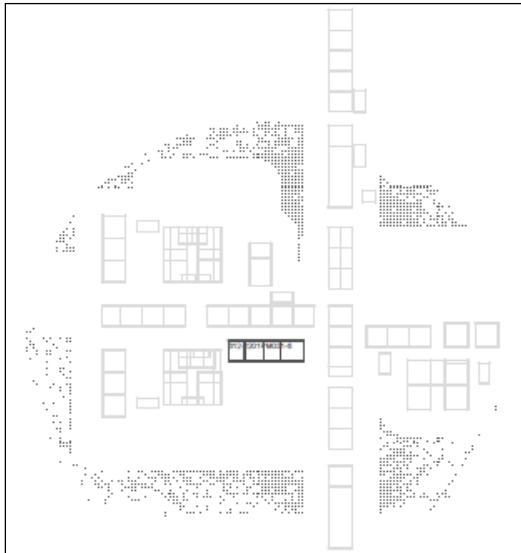
Figure 11: Possible Crane Locations for Installing One Module for Practical Case Study

Visual Basic for Application (VBA) in MS Access was used to implement the algorithm. Figure 12 provides an overview of the various steps of solution preparation process using the automation tool. In Step 1, project details are provided in the form of an information database. In Step 2, the project module list, obtained from the input information, is displayed. The user then specifies which modules are previously installed and which modules are to be sequenced for installation. Next, by clicking the "Start Planning" button, a list of feasible modules for Installation Iteration 1 are generated and ranked. In Step 3, the user selects a module as the first module for Installation Iteration 1 and clicks the "Generate Solution" button. In Step 4, solutions are iterated using the proposed methodology. Alternatively, the user can select to navigate through the installation iteration on a step-by-step basis using the "Next" and "Back" buttons. In Step 5, the solution is generated and stored in the database.

Using the methodology outlined in Section 3, a feasible solution was found for the practical problem. Figure 13 details part of the solution stored in the database. The plan indicated that module installation could be completed using 4 distinct crane foundation locations, 3 crane relocations, and a total of 898 feet of crane travel movement. Notably, crane locations within a 45-foot radius were assumed to make use of the same crane foundation. Relocation was considered to have occurred when the crane was required move a distance of 45 feet or more to a new location.

While the user or SME were heavily involved in method development, three validation methods stated in Section 4.4 were also used to ensure model correctness for the large-size practical problem. Validation methods included tracing the individual module behavior during the process, generating an animation illustrating the module installation plan, and face validation by knowledgeable individuals. A screenshot of the

animation at Installation Iteration 61 is displayed in Figure 14.
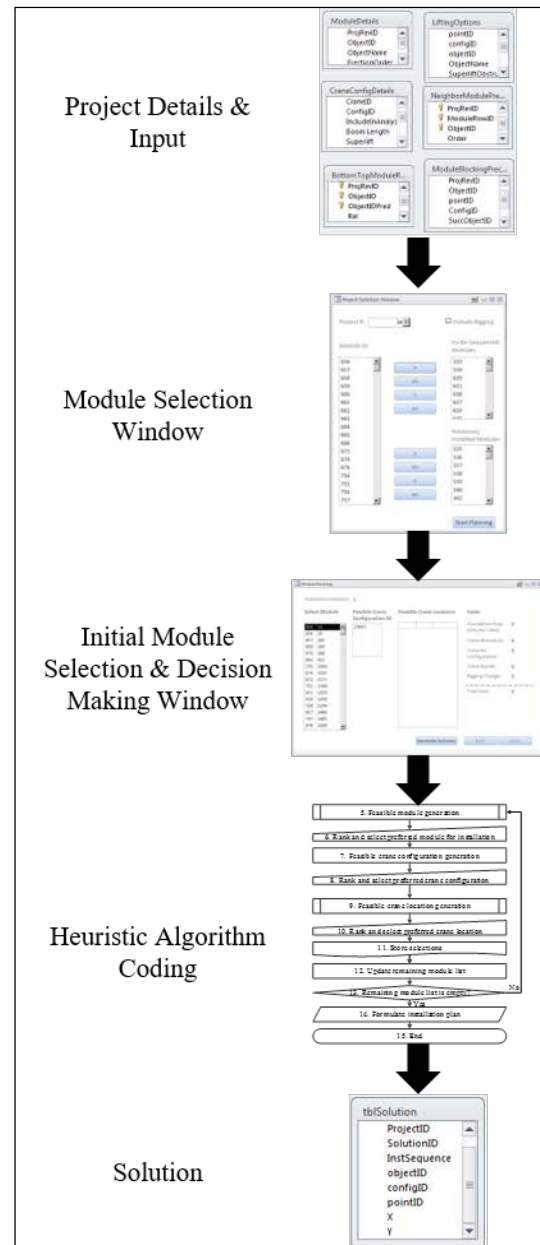
Figure 12: Planning Process using the Developed Automation Tool

Figure 13: Details of the Heuristics-Based Solution Stored in the Database
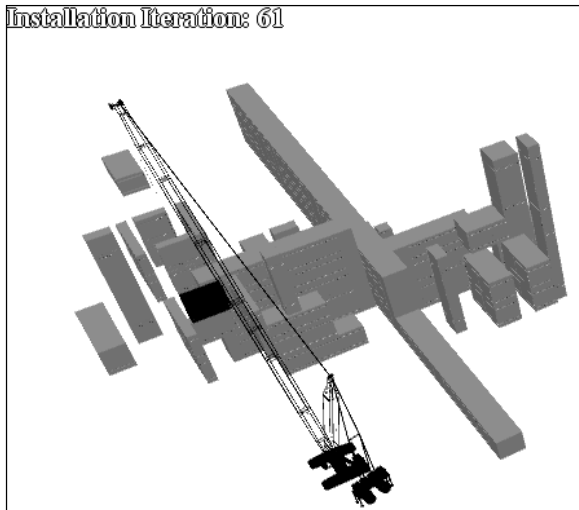
Figure 14: A Screenshot of the Animation of the Installation Plan Generated for the Practical Case Study

Table 5 compares the solution generated by the proposed methodology with the solution provided by the practitioners. In the experience-based installation plan, 8 distinct crane foundations, 14 crane relocations, and a total of 2270 feet of crane movement were required to complete module installation. The proposed installation plan based on the methodology developed here outperformed the experience-based installation plan by reducing the number of crane foundations, crane relocations, and total crane travel required.

Table 5: Comparison of heuristic-based decision support tool and the experience-based installation plan solutions

Comparison Item	Proposed Plan	Experience-based Plan
Number of Crane Foundations	4	8
Number of Crane Relocations	3	14
Total Crane Travel Distance (ft.)	898	2270

6. SUMMARY AND FUTURE WORK

Currently, there is no formalized framework or methodology in place for preparing and automating multi-lift site plans for modular construction. Current practice, which is based on trial-and-error-based approaches, is time-consuming and error-prone. The novel methodology presented in this work can be used to automate module installation planning in practice. The methodology developed uses project information (e.g. list of modules, module rigging requirements, crane availability, and available crane configurations) as inputs, enforces crane-module technological constraints, and ranks the sequencing options using heuristic-based rules. This facilitates the scheduling tasks involved in preparing an error-free plan for module installation on site. The proposed methodology ensures that a feasible installation plan is generated while minimizing crane operation costs by means of heuristic rules. The plan feasibility is ensured by enforcing: (i) bottom-top module precedence relationships, (ii) neighbor module precedence relationships, and (iii) module blocking precedence relationships. Crane operation costs are reduced using heuristic rules that minimize the number of distinct crane locations, number of crane relocations, and number of crane reconfigurations. A software system prototype was developed by implementing the proposed methodology using VBA for MS Access and was used to automatically schedule a real-world modular construction project. The software system developed effectively prepared a module installation plan that satisfied all indicated constraints.

There are four advantages to using the software tool developed in this work. First, the software system allows the planner to choose preferred installation options in terms of module installation sequence and crane location. Secondly, when the installation plan changes, the software allows the planner to investigate a potential path forward and update the project schedule. Thirdly, the software system also reduces the burden on the project team by ensuring all constraints are checked and satisfied. This represents a considerable advantage due to the large amount of project information and interdependency. Finally, the creativity and expertise of the planner can be incorporated in the planning and sequencing of module installation.

The work presented in this paper is limited to the use of one crane on site. Also, it does not consider the different rigging requirements of various modules. While the program can track the number of rigging changes required to complete module installation, it lacks the ability to minimize rigging changes while planning the module installation sequence. In future research, the proposed methodology can be expanded to include scenarios where multiple cranes are used simultaneously on site. Also, the possibility of taking into account module rigging requirements when sequencing module installation can be explored. Finally, the methodology can also be expanded to allow the preparation of an installation schedule with specific dates for module installation by incorporating the project start date and other project constraints, such as module delivery dates, into the input database.

REFERENCES

Al-Hussein M., Alkass S., Moselhi, O., 2001. An algorithm for mobile crane selection and location on construction sites. Construction Innovation, 1(2), 91-105.

Al-Hussein M., Alkass S., Moselhi, O., 2005. Optimization algorithm for selection and on site

location of mobile cranes. Journal of Construction Engineering and Management, 131(5), 579-590.

Boctor, F.F., 1990. Some efficient multi-heuristic procedures for resource-constrained project scheduling. European Journal of Operational Research, 49(1), 3-13.

Chen S.M., Chen P.H., Chang L.M., 2013. A framework for an automated and integrated project scheduling and management system. Automation in Construction, 35, 89-110.

Haas C.T., Lin K., 1995. An interactive database system with graphical linkage for computer aided critical lift planning. Proceedings of the 12th International Symposium on Automation and Robotics in Construction, pp. 313-324. Warsaw (Poland).

Han S., Lei Z., Bouferguène A., Al-Hussein M., Hermann U., 2014. Integrated visualization and simulation for lifting operations of modules under congested environment. Proceedings of the 31st International Symposium on Automation and Robotics in Construction, pp. 262-269. Sydney (Australia).

Hendrickson C., Zozaya-Gorostiza C., Rehak D., Baracco-Miller E., Lim P., 1987. Expert system for construction planning. Journal of Computing in Civil Engineering, 1(4), 253-269.

Hermann U., Hendi A., Olearczyk J., Al-Hussein M., 2010. An integrated system to select, position, and simulate mobile cranes for complex industrial projects. Construction Research Congress, pp. 267-276. Banff (Alberta, Canada).

Hornaday W.C., Haas C.T., O'Connor J.T., Wen J., 1993. Computer-aided planning for heavy lifts. Journal of Construction Engineering and Management, 119(3), 498-515.

Kolisch R., Hartmann, S., 1999. Heuristic algorithms for the resource-constrained project scheduling problem: Classification and computational analysis. In: Weglarz J., ed. Project Scheduling. New York, NY: Springer Science+Business Media, 147-178.

Kolisch R., Hartmann S., 2006. Experimental investigation of heuristics for resource-constrained project scheduling: An update. European Journal of Operational Research, 174(1), 23-37.

Koo B., Fischer M., Kunz, J., 2007. A formal identification and re-sequencing process for developing sequencing alternatives in CPM schedules. Automation in Construction, 17(1), 75-89.

Laporte G., Gendreau M., Potvin J.Y., Semet, F., 2000. Classical and modern heuristics for the vehicle routing problem. International Transactions in Operational Research, 7(4-5), 285-300.

Lei Z., Han S., Bouferguène A., Taghaddos H., Hermann U., Al-Hussein, M., 2014. Algorithm for mobile crane walking path planning in congested industrial plants. Journal of Construction Engineering and Management, 141(2), 05014016.

Lei Z., Taghaddos H., Han S., Bouferguène A., Al-Hussein M., Hermann U., 2015. From AutoCAD to 3ds Max: An automated approach for animating heavy lifting studies. Canadian Journal of Civil Engineering, 42(3), 190-198.

Lei Z., Taghaddos H., Olearczyk J., Al-Hussein M., Hermann U., 2013. Automated method for checking crane paths for heavy lifts in industrial projects. Journal of Construction Engineering and Management, 139(10), 04013011.

Lin K.L., Haas, C.T., 1996. An interactive planning environment for critical operations. Journal of Construction Engineering and Management, 122(3), 212-222.

Olearczyk J., Bouferguène A., Al-Hussein, M., Hermann, U.R., 2014. Automating motion trajectory of crane-lifted loads. Automation in Construction, 45, 178-186.

Olearczyk J., Lei Z., Ofrim B., Han S., Al-Hussein, M., 2015. Intelligent Crane Management Algorithm for Construction Operation (iCrane). Proceedings of the 32nd International Symposium on Automation and Robotics in Construction, pp. 1-8. Oulu (Finland).

Reddy H.R., Varghese K., 2002. Automated path planning for mobile crane lifts. Computer-Aided Civil and Infrastructure Engineering, 17(6), 439-448.

Reddy S.D., Varghese K., Srinivasan N., 2007. A computer-aided system for planning and 3D-visualization of multiple heavy lifts operations. Proceedings of the 24th International Symposium on Automation and Robotics in Construction, pp. 281-288. Kochi (India).

Sargent, R.G., 2005. Verification and validation of simulation models. Proceedings of the 2005 Winter Simulation Conference, pp. 130-143. Orlando (Florida, USA).

Taghaddos, H., AbouRizk, S., Mohamed, Y., & Hermann, U., 2010. Simulation-based multiple heavy lift planning in industrial construction. Construction Research Congress, pp. 349-358. Banff (Alberta, Canada).

Williams M., Bennett C., 1996. ALPS: the automated lift planning system. Computing in Civil Engineering, pp. 812-817. Anaheim (California, USA).

Wu I.C., Borrmann A., Beißert U., König M., Rank E., 2010. Bridge construction schedule generation with pattern-based construction methods and constraint-based simulation. Advanced Engineering Informatics, 24(4), 379-388.

THE OPEN ARCHITECTURE SCHEDULING SYSTEM FOR A SINGLE-ARMED CLUSTER TOOL WITH PM CLEANING OPERATIONS

Dong-Hyun Roh[(a)]**, Tae-Eog Lee**[(b)]

[(a),(b)] Department of Industrial & Systems Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea

[(a)]rdh@simlab.kaist.ac.kr, [(b)]telee@kaist.ac.kr

ABSTRACT

Cluster tools are a type of widely used semiconductor manufacturing equipment. Generally, a cluster tool is operated on a built-in schedule; however, it is impossible to modify or change the built-in schedule because of the closed architecture of the scheduling system for the tool. In this study, we propose a framework for an open architecture scheduling system for a single-armed cluster tool with PM cleaning operations. The scheduling system works by scheduling command files that can be modified or replaced from outside. As an application of the framework, performance comparison analysis between the backward and backward(z) sequences in a single-armed cluster tool with multi-period PM cleaning operations is conducted.

Keywords: cluster tool, open architecture scheduling system, virtual cluster tool, PM cleaning operation

1. INTRODUCTION

In modern semiconductor manufacturing systems, about 300 processes are required for the fabrication of a semiconductor product. Most processes, including etching, vapor deposition, and wafer cleaning, are performed using cluster tools (Yu and Lee 2017). Cluster tools are the most popular type of configurable semiconductor manufacturing equipment; they consist of several single-wafer process modules (PMs) and a material-handling robot called a transporting module (TM). PMs are modular and can be detached and attached; TM is responsible for transporting wafers. Usually, a cluster tool is operated according to a specified wafer flow pattern and several operational constraints. There have been numerous studies on the operation of cluster tools for various wafer flow patterns and constraints (Venkatesh et al. 1997; Geismar 2004; Lee et al. 2014; Wu et al. 2011; Kim et al. 2015; Kim et al. 2013; Kim et al. 2013; Yu and Lee 2017; Rostami et al 2001; Kim et al. 2003; Wu 2008).
Cluster tools have diverse structures depending on the types of PM and TM; structures include single-armed and dual-armed cluster tools, multi-slot cluster tools, and in-line multi-cluster tools (Lee 2008). Inter alia, the most commonly used structure is a radial cluster tool, in which PMs radially surround the tool. A radial cluster tool is called a single-armed or dual-armed cluster tool according to the number of robot grippers. It is known that the backward sequence and the swap sequence are optimal robot task sequences for the single-armed and dual-armed cluster tools with series-parallel wafer flow patterns, respectively (Lee 2008).
Generally, a semiconductor fabrication plant, a so-called "fab", purchases semiconductor manufacturing equipment including cluster tools from a variety of equipment manufacturing companies. Therefore, most cluster tools operate on the schedule built into the tool by the equipment manufacturing company. The built-in schedule usually consists of a benchmark schedule mainly based on the backward or the swap sequence and several exception-handling techniques.
However, there exists an issue that built-in schedules and tool schedulers are different for each equipment manufacturing company. In addition, a tool scheduler has a closed architecture; in other words, only the equipment manufacturing company can modify the internal structure and operating logic of the scheduler. This means that a fab cannot arbitrarily modify a tool schedule to induce high productivity and exception-handling techniques for quality control. Since, in recent years, the method of semiconductor production has changed from the existing high-volume low-mix production to high-mix low-volume production, a tool scheduler with closed architecture leads to difficulty in optimizing the operation of the tool in situations in which frequent schedule changes are needed. Furthermore, there also exists a disadvantage that it is difficult to investigate whether the existing built-in schedule provides optimal productivity or, if not, which parts need to be improved.
Therefore, a system capable of changing a tool schedule from the outside as desired, called an open architecture scheduling system, is needed. In order to implement the open architecture scheduling system, a standard scheduling command scheme of cluster tools that all equipment manufacturing companies and fabs can understand is needed. If all cluster tool schedulers follow the same scheduling command scheme, it becomes quite

easy for a fab to replace the old tool schedule with a new one. This idea is similar to the NC programming language or the NC code used in CNC systems. Since the NC code has a standardized command scheme, a work schedule of a CNC machine can be easily changed from outside, even though the NC code is generated using a variety of CAD/CAM software.

However, semiconductor manufacturing equipment, including cluster tools, is much more sophisticated and has diverse structures and operational constraints depending on the processes. In addition, even equipment that performs the same process may have different structures depending on the equipment manufacturing company. Therefore, establishing a standard scheduling command scheme for cluster tools requires much expert effort in the industrial setting.

For this reason, this study proposes a framework for an open architecture scheduling system. As an example, we develop a virtual cluster tool (VCT) for a single-armed cluster tool with PM cleaning operations and apply the proposed framework to describe how the open architecture scheduling system works. The framework can be extended to cluster tools of various architectures.

In this paper, we first describe the open architecture scheduling system framework. The framework improves the basic concepts of the open scheduling architecture of the cluster tool proposed by Lee and Lee (2010). Then, we define the structure and operation of a single-armed cluster tool with PM cleaning operations. To apply the open architecture scheduling system framework, we also suggest a reference model of the VCT. As an application, we conduct a performance analysis for the backward and the backward(z) sequences suggested by Yu and Lee (2017).

2. ARCHITECTURE AND OPERATION OF A SINGLE-ARMED CLUSTER TOOL WITH PM CLEANING OPERATIONS

2.1. Architecture of a Single-armed Cluster Tool
Cluster tools have a variety of structures. Among them, the single-armed cluster tool having a single-armed transporting module and radially located PMs is one of the most used pieces of manufacturing equipment in leading fabs. A single-armed cluster tool consists of a TM with a single gripper, several, usually two to six, PMs, two loadlock modules for loading a wafer cassette, and an aligner module for aligning a wafer unloaded from a loadlock (Lee 2008). Among the components, the TM and PMs are the most important for the operation.

2.2. Operation of a Single-armed Cluster Tool
When a cassette containing 25 wafers is loaded into a loadlock, a TM carries a wafer to the PMs one by one. The TM only performs wafer loading into the PM, wafer unloading from the PM, and moving operations. In addition, TM can perform one job at a time.

Each wafer is processed according to a predetermined recipe, which defines the sequence of process steps that the wafer should visit and the process time for each

process step. For each process step, to raise productivity, several PMs are assigned as parallel PMs. When a wafer is processed sequentially in a tool having parallel PMs, the wafer flow pattern is referred to as a series-parallel flow pattern. After all processes have been completed, the wafer is returned to the loadlock. This process continues until all wafers in the cassette have been processed.

In order for a wafer to be loaded into a PM, the chuck that serves to fix the wafer entering the PM should first be prepared. When the chuck is ready, the slit valve door opens and the TM loads the wafer. After wafer loading, the slit valve door is closed and the chuck firmly holds the wafer. Depending on the type of PM, pumping and venting operations may be required to prepare the process.

Therefore, the number of process steps, the number of parallel PMs for each process step, the process time for each PM, the required times for TM operations (loading, unloading, and moving), and the required task times of the slit valve door and chuck operations for each PM are needed to define the configuration of a single-armed cluster tool with a series-parallel flow pattern.
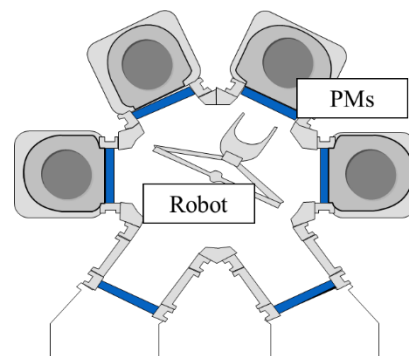
Figure 1: Single-armed Cluster Tool

2.3. Periodic PM Cleaning Operations
As the diameter of a wafer increases and the thickness of the wafer circuit becomes thinner, enhanced wafer quality control is required. Therefore, it becomes necessary to remove residual chemicals in the PM after the wafer process completes. A PM cleaning operation is a process of removing these impurities (Kim et al. 2013; Yu and Lee 2017). It is performed after a predefined number of wafer processes, which is called a cleaning period. During the cleaning operation, a wafer cannot be loaded.

In terms of operation, a cluster tool with periodic PM cleaning operations has differences from a typical single-armed cluster tool. To define an operation of a tool with periodic PM cleaning operations, the cleaning period and the cleaning time for each PM are additionally required. Consequently, the tool configuration is defined by information describing the tool operation which is shown in Table 1.
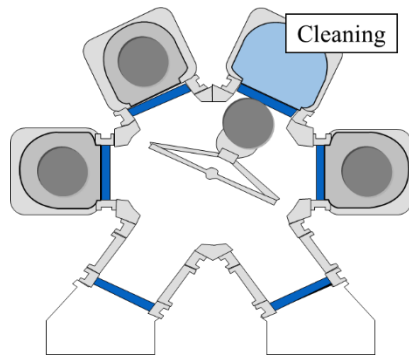
Figure 2: Single-armed Cluster Tool with PM Cleaning Operations

Table 1: Information Required to Define Tool Configuration

Type	Information Required to Define Tool Operation
Typical Single-armed Cluster Tool	Number of Process Steps
	Number of Parallel PMs for Each Step
	Process Time for Each PM
	Required Task Times for TM Operations (Loading, Unloading, Moving)
	Required Task Times for Slit Valve and Chuck
with PM Cleaning Operations	Cleaning Time for Each PM
	Cleaning Period for Each PM

3. OPEN ARCHITECTURE SCHEDULING SYSTEM OF SINGLE-ARMED CLUSTER TOOL WITH PM CLEANING OPERATIONS

The open architecture scheduling system of the cluster tool is a tool scheduler that can receive from the outside information necessary for scheduling. However, existing cluster tool controllers and schedulers have been developed independently for each equipment manufacturing company and it is not easy to implement them because there is no standardized scheduling model or rule. Therefore, a standard for the cluster tool scheduling commands should be given priority. It is difficult to standardize the scheduling commands of all cluster tools because of the various types of cluster tools depending on their usage, function, and required performance level. Furthermore, there exist differences in tools due to technology gaps among manufacturing companies. Consequently, this study suggests an open architecture scheduling system for a single-arm cluster tool with PM cleaning operations.

Scheduling a cluster tool requires three pieces of information: the tool architecture information, the benchmark TM task schedules, and exception-handling techniques. The tool architecture information is the information about the tool configuration and the tool operation, listed in Table 1. The benchmark TM task schedules refer to the work schedule of a TM when the cluster tool is in a steady state. A TM task schedule consists of a TM task sequence and a timing rule. Lastly, exception-handling techniques refer to TM task schedules when a tool is in transient state or when an exception occurs. The types of exceptions, criteria, and coping methods are included.

Lee and Lee (2010) suggested an XML-based file formats that contains integrated scheduling information, called the Scheduling Command File (SCF). In this study, we subdivide the SCF into three categories so that the schedule can be more effectively adjusted. Using SCFs, scheduling is performed by referencing the information from each SCF at each necessary moment. Each SCF can be replaced or modified as needed. Therefore, proposing standardized SCFs is the most important task in implementing the open architecture scheduling system. We briefly describe the information contained in each SCF and how the scheduler refers to each SCF.
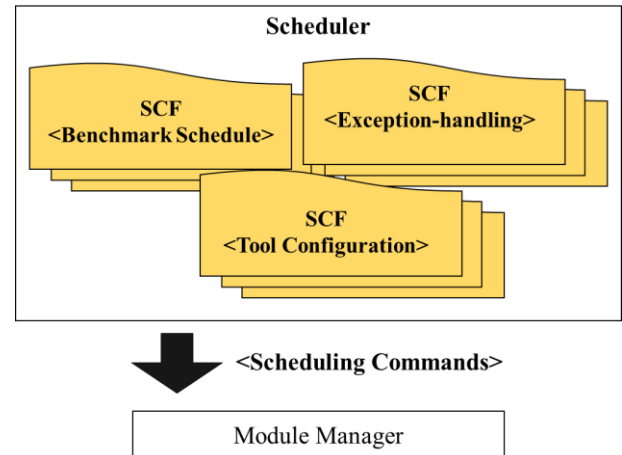
Figure 3: Scheduling Command Files in a Tool Scheduler

3.1. Tool Architecture Information SCF

In order to determine a tool schedule, the tool scheduler must first have information on the tool configuration. It also must have information on the constraints that must be met in the operation of the tool. A 'tool architecture information SCF' contains the information on tool configuration and operational constraints. In this study, the information is listed in Table 1.

3.2. Benchmark TM Task Schedules SCF

Most previous studies have expressed a benchmark schedule through a Petri net or a timed event graph (TEG) (Lee 2008). Therefore a 'benchmark TM task schedules SCF' contains the information for the Petri net model of the schedule, usually expressed by an incidence matrix and a token marking vector, the firing sequence, and the corresponding firing rule.

Among numerous benchmark TM task schedules, the most commonly used schedule in the single-armed cluster tool is the backward sequence and the earliest starting rule (Lee 2008). The earliest starting rule lets a TM perform a task as soon as possible.

3.3. Exception-handling Techniques SCF

An 'exception-handling techniques SCF' contains information about types of exceptions that can be controlled, a precise criteria for each exception, and a TM schedule for each exception. In the types of exceptions, not only hardware problems such as a PM or

a TM failure but also problems in a tool schedule such as transient periods of the tool or a K-periodic schedule are included. As cluster tools become increasingly sophisticated, more exceptions arise. There are also many cases in which the exceptions occur in combination. Therefore, it becomes an issue to establish precise criteria for all exceptions and corresponding coping methods.

4. DEVELOPMENT OF A VIRTUAL SINGLE-ARMED CLUSTER TOOL WITH PERIODIC PM CLEANING OPERATIONS

The open architecture scheduling system proposed in this study is a system capable of operating a tool through an external input schedule; thus, a fab can arbitrarily adjust the schedule. However, there exist significant time and money costs to apply the system to a real cluster tool. Therefore, we develop a VCT that operates in a manner similar to the actual cluster tool and apply the framework to the VCT. Among numerous studies about VCTs (Shin et al. 2001; Joo and Lee 2004; Kim and Lee 2013; Min and Lin 2013; Niedermayer and Rose 2003; Pan and Bao 2012), we implemented the virtual cluster tool based on the VCT framework suggested by Joo and Lee (2004).
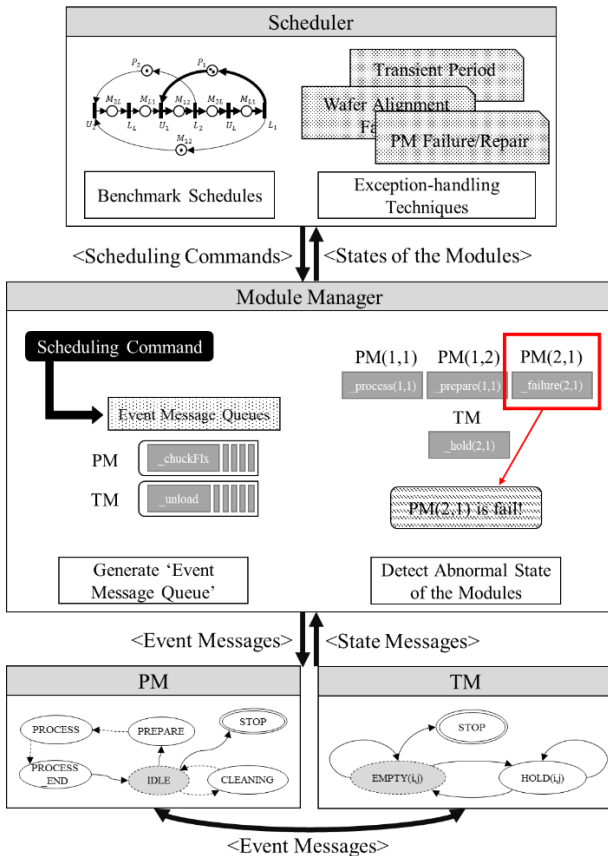
Figure 4: Reference Model of the VCT

4.1. Reference Model of the VCT
The core components of the cluster tool are PMs, a TM, the tool scheduler, and the module manager. The PM and the TM are devices that process and transport wafers; the tool scheduler is responsible for scheduling tasks for the

modules, especially for the TM. The module manager is a kind of interface module that interprets scheduling commands coming from the scheduler into a language understood by the modules and transfers the messages to each module.

The VCT also consists of PMs, a TM, the tool scheduler, and the module manager; the components in the VCT are modeled to perform the same functions as the components of the real cluster tool. Fig. 4 demonstrates the reference model of the VCT. The reference model shows what types of messages are exchanged between the components and what functions and information each component has.

4.2. Modeling of Cluster Tool Components
In order to implement the VCT, it is necessary to accurately model the internal operating logic of each component. In this study, we model the internal operating logics of the PMs and the TM as state graphs. Brief explanations of the tool scheduler and the module manager are also included.

4.2.1. State Graph Models
We first briefly explain the state graph model. A state graph is an extended version of timed automata having state variables, system variables, and timers (Choi and Kang 2013). In addition, a state graph allows transition conditions and three types of actions: entry, input, and transition actions. In the graph, "?" denotes an input event or message, and "!" denotes an output event or message. A transition condition is denoted by "~". Among the states of the graph, the initial state and the final state are expressed differently. Fig. 5 shows an example of a simple state graph.

Each state has its own name, entry actions, and a timer $[\Delta(t_0)]$. A state graph model can be described by a state transition table. In a state transition table, all components for constructing the state graph model are specified. The state transition table shown in Table 2 represents the state graph model in Fig. 5.
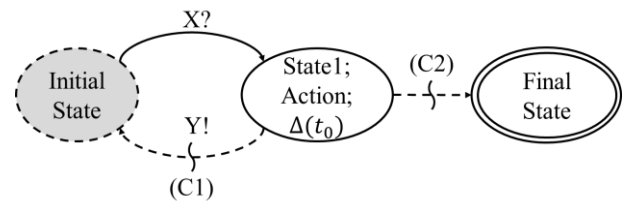
Figure 5: Example of State Graph Model

Table 2: State Transition Table of Model in Fig. 5

State			Input		Transition		Next State
Name	Action	Timer	Event	Action	Condition	Action	
Initial	-	-	X?	-	-	-	State1
State1	-	$\Delta(t_0)$	-	-	C1	Y!	Initial
			-	-	C2	-	Final
Final	-	-	-	-	-	-	-

4.2.2. State Transition Diagram of PM(i,j)

A PM consists of a process part, a chuck, and a slit valve. Therefore, in order to model the operating logic of a PM, the three state graphs for the components are required. In the model, PM(i,j) denotes the j-th parallel PM for the i-th process step.

For the process part, the state graph has six states: 'IDLE', 'PREPARE', 'PROCESS', 'PROCESS_END', 'CLEANING', and 'STOP'. When a wafer is loaded into a PM and a slit valve door is closed, the state of the PM is changed from 'IDLE' to 'PREPARE'. In this study, we assume that a tool follows the earliest starting policy that there are no intentional delays in the tool; hence, the state is changed instantly from 'PREPARE' to 'PROCESS'. If a tool follows another process-start rule, there might be a time delay h_t before entering the process state. The state becomes 'PROCESS_END' after the process time p_t has elapsed. Then the state becomes 'IDLE' after TM unloads the wafer and the slit valve door is closed. A PM cleaning operation can be performed only when the state is 'IDLE'. In order for the cleaning operation to start, the cleaning condition should be satisfied. For a multi-period cleaning operation, the operation starts after the specified number of wafers are processed. For a non-periodic cleaning operation, the operation starts when the degree of contamination inside the PM exceeds the reference value. Such conditions are referred to in the model as (C1). If a cleaning operation starts, the state becomes 'CLEANING'. After a required cleaning time c_t has elapsed, the state returns to 'IDLE'. When a PM receives a '_stop' event message, the state becomes 'STOP' and the simulation is terminated.

The state graph of a chuck has three states: 'UNFIXED', 'FIXED', and 'STOP'. The state graph of a slit valve door also has three states: 'CLOSED', 'OPENED', and 'STOP'. The chuck and the slit valve door work closely together. In order for a wafer to be loaded, the state of the slit valve door should be 'OPENED'. The wafer is then loaded, and the state of the chuck becomes 'FIXED' and the state of the slit valve door becomes 'CLOSED'. Similarly, in order to unload a wafer, the states of the chuck and the slit valve door respectively become 'UNFIXED' and 'OPENED'. After the wafer is unloaded, the state of the slit valve door again becomes 'CLOSED'. When the chuck and the slit valve door receive '_stop' event messages, the states become 'STOP'.

Every operation of the chuck and the slit valve door has its own task time. Therefore, the chuck and the slit valve door send their state transition messages after required task times have passed.

In this study, we propose a model to consider only simplified PM operating logics and PM cleaning operations. However, in actual tools, motion control and wafer alignment check are conducted inside the PM. Furthermore, there are various exceptions such as PM failure and repair, time variations, and wafer alignment failures. An improved model has been developed in order to consider all realistic situations.
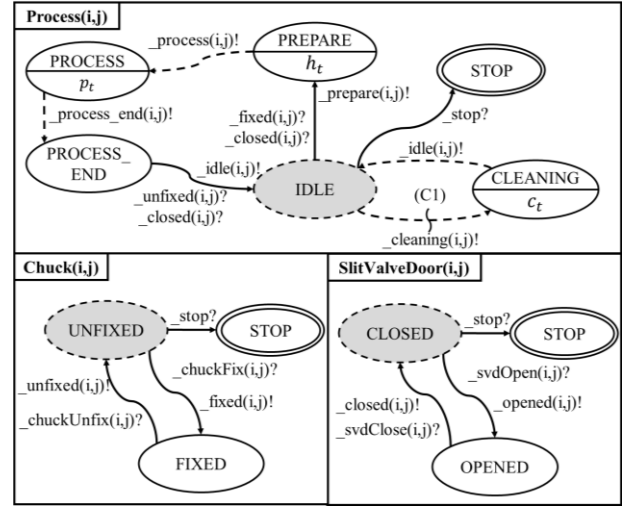
Figure 6: State Transition Diagram of PM(i,j)

Table 3: State Transition Table of PM(i,j)

State		Input	Transition		Next State
Name	Timer	Event	Condition	Action	
IDLE	-	_fixed(i,j)? _closed(i,j)?	-	_prepare(i,j)!	PREPARE
		-	C1	_cleaning(i,j)!	CLEANING
PREPARE	h_t	-	-	_process(i,j)!	PROCESS
PROCESS	p_t	-	-	_process _end(i,j)!	PROCESS _END
PROCESS _END	-	_unfixed(i,j)? _closed(i,j)?	-	_idle(i,j)!	IDLE
CLEANING	c_t	-	-	_idle(i,j)!	IDLE
UNFIXED	-	_chuckFix(i,j)?	-	_fixed(i,j)!	FIXED
FIXED	-	_chuckUnfix(i,j)?	-	_unfixed(i,j)!	UNFIXED
CLOSED	-	_svdOpen(i,j)?	-	_opened(i,j)?	OPENED
OPENED	-	_svdClose(i,j)?	-	_closed(i,j)?	CLOSED

4.2.3. State Transition Diagram of TM

The state transition diagram of a TM is shown in Fig. 7 and Table 4. The state graph model of the TM has three states: 'EMPTY', 'HOLD', and 'STOP'. If the TM holds a wafer, then its state is 'HOLD'. If not, then the state is 'EMPTY'. Each state has a state variable (i,j) that denotes the ID of the PM to which the TM is directed. So that the TM can unload the wafer, the states of the chuck and the slit valve door are 'UNFIXED' and 'OPENED', respectively. After the TM completes its unloading task, it sends an event message '_svdClose' so that the slit valve door can be closed. Likewise, the state of the slit valve door should be 'OPENED' before the TM loads a wafer. After the wafer loading is completed, the TM sends event messages '_chuckFix' and '_svdClose' to the chuck and slit valve door, respectively. When the TM moves to another PM, it updates its state variable (i,j). Finally, the state becomes 'STOP', which is the final state, if the TM receives a '_stop' event message.

Each robot operation has its own task time. Thus, like the chuck and the slit valve door, the TM sends its job completion messages after the simulation time has elapsed the required task time.

In addition to the single-armed robot considered in this study, there exist diverse TMs such as a dual-armed robot, a quad-armed robot, and a robot with independent arms.

Furthermore, TM malfunctions occasionally occur. Future work will address a more comprehensive TM model.
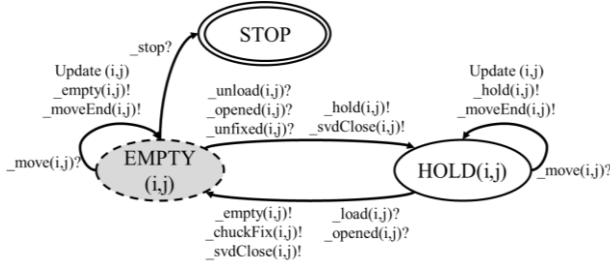
Figure 7: State Transition Diagram of TM

Table 4: State Transition Table of TM

State		Input			Next State
Name	Timer	Event	Condition	Action	
EMPTY(i,j)	-	_unload(i,j)? _opened(i,j)? _unfixed(i,j)?	-	_hold(i,j)! _svdClose(i,j)!	HOLD(i,j)
		_move(i,j)?	-	Update (i,j) _empty(i,j)! _moveEnd(i,j)!	EMPTY(i,j)
HOLD(i,j)	-	_load(i,j)? _opened(i,j)?	-	_empty(i,j)! _chuckFix(i,j)! _svdClose(i,j)!	EMPTY(i,j)
		_move(i,j)?	-	Update (i,j) _hold(i,j)! _moveEnd(i,j)!	HOLD(i,j)

4.2.4. Tool Scheduler and Module Manager

The tool scheduler determines when each module, especially the TM, should perform what tasks. We apply the open architecture scheduler to the VCT. The tool scheduler issues scheduling commands by referring to the SCFs.

The module manager is responsible for interpreting the scheduling commands into language that PMs and TM can understand and delivering the commands to the modules. In this study, state transitions of the modules occur via event messages; the module manager interprets scheduling commands into several event message sequences and transfers the commands to the modules. For example, when the module manager of the VCT receives a scheduling command 'unload the wafer #1 from PM(1,1)', the module manager sends an event message '_unload(1,1)' to TM and event messages '_svdOpen(1,1)' and '_chuckUnfix(1,1)' to PM(1,1), particularly to SlitValveDoor(1,1) and Chuck(1,1), respectively. If the tool follows other job timing rules, rather than the earliest starting policy, the event messages should include the timing information for each event message.

The module manager is highly related to the tool configuration and the internal operating logic of the modules. The presence of the module manager allows the tool scheduler to operate independently of the tool architecture. Even when two cluster tools have different architectures, it is possible to operate them through a single scheduler when the two module managers are able to interpret the same scheduling commands.

In addition, the module manager informs the scheduler of the states of the modules. The module manager analyzes the state messages sent from the modules and informs the scheduler as to whether each module is abnormal or not.

5. APPLICATION: PERFORMANCE ANALYSIS IN A SINGLE-ARMED CLUSTER TOOL WITH PM CLEANING OPERATIONS

As an application of the proposed framework, we conduct a performance analysis on benchmark schedules in a single-armed cluster tool with PM cleaning operations. The schedules to be analyzed are the backward sequence which guarantees optimal productivity in the basic single-armed cluster tool and the backward(z) sequence which is known to give good performance in a single-armed cluster tool with 1-periodic PM cleaning operations, as proposed by Yu and Lee (2017).

5.1. SCFs for Performance Analysis

For the performance analysis, we first create tool architecture information SCFs. In the analysis, we set all task times of the slit valve door and the chuck 1. TM task times are also set to 3. The number of process steps, the number of parallel PMs, the process times, the cleaning times, and the cleaning periods are set differently according to the experiment.

After this, the benchmark TM task schedule SCFs for the backward and the backward(z) sequences should be created. Both sequences are basically similar in that they are pull-type tool operation methods; however, the initial state of the tool when the tool is operated in the cyclic schedule is different. The conventional backward sequence proceeds at the full loading state while all PMs are in progress and then returns to the full loading state. On the other hand, in order to keep the WIP constantly and increasing the productivity, the backward(z) sequence is used to allocate some of the parallel PMs to PM cleaning operations, where the number of PMs to allocate is denoted by 'z'. This is called a partial loading strategy. Therefore, although both schedules give the same schedule messages, those show a large difference in productivity due to the different operation strategies. For example, consider a tool with six PMs and three process steps. Assume that PM1, PM2 - PM4, and PM5 - PM6 are assigned as parallel PMs for the first, the second, and the third process steps. Table 5 demonstrates the initial states of both sequences when a tool is operated on a cyclic schedule. One way of finding an optimal z value is introduced in Yu and Lee (2017). Fig. 8 shows Petri net models for both schedules. The model has a different net shape and token marking vectors depending on the tool configuration and the cleaning period. The earliest starting rule is applied in the analysis.

In this experiment, we do not consider any exceptions in our analysis of the impact of the cleaning period and time on the performance of each TM schedule.

Table 6: Initial States for Backward and Backward(z) Sequences

Initial State for the Backward Sequence					
PM1(1)	PM2(2)	PM3(2)	PM4(2)	PM5(3)	PM6(3)
Process	Process	Process	Process	Process	Process

Initial State for the Backward(z) Sequence (optimal z = [0, 1, 1])					
PM1(1)	PM2(2)	PM3(2)	PM4(2)	PM5(3)	PM6(3)
Process	Cleaning	Process	Process	Cleaning	Process
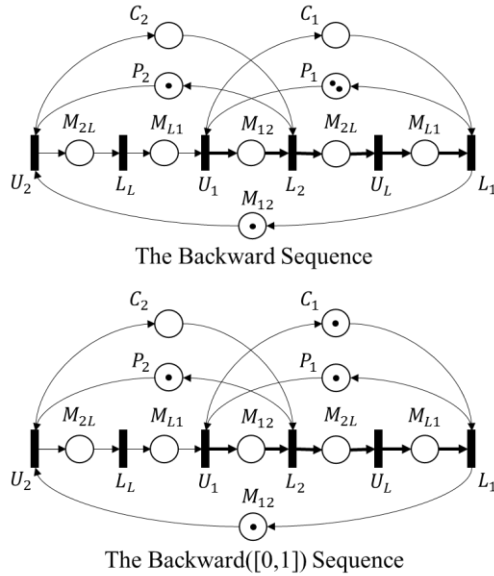
Figure 8: Petri Net Models for the Backward and the Backward(z) Sequences

5.2. Experiment Results

In the experiments, we computed the average cycle times of the backward and backward(z) sequences for the SCFs that we created for diverse tool configuration and both TM schedules.

Problem instances for the experiment and the results are shown in Table 7. In the table, the wafer flow pattern $[m_1, \ldots, m_n]$ means that there exist m_i parallel PMs for the i-th process step. This is similar to the process times $[p_1, \ldots, p_n]$ and the cleaning times $[c_1, \ldots, c_n]$. The cleaning periods are assumed to have the same value for all PMs; this value is larger than one. Since the average cycle time indicates that the average time elapsed to produce a wafer, a lower average cycle time means better performance. The gap shows how the average cycle time of the backward(z) sequence is lower than the average cycle time of the backward sequence. Thus, the negative gap means that the backward sequence yields better performance.

From Table 7, we can see that, for the most cases, the backward(z) sequence provides better performance than the backward sequence. Particularly, the gap becomes larger when the PMs have a short cleaning period and long cleaning times. This indicates that the greater the influence of the PM cleaning processes on the tool operation, the better the productivity of the backward(z) sequence. However, the backward sequence provides better performance for several cases with long cleaning period and short cleaning times. Therefore, in order to increase the productivity, it is important to apply an appropriate schedule according to the tool configuration, the cleaning time, and the cleaning period. Since adaptive scheduling is possible in response to environment changes and time-disruptive exceptions, the suggested open architecture scheduling system for the cluster tool will be competitive.

Table 7: Performance Analysis of the Backward and Backward(z) Sequences.

Tool Configuration				Average Cycle Time		
Wafer Flow Pattern	Process Times	Cleaning Times	Cleaning Periods	Backward	Backward(z)	Gap (%)
[4]	[100]	[50]	2	190	150.5	20.8
			3	169.7	141.5	16.6
	[100]	[100]	2	199.5	168.8	15.4
			3	178.6	161.33	9.7
[1, 3]	[70, 300]	[35, 150]	2	152.8	164.5	-7.7
			3	129.1	164.5	-27.4
		[70, 300]	2	221.2	181.5	17.9
			3	188.4	171.8	8.8
[1, 4]	[70, 400]	[35, 200]	2	168	148.4	11.7
			3	150.7	148.4	1.5
		[70, 400]	2	275.6	154.9	43.8
			3	215.5	147.3	31.6
[2, 3]	[150, 300]	[75, 150]	2	298	268.5	9.9
			3	278	238.7	14.1
		[150, 300]	2	459.1	279.5	39.1
			3	385	250.8	34.9
[1, 3, 2]	[50, 300, 150]	[25, 150, 75]	2	158.3	164.5	-3.9
			3	144.7	164.5	-13.7
		[50, 300, 150]	2	241.4	191.2	20.8
			3	203	182	10.3

6. CONCLUSION

We have suggested a framework of an open architecture scheduling system for a single-armed cluster tool with PM cleaning operations. In the system, the tool scheduler refers to SCFs and performs scheduling. SCFs can be freely replaced from outside. To apply the framework, we have developed a VCT similar to the actual cluster tool. The scheduler, the module manager, the PMs, and the TM are modeled to describe the VCT. Using the VCT, we have also conducted performance analysis on the backward and backward(z) sequences for a single-armed cluster tool with multi-period PM cleaning operations. As a result, we found that the open architecture scheduling system that can modify or change the tool schedules based on various tool environments is effective at increasing productivity.

In future works, we will specify SCFs to enable scheduling of various tool structures and operational constraints. Furthermore, exception-handling techniques for various exceptions will be examined.

REFERENCES

Choi B.-K. and Kang D., 2013. Modeling and Simulation of Discrete Event Systems. John Wiley & Sons.

Geismar H. N., Dawande M., and Sriskandarajah C., 2004. Robotic cells with parallel machines: Throughput maximization in constant travel-time cells. Journal of Scheduling, 7 (5), 375–395.

Joo Y.-J. and Lee T.-E., 2004. A Virtual Cluster Tool for Testing and Verifying a Cluster Tool Controller and a Scheduler. IEEE Robotics & Automation Magazine, 11 (3), 33-49.

Kim C. and Lee T.-E., 2013. Modelling and simulation of automated manufacturing systems for evaluation of complex schedules. International Journal of Production Research, 51 (12), 3734-3747.

Kim H., Kim H.-J., Lee J.-H., and Lee T.-E., 2013. Scheduling dual-armed cluster tools with cleaning processes. International Journal of Production Research, 51 (12), 3671–3687.

Kim H.-J., Lee J.-H., and Lee T.-E., 2015. Noncyclic scheduling of cluster tools with a branch and bound algorithm. IEEE Transactions on Automation Science and Engineering, 12 (2), 690–700.

Kim H.-J., Lee J.-H., Jung C., and Lee T.-E., 2013. Scheduling cluster tools with ready time constraints for consecutive small lots. IEEE Transactions on Automation Science and Engineering, 10 (1), 145–159.

Kim J.-H., Lee T.-E., Lee H.-Y., and Park D.-B., 2003. Scheduling analysis of time-constrained dual-armed cluster tools. IEEE Transactions on Semiconductor Manufacturing, 16 (3), 521–534.

Lee J.-H., Kim H.-J., and Lee T.-E., 2014. Scheduling cluster tools for concurrent processing of two wafer types. IEEE Transactions on Automation Science and Engineering, 11 (2), 525–536.

Lee, J.-H., and Lee, T.-E., 2010. An open scheduling architecture for cluster tools. In Proceedings of the 6th IEEE Conference on Automation Science and Engineering, pp. 420-425, Toronto (Ontario, Canada).

Lee, T.-E., 2008. A review of scheduling theory and methods for semiconductor manufacturing cluster tools. In Proceedings of the 40th Conference on Winter Simulation, pp. 2127-2135, Miami (Florida, USA).

Min Y. and Lin X.-R., 2013. A Simulation-Based Analysis of Cluster Tools Scheduling with Plant Simulation. In Proceedings of the 19th International Conference on Industrial Engineering and Engineering Management, pp. 71-80. Berlin (Germany),

Niedermayer H. and Rose O., 2003. A simulation-based analysis of the cycle time of cluster tools in semiconductor manufacturing. In Proceedings of the 15th European Simulation Symposium, pp. 349-354, Delft (Netherlands).

Pan C., and Bao N., 2012. An eM-Plant-based virtual single-arm cluster tool. In Proceedings of the 9th IEEE International Conference on In Networking, Sensing and Control, pp. 40-45, Beijing (China).

Rostami S., Hamidzadeh B., and Camporese D., 2001. An optimal periodic scheduler for dual-arm robots in cluster tools with residency constraints. IEEE Transactions on Robotics and Automation, 17 (5), 609–618.

Shin Y.-H., Lee T.-E., Kim J.-H., and Lee H.-Y., 2001. Modeling and implementing a real-time scheduler for dual-armed cluster tools. Computers in Industry, 45, 13-27.

Yu T.-S., Kim H.-J., and Lee T.-E., 2017, Scheduling single-armed cluster tools with chamber cleaning operations. IEEE Transactions on Automation Science and Engineering, DOI: 10.1109/TASE.2017.2682271.

Venkatesh S., Davenport R., Foxhoven P., and Nulman J., 1997. A steady-state throughput analysis of cluster tools: Dual-blade versus single-blade robots. IEEE Transactions on Semiconductor Manufacturing, 10 (4), 418–424.

Wu N. Q., Chu F., Chu C., and Zhou M. C., 2011. Petri net-based scheduling of single-arm cluster tools with reentrant atomic layer deposition processes. IEEE Transactions on Automation Science and Engineering, 8 (1), 42–55.

Wu N. Q., Chu C., Chu F., and Zhou M. C., 2008. A petri net method for schedulability and scheduling problems in single-arm cluster tools with wafer residency time constraints. IEEE Transactions on Semiconductor Manufacturing, 21 (2), 224–237.

ACO TOPOLOGY OPTIMIZATION: THE GEOMETRICAL CONSTRAINT BY LEARNING OVERLAID OPTIMAL ANTS ROUTE

Nanami Hoshi[(a)], Hiroshi Hasegawa[(b)]

[(a)(b)] Graduate School of Engineering and Science, Shibaura Institute of Technology, Japan

[(a)]mf17061@shibaura-it.ac.jp [(b)]h-hase@shibaura-it.ac.jp

ABSTRACT

Topology optimization commonly has performed minimization of the mean compliance under a volume constraint. On the other hand, mechanical product designers are considering "a weight minimization under a stress constraint" as an objective and constraints for generating new optimal structure. Moreover, for obtaining this objective, a mechanical structure design has performed to minimize weight of its structure by checking the principal stress vectors as the force's flow, and speculating its desirable structure under maintaining its stiffness, iteratively. These design processes' difference has generated mismatch between actual design practice and the conventional topology optimization theory. Therefore, we have proposed ACO using the principal stress vector for overcoming mismatch of the topology optimization theory. In this paper, ACO Topology Optimization with Geometrical Constraint (ACTO with GC) is proposed to improve unnecessary structures elements problem. Our proposal is new geometrical constraint method which overlays obtained optimal ants route as the shape feature pattern, learns it for next optimization process.

Keywords: Ant Colony Optimization, Topology Optimization, Mechanical Structure Design, Principal Stress Vector

1. INTRODUCTION

 Topology optimization has been used for structural optimization, and the various techniques and approaches of topology optimization have been developed and researched since about 1985 (Nishiwaki, Izui and Kikuchi 2012). Topology optimization can change shape, size and number of holes, therefore topology optimization is the most flexible methodology in structural optimization. Topology optimization has generally performed minimization of the mean compliance under a volume constraint until now. First topology optimization's CAE software, which is named OPTISHAPE, has been sold in Japan from 1989. OPTISHAPE, based on minimization of the mean compliance, has been studied and developed (Bendsoe and Kikuchi 1988, Suzuki and Kikuchi 1991). OPTISHAPE has been utilized in various industries, such as machine, aircraft, building and automobile industries have been used. However, obtained optimal topological structure has complex shapes and layouts. It was difficult to manufacture it efficiently, because it is required that precision technology and great cost should be supplied. Additive Manufacturing (AM), which is a rapidly evolving field, solves the problem between topology optimization and an engineering and manufacturing standpoint. AM via including the 3D printer has changed this situation, because production by AM has flexible and to be able to produce optimal structure introduced by topology optimization. A method of topology optimization focused on AM was suggested in 2011 (Brackett, Ashcroft and Hague 2011).

 On the other hand, mechanical product designers often consider "a weight minimization under a stress constraint" and the force flow *i.e.* principal stress vector when design optimal structure. Moreover, for obtaining this objective, a mechanical structure design has performed to minimize weight of its structure by checking the principal stress vectors as the force's flow, and speculating its desirable structure under maintaining its stiffness, iteratively. These design processes' difference has generated mismatch between actual design practice and the conventional topology optimization theory. However, topology optimization has generally performed minimization of the mean compliance under a volume constraint. Moreover, topology optimization considering principal stress vector is not much.

 Structure optimization using ACO has been suggested by Champ et al. in 2004 (Camp and Bichon 2004), and topology optimization using ACO was applied in 2008 (Kaveh, Hassani, Shojaee and Tavakkoli 2008). However, these applied optimization only have been introduced a basic principle of ACO theory to for minimization of the mean compliance via a density method (Takada 2006). Subsequently, new topology optimization using ACO, which is called ACO Topology Optimization (ACTO), has been proposed and developed (Guan and Chun 2009, Ito Hoshi and Hasegawa 2016). These methods have versatility of design variable and, optimize in elements of the discretized design domain

that is ants are moving in design domain. Especially, the method developed by Ito considered a principal stress vector as a design variable for filling this mismatch between the topology optimization theory and mechanical product designers. However, these optimization has an improvement problem, and it is that an optimal topological shape and layout in which many unnecessary structures elements were included is obtained. In this paper, ACTO with a geometrical constraint (ACTO with GC) is proposed to solve the unnecessary structures elements problem. Our proposal is new geometrical constraint method which overlays obtained optimal ants route as the shape feature pattern by ACO, and learns its pattern for next optimization process. This paper discusses the effect of new proposal method through a trial of simple cantilever problem.

2. ANT COLONY OPTIMIZATION

ACO algorithm was inspired by a behavior of ant swarm intelligence. ACO consists of three steps (Figure 1). First step is a setting the initial pheromone of ants on the route. Next step is to add pheromone on the route where ants selected. Final step is to update the pheromone on overall routes. By repeating these steps, ACO obtains the optimal route by moving of ants which are guided to its pheromone. This pheromone update procedure consists of addition and evaporation of pheromone. The addition of pheromone means that pheromone is added on the route where ants passed, and then, the evaporation of pheromone means that a pheromone evaporates with rain, respectively. Moreover, ants are able to communicate with a number of ants and can obtain optimal route through pheromone update.

To compute pheromone updating:

$$\tau(t+1) = \mu\tau(t) + \sum_{i=1}^{m} \Delta\tau_i \qquad (1)$$

μ: Reduction factor   m: individual Number
i:selected route $\Delta\tau_i$:addition pheromone
t: Generation Number

Equation(1) shows update rule of the pheromone. Where, Reduction factor μ become decimal value, and then the pheromone is reduced via this factor. It means evaporation of pheromone.
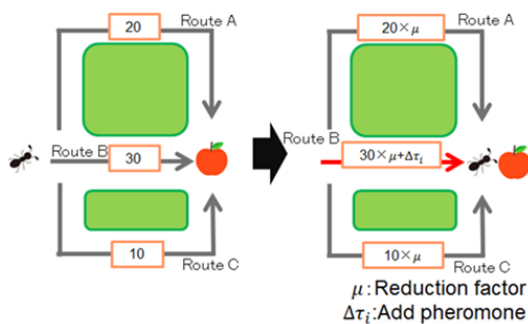
Figure 1: Process of ACO

3. ACO TOPOLOGY OPTIMIZATION

First, several ants generate routes in a design region. The design region is set from Von Mises stress as first pheromone of first generation. The generation means a repeat of ACTO. Routes of generated by ants become a structure one of structures of first generation. This study regards the structure as an individual. Optimal individual in first generation is chose by elite strategy when ACTO satisfy the requirement of individual. ACTO update first phenome of next generation base on the optimal individual of first generation. Moreover, the design space of a mechanical structure is divided by finite element where each element shows material or void as Figure 2 (exist: 1, not exist: 0). In addition, ants explore optimal route in the design space by considering finite elements as a route. A topology representation is created by an ant's routes via setting 1 as passed route and 0 as non-passed of an existence for a finite element. These elements become design variables of ACTO. The structure can be described by a discrete function ρ, as below:

$$\rho = \{\rho_1, \rho_2 \dots \rho_i\} \qquad (2)$$

$$E_\rho = \begin{cases} E_1 : \text{if } \rho_i = 1 \\ E_2 : \text{if } \rho_i = 0 \end{cases} \qquad (3)$$

$$\rho \in 1, 0 \qquad (4)$$

ρ: density function   i:element number
E_1: Young's modulus(material exist)
E_2: Young's modulus(material don't exist)

In this case, E_1 is adopt as young's modulus when the element has material but E_2 is adopt as young's modulus when the element is void. Therefore, the element chose as route by ants has material, hence optimal structure of ACTO only has 1 or 0 of density (Figure 3).
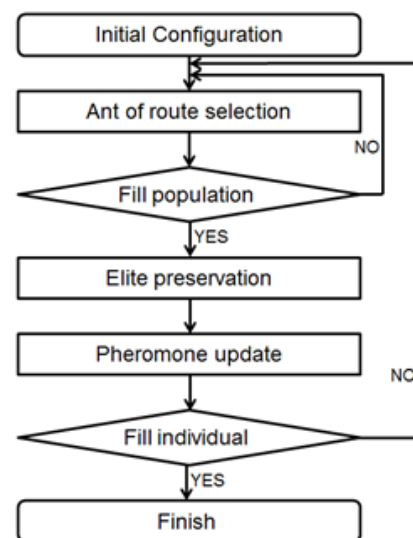
Figure 2: Flow chart of ACO topology optimization
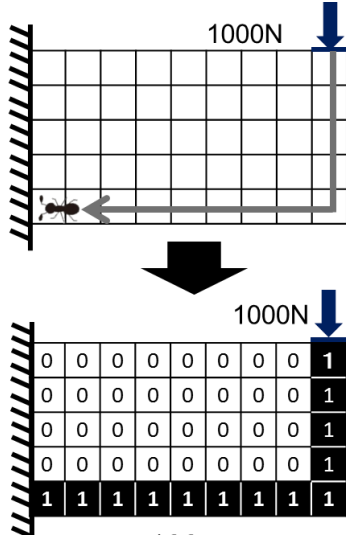
Figure 3: Object model diagram and route of ants

1) Initial Configuration

The first pheromone is set for each elements. Von Mises stress is used to set the first pheromone.

2) Ant of route selection

The set first pheromone is use by ants to generate routes. This section use some method, as follow:

- Linear rank method

The linear rank method has been proposed by Baker in 1985(Baker 1985, Mitchell and Iba 1997). In the linear rank method, its pheromone are ranked normal ascending order from 1st to N^{th}, redistribute values based on rank order (Figure 4). Moreover, the Max value of the linear rank method use Table 1.

Equation of linear rank method, as follow:

$$\text{Exp Val(r, t)} = \text{Min} + (\text{Max} + \text{Min})\frac{\text{rank(r, t)-1}}{\text{N-1}} \quad (5)$$

$$1 \leq \text{Max} \leq 2 \quad (6)$$

$$\text{Min} = 2 - \text{Max} \quad (7)$$

t: time(iteration number) Max: N^{th} redistribution value
Min: 1st redistribution value r: rank
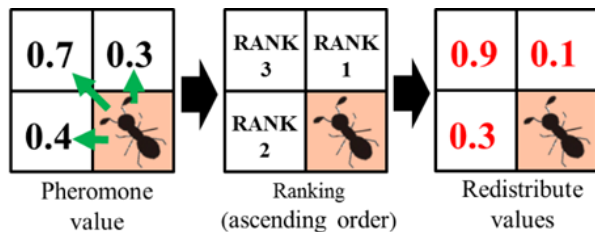
Figure 4: Linear rank method

Table 1: Max value of the linear rank method

Generation number	Max value of liner rank method
1-100	$1 \leq Max \leq 1.5$
101-200	$1.5 < Max \leq 2$

- Roulette wheel selection

Roulette wheel section (Lipowski and Lipowska 2012) is the probability of selection is proportional to redistribute values of the liner lank method. The better fitted redistribute values of the liner lank method, the larger the probability of selection (Figure 5). This method considers N individuals, each characterized by redistribute values of the liner lank method. Selection of an individual choose randomly. The selection probability of i-th individual P_i is follow as:

$$P_i = \frac{\text{Exp Val(r, t)}}{\sum_{i=1}^{N} \text{Exp Val(r, t)}} \quad (i = 1,2,3 \dots, N) \quad (8)$$

Exp Val(i,t): redistribution value of liner lank method
i: individual r: rank of liner lank method
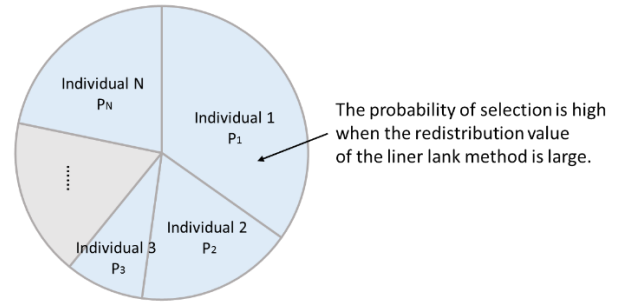
Figure 5: Roulette wheel selection

-Consider with the principal stress vector

The green arrow in Figure 6 is the synthetic vector of the maximum principal stress and the minimum seed stress. The probability of selected the element increase when the element has the principal stress vector (Ito, Hoshi and Hasegawa 2016).
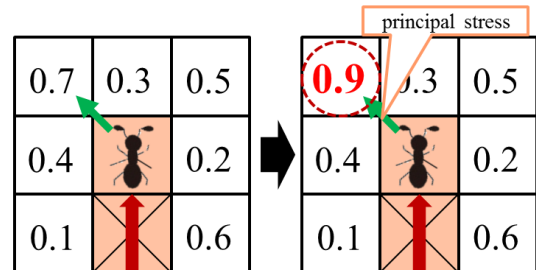
Figure 6: Consider with the principal stress vector

3) Fills population

Ants repeat route selection in design domain until the set number of individuals is satisfied.

4) **Elite preservation strategy**

Elite preservation strategy is able to replicate optimal individuals (*i.e.* optimal structure) in next generation, therefore this method prevents deterioration of volume of optimal structure (Kenneth 1975). Figure 7 shows the flow of Elite preservation strategy.
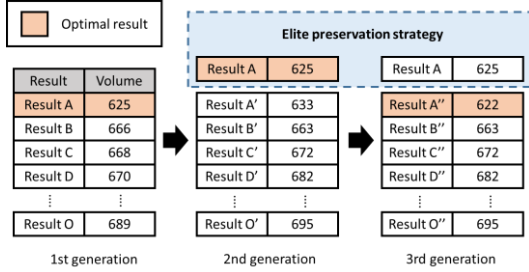
Figure 7: Elite preservation strategy

5) **Pheromone update**

This section executes pheromone update where paths' ants have walked. Also, existing pheromones decrease because of evaporation of pheromone. Reduction rate of evaporation of pheromone M is often use 0.8 to 0.98.

6) **Fill individual**

ACTO repeats generating optimal structures until the set number of generations is satisfied.

4. ACO TOPOLOGY OPTIMIZATION WITH GEOMETRICAL CONSTRAINT

An optimal structure, is introduced by ACTO, has not been able to generate the intermediate element of density. This is a strong point of ACTO, but many unnecessary structures elements have comprised a large percentage of an optimal structure (Figure 8).

(a) Optimal structure of ACTO

(b) Optimal structure of other method
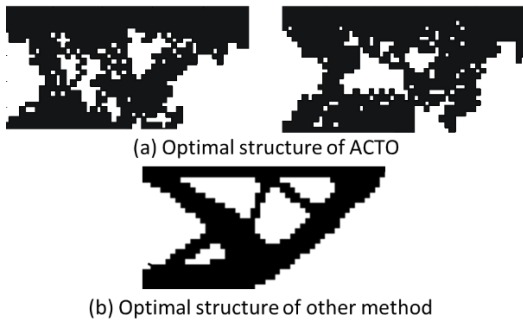
Figure 8: Analysis result of ACTO

Therefore, we propose new geometrical constraint method (*i.e.* learning function) by learning overlaid ants route into ACTO, which is named ACTO with GC, to improve unnecessary structures problem. Figure 9 shows the process of ACTO with GC. First step is to perform ACTO of the inner loop using Von Mises stress as the first pheromone value. This is repeated until iteration count in Table 2. Next, the shape feature pattern is made in learning function. Figure 10 shows a way of making the shape feature pattern in learning function. Red part of this figure is necessary structures for optimal structure of

ACTO. On the other hand, blue part of this figure is unnecessary structures for optimal structure of ACTO. In this function, optimal ants route is overlaid. After that, elements with small number values change the value to 0 because this elements is unnecessary structures for optimal structure of ACTO. This overlaid route becomes the first shape feature pattern. We repeat ACTO based on the pheromone that was calculated from the first shape feature pattern and its Von Mises stress into outer loop. This outer loop terminates the iteration count of outer loop in Table 2.
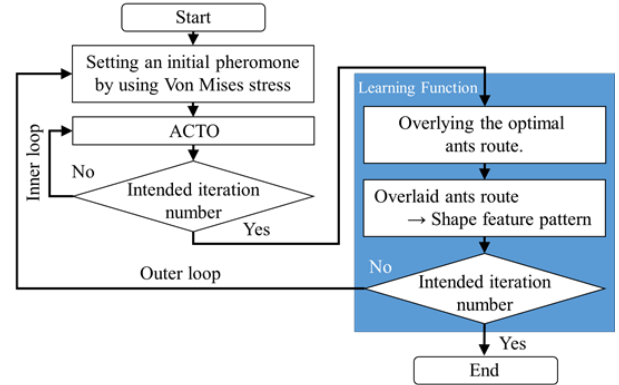
Figure 9: Flowchart of ACTO with GC

Table 2: Iteration count of inner and outer loops

Outer loop	3 times
Inner loop	10 times
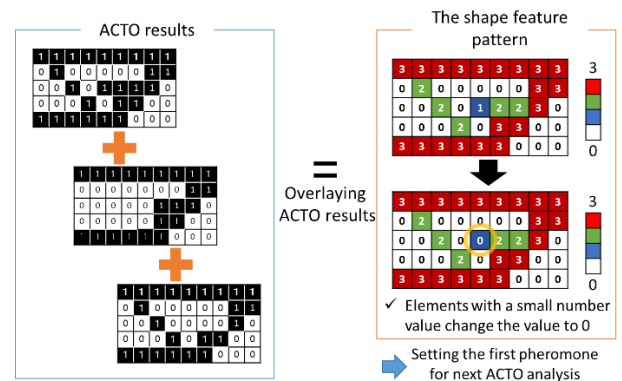
Figure 10: making the shape feature pattern in learning function

5. EVALUATION

Initial pheromone distribution of ants is created by Von Mises stress distribution. The optimization definition is described as the following.

$$f_P(\rho) = V(\rho) + rP(\rho) \rightarrow \text{Min} \qquad (5)$$

$$V(\rho) = \int_s \rho \, ds \qquad (6)$$

$$P(\rho) = \begin{cases} 0, & \text{if} \quad \sigma^{max} < \sigma^{all} \\ 1, & \text{otherwise} \end{cases} \qquad (7)$$

Where $f_P(\rho)$, $V(\rho)$, and $P(\rho)$ denote the modified objective function, the volume function for mechanical structure, and the penalty function, respectively. The design variables are defined by density variable ρ. The local stress constraint is consisted of maximum stress σ^{max} and allowable stress σ^{all}.

6. ANALYSIS SETTING

Figure 11 shows object model in using this study. In addition, the analysis settings of this paper are shown in Table 3. In this study, we analyze 2 approaches to consider the shape feature pattern in ACTO with GC. First, analysis type 1(no normalization) does not change the shape feature pattern in setting the first pheromone. On the other hand, analysis type 2 (normalization) generate the first pheromone by binarizing the shape feature pattern (Figure 12).
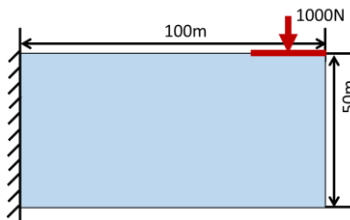
Figure 11: object model

Table 3: Experiment condition

Mesh division Vertical	25
Mesh division Horizontal	50
The number of generations individuals	10
Ant to route search per individual	10
Generation number	200
Add pheromone value	0.0002
Pheromone reduction factor	0.99
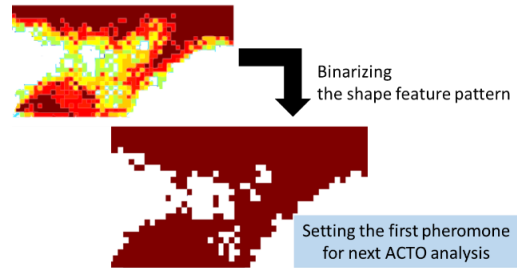
Figure12: Binarizing the shape feature pattern

7. RESULTS AND DISCUSSION

We applied ACTO with GC to the simple cantilever problem. Figure 13 shows the shape feature pattern of cantilever structure in each outer loop of learning function and the optimal structure. Red elements of this figure have strong characteristics of optimal structure, especially these elements are necessary structures for optimal structure of ACTO. On the other hand, yellow and green elements of this figure has weak characteristic of optimal structure, seemingly this part is unnecessary structures for optimal structure of ACTO. In addition, the optimal structure extracted characteristics from the shape feature pattern and made by using modeling tool.

The shape feature pattern of analysis type 1 have red elements all over this shape. Especially, the third shape feature pattern (Figure13(c)) has its tendency. Hence, ACTO with GC is possible to delete unnecessary structures and generate clearly optimal structure from the shape feature pattern, such as Figure13 (d). However, the optimal structure (Figure 13(d)) have scraggly paths, because the shape feature pattern of analysis type 1 has a tendency that the structure of the upper part of the shape feature pattern is derived thickly. Analysis type 1 need to lessen its tendency.
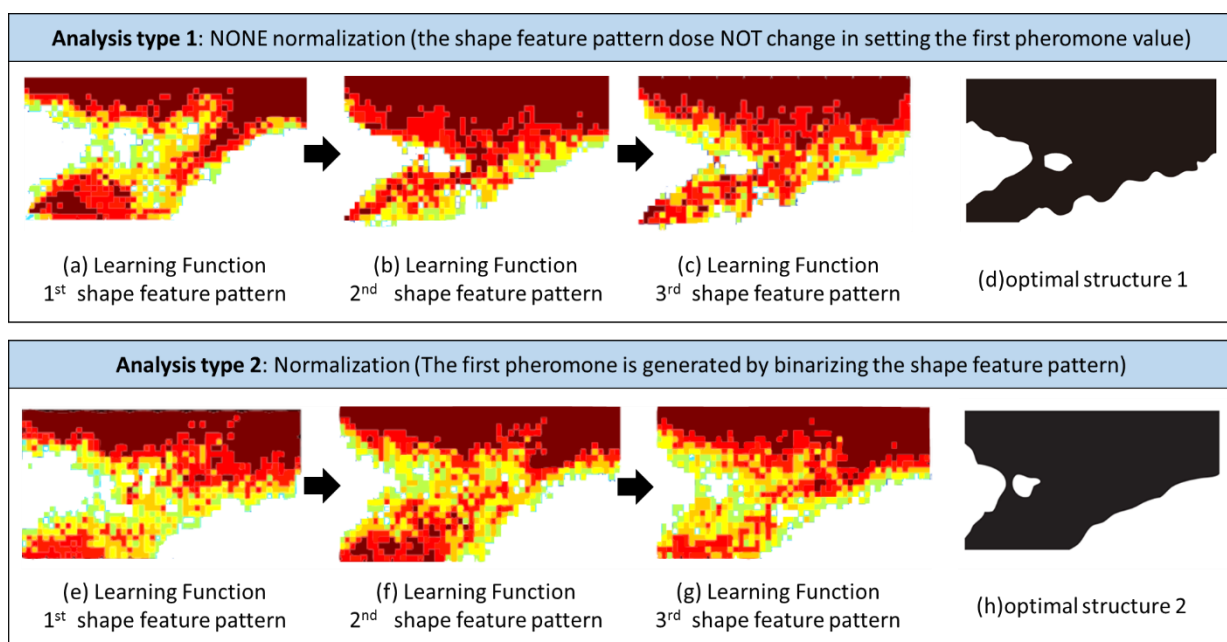
Figure 13: Analysis result of ACTO with GC

On the other hand, the shape feature pattern of analysis type 2 has not the tendency that the structure of the upper part of the shape feature pattern is derived thickly. In addition, the shape feature pattern of analysis type 2 has a tendency to generate the internal structure finely. However, the internal shape of the shape feature pattern of analysis type 2 has Red elements less than the shape feature pattern of analysis type 1, because of a global route selection by ants. Therefore, the optimal structure (Figure 13(h)) has not tendency of the internal structure finely. Analysis type 2 need to increase iteration count of inner and outer loops of learning function.

From these facts, the optimal structure is change by a way of setting initial pheromone value of the shape feature pattern in ACTO with GC. Moreover, count of inner and outer loops of learning function is important to improve the optimal structure.

8. CONCLUSION

In this paper, to solve the unnecessary structures problem of ACTO, the geometrical constraint method by learning overlaid optimal ants route have been introduced in ACTO. As the result, we confirmed ACO with GC is able to obtain the optimal structure, and to reduce unnecessary structures elements. However, unnecessary structures elements remain in the shape feature pattern. To remove unnecessary structures elements, it is necessary to increase iteration count of inner and outer loops. In addition, a way of setting initial pheromone value of the shape feature pattern in ACTO with GC is important to improve the optimal structure. In the future work, we plan to reconsider about iteration count of inner and outer loops and way of setting initial pheromone value of the shape feature pattern in learning function. After that, we try to perform ACTO with GC using new iteration count and a way of setting initial pheromone value.

REFERENCES

Nishiwaki, S., Izui, K., and Kikuchi, N., 2012. Topology Optimization, The Japan Society for Computational Engineering and Science, MARUZEN-YUSHODO Company, (in Japanese).

Bendsoe, M.P. and Kikuchi, N., 1988. Generating optimal topologies in structural design using a homogenization method, Comput, Meth. Appl. Mech. Eng, Vol.71 No.2: pp197-224.

Suzuki, K. and Kikuchi, N., 1991. A homogenization method for shape and topology optimization, Comput. Meth. Appl. Mech. Eng, Vol. 93: pp.291-318.

Brackett D., Ashcroft I., Hague R., 2011. Topology Optimization for additive manufacturing, Proceedings of the solid freeform fabrication symposium, Austin, Vol. 1:pp. 348-362.

Camp, C.V. and Bichon, B.J., 2004. Design of Space Trusses Using Ant colony Optimization, Journal of Structural Engineering, Vol. 130, Issue 5.

Kaveh, A., Hassani, B., Shojaee, S., and Tavakkoli., 2008. S.M., Structural topology optimization using ant colony methodology, Engineering Structures Vol. 30, Issue 9.

Takada, T., 2006. An optimization method for truss structures with objective function of volume and compliance, Transactions of Japan Society for Computational Engineering and Science (JSCES), No. 20060004(in Japanese).

Ito Y., Hoshi N., Hasegawa H., 2016. Ant Colony Topology Optimization for Mechanical Structure with the Local Stress Constraint, Asian Congress of Structure and Multidisciplinary Optimization 2016, May 22-26, Nagasaki, Japan.

Guan-Chun Luh., Chun-Yi Lin., 2009. Structural topology optimization using ant colony optimization algorithm, Applied Soft Computing, Vol.9:pp.1343-1353.

Baker, J.E 1985. Adaptive selection methods for genetic algorithms, In Grefenstette, J.J. ed., Proceedings of First International Conference on Genetic Algorithms and Their Applications, Erlbaum.

Mitchell, M., and Iba, H. 1997. Introduction to Genetic Algorithms, Tokyo Denki Univaesity Press, (in Japanese).

Lipowski A. and Lipowska D. 2012. Roulette-wheel selection via stochastic acceptance. Physica A: Statistical Mechanics and its Applications, 391(6), 2193-2196.

Kenneth Alan De Jong, 1975. An Analysis of the behavior of a class of genetic adaptive systems, Ph.D. thesis, University of Michigan, Ann Arbor.

AUTHORS BIOGRAPHY

Nanami Hoshi received her B. E.(2017) from Shibaura Institute of Technology(SIT), Japan. She is a master student at Division of Systems Engineering and Science-Graduate School of Engineering and Science, SIT. Her research interests are optimization methods.

Hiroshi Hasegawa received his B.E. (1992) and M.E.(1994) from Shibaura Institute of Technology(SIT), Japan. He received PhD (1998) in Mechanical Engineering from Tokyo Institute of Technology, Japan. He has been working at SIT, and currently is a Professor. He is member of JSME, ASME, JSST, JSCES and JSDE. His research interests include computer-aided exploration, especially multi-peak optimization, robust design and multi-disciplinary optimization.

SIMULATION STUDY FOR IMPROVING THE PERFORMANCE OF A PRODUCTION LINE IN THE ELECTRONICS INDUSTRY

Sara Costa[(a)], Ana Luísa Ramos[(b)], José Vasconcelos Ferreira[(c)]

[(a)] DEGEIT, University of Aveiro, Campo de Santiago, 3810-193 Aveiro, Portugal
[(b) (c)] GOVCOPP/DEGEIT, University of Aveiro, Campo de Santiago, 3810-193 Aveiro, Portugal

[(a)]costasara@ua.pt [(b)]aramos@ua.pt [(c)]josev@ua.pt

ABSTRACT

Due to the complexity of the current production systems in the industrial sector, more complex simulation models are needed. This work describes the study and the analysis of the production process of one of the main products produced by the electronics Company X. For this study, a simulation model was developed to mimic the current operation of the production line. The simulation study key objective was to evaluate the dynamic behaviour of the production process of the Product Y based on some performance measures such as cycle time, lead time, utilization rate of resources and work in process' statistics. They are also presented three scenarios that show improvement suggestions in relation to the system current configuration. The use of computer simulation revealed the importance of this tool in process control and in the analysis of improvement strategies that make the production system more efficient. Therefore, simulation can be used as a scientific basis to help in decision making with the considerable gain of avoiding the interference with the regular operation of the system.

Keywords: cycle time, lead time, simulation, Arena software

1. INTRODUCTION

In order to understand and cope with the challenges that businesses face today, empowered and able people are needed to make decisions in uncertain environments, because the ability of making system analysis and to find constraints or opportunities for improvement can make the difference in the production system performance.

There are several tools that can be used to study a system and to help the decision-making process but simulation is probably the only tool able to mimic dynamic and complex environments with considerable interdependencies and stochastic behavior. With simulation, it is also possible to analyze several scenarios and to consider a wide range of performance measures.

In this case, the production line simulated belongs to an electronics industry, the Company X. Operating in this type of industry has become increasingly difficult, because companies compete with high quality standards, with rapid technological changes and with short production cycles. It is known that electronics products are the most complex to produce. The Product Y was the elected product for simulation because is the representative one. The production process can be divided into six phases: manual assembly 1, welding, rework, manual assembly 2, visual inspection and electrical test and packing.

This paper is organized as follows: Section 2 presents an overview of simulation in order to describe the theoretical work context. Section 3 describes the current state of the system and presents the main steps to develop the simulation model. To create the logical models it was used the Arena simulation software. After verification and validation steps, the main results are presented as well as three proposed scenarios for process improvement. Section 4 presents the key conclusions of this work.

2. SIMULATION OVERVIEW

Over the last thirty years, numerous books and papers have focused on the topic simulation because this has been useful and important as a decision support tool. With simulation, it is possible to build, quickly and almost inexpensively, virtual models of complex systems and to do the analysis of different perspectives before making a decision on the actual system (Seleim et al, 2012).

According to Altiok and Melamed (2007) simulation modelling "is a common paradigm for analysing complex systems". One should also refer that simulation modelling "involves the development of descriptive computer models of a system and exercising those models to predict the operational performance of the underlying system being modelled", according to Smith (2003).

Simulation can be described as the process of building a model that represents a real system and allows users to perform experiments with this model, in order to learn about its behaviour and, thus, evaluate the impact of each alternative operation strategy.

The simulation offers benefits such as low cost, rapid and safe analysis system (Wang et al, 2009) and numerous other advantages, such as (Shannon, 1998):

- the ability to identify bottlenecks in information, material and product flows;
- the study of complex real systems, which would be difficult to represent by analytical models;
- the study of alternative layouts without any cost of implantation.

Nonetheless, simulation has also disadvantages, like (Shannon, 1998 and Banks, 1999):

- the (statistical) simulation results are difficult to interpret;
- the collection of data that shows confidence can become a very slow process;
- the simulation alone does not solve problems, only shows the solutions that can solve the problem, so someone must implement the proposed changes.

Negahban and Smith (2014) provide a comprehensive review on manufacturing simulation studies highlighting the application of DES (Discrete Event Simulation) in this context. Simulation models are used for a wide range of complex manufacturing scenarios, from system design to daily operations, as well as covering an extensive set of manufacturing sectors.

In this work, Arena software (a DES simulator) was used as a modelling and simulation tool to study a production line for an electronic product. This software was developed by Rockwell Automation Company. Using SIMAN processor and its simulation language, this software is commonly used to simulate manufacturing processes or services whose purpose is to study the current system performance (Wang et al, 2009).

In the manufacturing Portuguese context, simulation is not a widespread tool for decision-support. This work intends to contribute to evidence the benefits that Portuguese factories can have when adopting simulation practices.

3. CASE STUDY

3.1. Current state of the system and problem formulation
To start a simulation study is necessary to formulate the problem and define the objectives. Relatively to problem formulation, it is intended to develop a simulation model that represents the production system of Product Y. Regarding the objectives, it is important to identify bottlenecks, identify the lead time and the cycle time (because these two indicators are unknown for the Company X) and look for improvements. So, in order to facilitate the analysis of the simulation model, it is intended to gather the following performance measures: throughput, lead time, cycle time, utilization rates, number in queue and time in queue.

3.2. Data collection and information and conceptual model definition
To develop the initial model, data were collected on the production line. Several observations were made in order to observe and measure the operations and to detect possible failures and/or maintenance procedures (e.g., in the welding machine). These observations were also conducted to better understand the details associated with the line. The first annotations included time measurements of each operation, time between failures on the welding machine, operators' work schedule, rework and product rejection rates, number

of daily produced units and product transfer time between different phases. The chosen line has eight operators and a single welding machine.

The data gathered will serve as input to the simulation model and will be treated using the *Input Analyzer* (Arena tool), in order to identify the distributions that best fits the data collected on the shop floor.

The definition of the conceptual model was possible through direct observation of the production process and through process sheet for the product Y. This means that both were analyzed in detail. The conceptual model is illustrated in the following figure and shows, in a simplified and structured way, the sequence of operations required to produce the product. The conceptual model will serve as basis for modeling the real system in the Arena *software*.
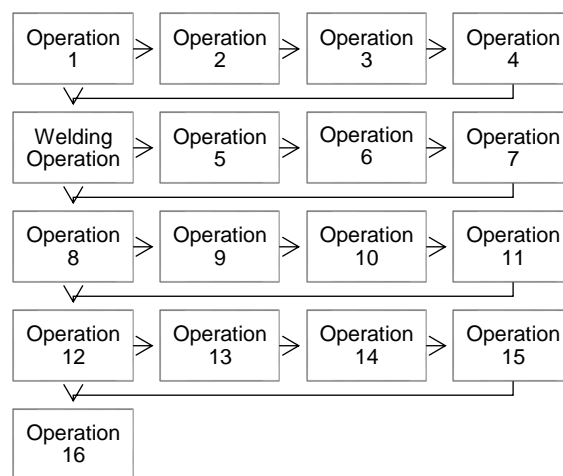
Figure 1: Conceptual model

After several discussions with the area leaders, it was concluded that the conceptual model is correct and complete.

3.3. Logic model construction and verification
After collecting data and some other information, the next step comprises the development of the current logical model. The operation processing times collected on the production line are used in this step.

For the construction of this model, it is intended to codify the conceptual model on a set of logical statements that reflect the real system behavior. The Arena software has several templates that help building the model, such as the *Basic Process*, the *Advanced Process* and the *Advanced Transfer*. To this study, the three templates were used as well as other modules.

An additional relevant fact was the definition of the operators work schedule. For this simulation run, it was considered that a working day has 9 hours, however discounting the lunch break (1 hour) and the snack breaks (20 minutes in total) the production period is 7 hours and 40 minutes. Only the *WeldingMachine* resource is operational during 9 hours, and every 2 hours, this resource stops for, approximately, 5 minutes.

It was decided to simulate a typical planning period of 8 working days. However, it is important to note that the phases are not simultaneously in operation. The following

Gantt diagram (figure 2) shows the days when each phase is being operated. At the end of the day 8, the company has 11 boxes of Product Y.
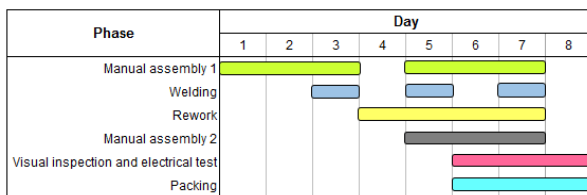
Figure 2: Gantt diagram

To run the model, 50 replications were made. The replication length was 8 days and 9 hours per day. The base time units are minutes. It should be noted that there is a resource, the *SubstituteAF*, which only works in this production in *AF* resource schedule breaks. This resource will not be considered for further analysis.

Manual assembly 1 phase
To build the simulation model, in this first phase, the following modules were used: *create*, *assign*, *batch*, *hold*, *match*, *record*, *process*, *station* and *route*. MF and LB operators carry out the first four operations. The *MF* operator is responsible for *Operation 1*, *Operation 2* and *Operation A*. Two operators perform the Operation 3: *MF* and *LB*.

The model starts with nine *create* modules that create the entrance of the following components: *panels*, *switches B*, *orange cables1*, *coloured cables*, *accessories1, accessories2, switches A1, switches A2 and orange cables2.*

The *Operation 1* (*process* module), performed by MF operator, needs *panels* and *switches B*. In one panel are assembled 100 *switches B*. The resulting entity of the group is synchronized by the *match* module ("*Match 1*") to the panel. After the entities being synchronized, a permanent *batch* ("*Batch 1*") was created. The *record* modules were used to collect the number of entities before and after each operation.

In this phase were used two *hold* modules: one before the *Operation 1* and the other after the *Operation 3*. The entities arriving this module wait in the line until the necessary condition to advance for the next module is verified. It is given the example of the "*Hold welding*", where the following condition was defined: "*(CalDayOfMonth(TNOW)==3)* || *(CalDayOfMonth(TNOW)==5)* // *(CalDayOfMonth(TNOW)==7)*. The entities locked in this module only advance when the day of month is equal to three, five or seven. Otherwise, the entities wait in the line until the necessary day is confirmed. The remaining *hold* modules used in the model, follow the same logic explained here. The entities that are in this module wait until the necessary condition is verified. When it happens, the panel is transferred to the next phase by the *station* and *route* modules.

Here, the "*Entry time*" *assign* module, along with the "*Lead time*" *record* module in the packing phase, has the goal to collect the production lead time.

Welding Phase
This phase is constituted by two operations: *Operation 4* and *Welding operation*. The *Operation 4* has two operators: *AF* and *SubstituteAF*. The *Welding operation* has two human resources (*AF* and *SubstituteAF*) and a welding machine (*WeldingMachine*). To represent the operations, two *process* modules were used. Like the previously stage, the *record* modules here were used with the same goals. As once said, in this phase, it is also present a *hold* module ("*Hold Rework*"), that hold the entities until all of the conditions are checked. Once verified, the entities are transferred to the next workstation ("*Station Rework*") through the "*Route Welding_Rework*" *route* module.

Rework phase
In this phase, the *PP* operator is responsible for two operations (*Operation 5* and *Operation 6*). The construction logic is similar to the previous one.

Manual assembly 2 phase
Here, the following modules were used: *record*, *batch*, *match*, *process*, *decide*, *hold*, *station*, *route*, *dispose* and *separate*. The *RJ* operator is responsible for the *Operation 7* and *Operation 8*. The *Operation 9* includes the *PP* operator and the *Operation 10* contains the *PS* operator. Finally, the last operation in this phase (*Operation 11*) includes the *PS* and *PP* operators. The *PP* operator is also responsible for *Operation 8_1*.

The first operation of this phase is the *Operation 7*. In order to break the entity (panel) in its 25 PCB's units it was used the *separate* module ("*Separate panel in 25*"). From now on, the entity name that runs the system is PCB unit or only PCB and not panel. After *Operation 8*, the decide module is introduced. In the "*Rework operation 8?*" *decide* module, 2% of PCB's units require rework, which means that the remaining units follow for the next operation (*Operation 9*). If the PCB's units need rework, then they go to the *Operation 8_1*. After this operation, another module *decide* (*"Units recovered?"*) is presented. This module is used to decide the path of the units after the *Operation 8_1*. The PCB's units that are rejected, they go out of the system by *the* "*Exit operation 8_1 NOK*" *dispose* module. The recovered units follow to the *Operation 9*.

In this phase, the representative unit became a box containing 25 PCB's units. A temporary *batch* was created and then, immediately, a *separate* module was created too. These two modules ensure that to each operation arrive batches with 25 PCB's units and each batch is separated before being processed, to ensure the individual PCB unit processing.

Visual inspection and electrical test phase
To construct this phase the modules *record, process, decide, dispose, create, match, batch* and *separate* were used. The *PP* operator is responsible for *Operation 12_1* and the *SP* operator executes the remaining operations. To

add a new component to the system it is necessary to create a new entity. It was created the "*Labels*" and the "*Accessories3*" entities from *create* modules. Each of these entities is synchronized with the existing entity in the system by the same logic already explained in the previous phases. The *decide* module is used again. This module determines the need of rework after the *Operation 12*. In this case, there are three possible paths for the PCB's units:

- 2% of the units follow to the *Operation 12_1*.
- 3% of the units cannot be reused and so their final destiny is disposal. These units leave the system through the *"Rejection" dispose* module.
- the remaining units follow to *Operation 13*.

100% of the entities that go to *Operation 12_1* are recovered, which means that these entities also go to *Operation 13*.

Packing phase
To construct this last stage, the following modules were used: *create, process, batch, match, record, assign* and *dispose*. The *Operation B* and *Operation 16* are the only two operations in this phase and the operator allocated to them is *SP*. At this time, card boxes enter in the system. It was used a *create* module ("*Card boxes*") to give the reference to the component input. After the *Operation B* (that consists in assembling the boxes), each box is synchronized with the previous entity through the same construction logic.

According to Sargent (2013), "model verification is defined as ensuring that the computer program of the computerized model and its implementation are correct". The verification was carried out in phases allowing the identification and correction of code errors.

It was also developed a three dimensional animation (3D) for a better analysis of the system. The animation is a very important aspect in this work, because it allows, in a graphical form, to see the whole production process of the Product Y. The animation was also important for model verification and credibility (when showing the results to company managers). Figure 3 depicts a screen shot of the 3D animation model for the current system.
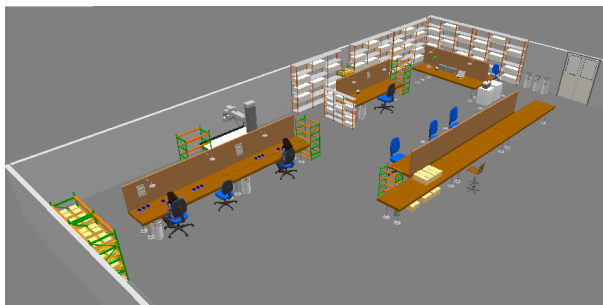
Figure 3: 3D animation model

3.4. Validation and output analysis
Several runs of the model were made. As previously mentioned, to run the model, 50 replications were made, the replication length was 8 days and the working time was 9 hours per day.

In relation to model validation, the author Sargent (2013) refers that the validation is defined "as the substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model". The existence of historical data contributed positively to model validation. To validate the model was considered the throughput performance indicator. Based on this performance indicator, it was found that the results returned by the simulation model (12.28 ± 0,13 boxes, 95% confidence interval) were similar to the real system (11.25 boxes, in average). It can be conclude that the model contains an acceptable adjustment because the performance indicator referred to validate the model is similar to the reality. Relatively to the unknown lead time indicator, the current simulated value is 4431.02 ± 10,50 minutes, which corresponds approximately to 8 days of work. Concerning to cycle time (also an unknown value), the value is 63.82±1,96 minutes. These results seem reasonable and were validated by working operators.

The simulation results in table 2 show the resources utilization rates. It is possible to verify that in general the resources utilization rates are high, except for *AF* and *WeldingMachine* resources. In relation to *AF* the observed low utilization rate (0.26 ± 0,00) is not relevant because this operator is a polyvalent resource and he gives assistance to other productions. The low utilization rate of the *WeldingMachine* (0.17 ± 0,00) it is not also a concern because this resource is used to weld others PCB's (is not fully dedicated to Product Y). It was noticed that the *RJ* resource is underutilized in comparison with the other resources.

Table 1: Resources utilization rate for current model

Resources	Utilization rate
AF	0.26 ± 0,00
LB	0.92 ± 0,00
WeldingMachine	0.17 ± 0,00
MF	0.93 ± 0,00
PP	0.88 ± 0,00
PS	0.88 ± 0,01
RJ	0.50 ± 0,00
SP	0.71 ± 0,01
SubstituteAF	0.20 ± 0,00

After carefully analysing all operations, it is possible to see in table 2 that the bottlenecks are in *Operation 2* and *Operation 10*, because these operations have the highest number waiting and waiting time of entities in the queue. The *MF* resource (allocated to *Operation 2*) has a utilization rate of 0.93 ± 0,00. The resource allocated to *Operation 10* is *PS* has a utilization rate of 0.88 ± 0,01.

Table 2: Entities waiting time and number waiting in queue

Operation	Number waiting (panels and PCB's)	Waiting time (minutes)
Operation 2	58.94 ± 0,13	1620.77 ± 3,49
Operation 10	189.14 ± 2,22	236.73 ± 3,33

3.5. Conduct experiments

After analyzing the output data, three scenarios were considered which consisted of making deliberate variations in the initial model in order to observe the system behavior. For each simulated model, 50 replications were done, the replication length was 8 days and the working time was 9 hours per day. The three simulated scenarios are now presented.

Scenario A: Act on bottleneck operations

The first improvement suggestion is the elimination of the bottlenecks that were identified. To achieve this scenario, the *RJ* resource was allocated to the bottleneck operations, because this is the resource with the lowest utilization rate (without considering the *AF* and *WeldingMachine* resources). It was possible to verify an improvement in the throughput indicator in relation to the actual system. Although the lead time and the cycle time have slightly increased, it is not considered a concern, because the throughput increased about 35%. In relation to the resources utilization rates allocated to the bottleneck operations, its utilization rates decreased. On the other hand, and as expected, the *RJ* utilization rate increased. The *AF* and *WeldingMachine* utilization rates increased too, although it was not significant. The most significant growth was in the *SP*, *PP* and *RJ* resources. It can be concluded that a better use of resources leads to a productivity increase (throughput=16.56 ± 0,17 boxes). It was also visible a decrease in number waiting and waiting time of entities in the bottleneck operations (*Operation 1* and *Operation 10*).

Scenario B: Reduce from 8 to 7 the production days

In order to analyse the impact that it will have the elimination of a production day, it was tested a scenario with this suggestion. To test this scenario, it was removed the "*Hold visual inspection_et*" *hold* module that was holding the entities immediately before the last two phases of the process. The operations of the last phases were realized on days 6, 7 and 8 being now executed on days 5, 6 and 7 (Gantt diagram in figure 4).
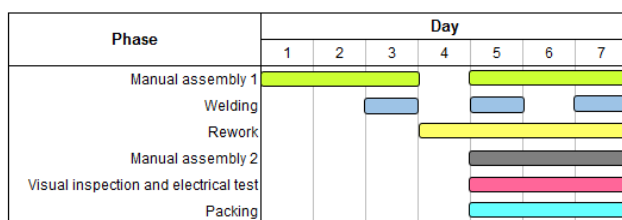
Figure 4: Gantt diagram for scenario B

With this scenario, it was possible to conclude that is viable to decrease one production day leading to better results in lead time (as expected) and cycle time performance indicators. In relation to throughput, the value obtained is good (11.84 ± 0,11) and still acceptable, because the value remains above the production goal (11.25 boxes, in average).

Relatively to the *LB*, *MF* and *PP* resources, it was found that these resources utilization rates increased. *LB* and *MF* presents a utilization rate of 100%, which means that they are well used (perhaps at the limit), such as the *PP* resource whose utilization rate is near to 100%. The other resources utilization rates are similar to the current model.

Scenario C: Situation make-to-stock and make-to-order

The last scenario comprises the make-to-stock and make-to-order logic. In make-to-stock mode the intermediate product is produced for stock and then in a make-to-order phase is when it happens the product differentiation to satisfy the customer's request (Gupta and Benjaafar, 2004). This strategy is known as postponement. With this scenario only the manual assembly 1 phase produces to stock during 7 days. When the company receives the customers' requests, the remaining phases enter in production during the period of 4 days. In order to take advantage of resources utilization rates and maximize the throughput, the following adjustments were made:

- A *hold* module was added after the *Operation 3* and before de *Welding operation*. This module is used to create stock.
- The *LB* resource also supports the *Operation 2*.
- The *Helper1* resource is used to support the following operations: *Operation 6*, *Operation 9*, *Operation 11* and *Operation 14*.
- The *Helper2* is assigned to the *Operation 10*.

With this scenario, the results were interesting as depicted in table 3.

Table 3: Results of scenario C

Throughput (number of boxes)	**29.26 ± 0,13**
Cycle time (minutes)	66.31 ± 0,28
Lead time (minutes)	5177.66 ± 4,78

This scenario takes advantages of resources capacity, because they evidence higher utilization rates, some even close to 100%, with the exception of the *AF* and *WeldingMachine* resources (table 4).

Table 4: Resources utilization rates of scenario C

Resources	Utilization rate
AF	0.27 ± 0,00
LB	1.00 ± 0,00
MaqSoldar	0.19 ± 0,00
MF	1.00 ± 0,00
PP	0.90 ± 0,00
PS	0.88 ± 0,00
RJ	0.88 ± 0,00
SP	0.93 ± 0,00
Helper1	0.91 ± 0,00
Helper2	0.86 ± 0,00
SubstituteAF	0.30 ± 0,00

3.6. Analyze output data

In this step, the results of the current model were compared with the proposed scenarios. In both alternative scenarios, the responses of the system were analyzed. The discussion of the results is based on the elected performance measures.

In figure 5, it is clear that in the three tested scenarios the throughput is above the production objective, which is 11.25 boxes. The scenario A shows that when we act on bottleneck operations, there is an increase in the throughput of about 35%. The scenario B is interesting because it has been found that the reduction in the number of production days does not interfere with the desired throughput (the decrease was approximately 4%). Even in this scenario, there is an improvement on cycle time and lead time. The scenario C shows an increase in the number of produced boxes in the order of 138% over the value obtained by the current model.
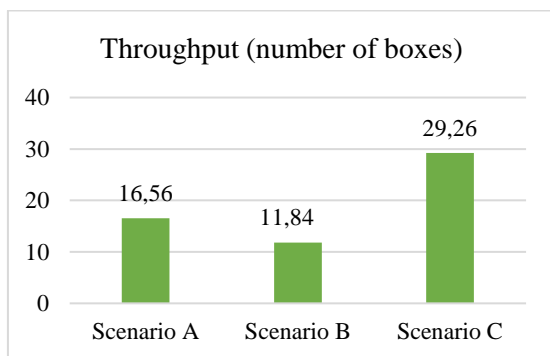
Figure 5: Throughput for the different scenarios

From the analysis of figure 6, it does not exist large discrepancies between the lead times of the different scenarios when compared to the current model. The largest discrepancy occurs in the scenario C, which it shows an increase of around 17%. However, scenario B shows improvements on lead time. The reduction of this performance indicator in relation to the current model is in the order of 6%. On the other hand, in scenario A there was an increase of approximately 7% of this indicator.
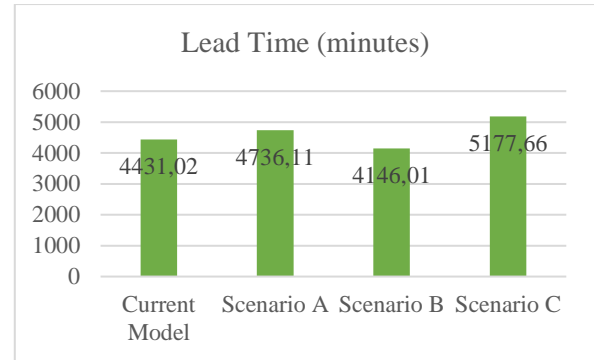
Figure 6: Lead time for the different scenarios

In relation to cycle time, it is possible to verify that just in scenario B was a reduction of this indicator in the order of 4%. This reduction is related to the decrease of a working day. On the other hand, scenarios A and C show an increase in cycle time of approximately 10% and 4%, respectively (figure 7).
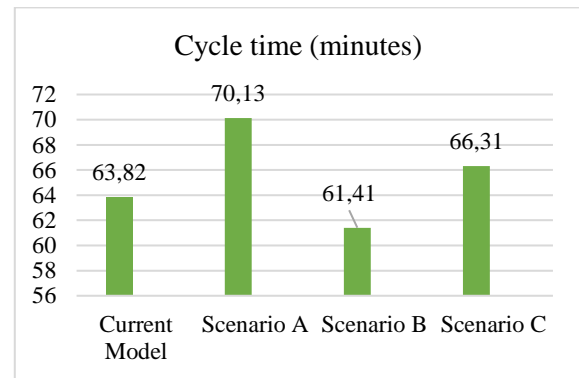
Figure 7: Cycle time

The graph of figure 8 illustrates the resources utilization rates, in percentage, of the current model and of the three considered scenarios. Scenario C is the alternative that shows similar resources utilization rates (with the exception of *AF* and *WeldingMachine*). There are not noteworthy differences and the load is more balanced. Concerning scenario A and B it can be seen that these two scenarios have utilization rates alike to the current model.
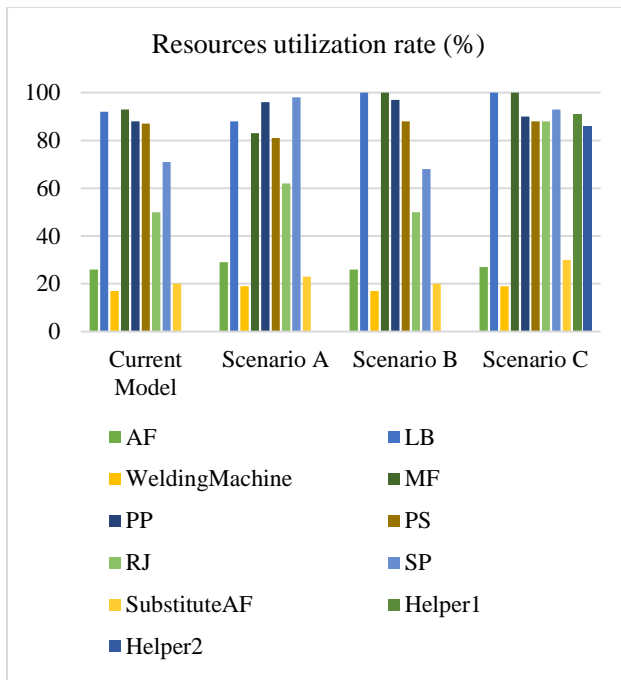
Figure 8: Resources utilization rates

4. CONCLUSIONS

Over the time, simulation has been proving that is one of the most flexible and most used tools in operations management and manufacturing systems, as well as a decision support tool.

In this work, the production system of electronic Product Y was dynamic analysed through modelling & simulation. The quantification of cycle time and lead time values had a huge importance to the Company. The information of cycle time allowed to know the time between the production of successive boxes and the lead time expresses the time that is required for the product Y to go through all process phases, from the beginning to the end. In relation to the three scenarios proposed, it was found that better performances than the current model can be obtained.

The use of computer simulation in a work of this nature, revealed the importance of it in process control and in presenting improvement strategies that make the production system more efficient.

In conclusion, the simulation has proven that is an effective tool for analyzing the production process and to help the decision making process by evaluating several different scenarios and the corresponding impact on a broad set of performance measures. Therefore, simulation can be used as a scientific basis to help in decision making with the advantage that it is not necessary to interfere with the real system. So it is proven the importance of simulation in improving the efficiency of the production process of Product Y. Wherefore, it is justified the use of this tool for analyzing complex systems/processes characterized by numerous interdependencies and stochastic behavior.

The Company X is analysing the simulation results and intends to implement some of the suggested improvements. Other scenarios are also in analysis including different

factors such as production volumes. The 3D animation was a critical factor concerning model results' credibility.

REFERENCES

Altiok, T, and Melamed, B (2007). Simulation Modeling and Analysis with Arena. ISBN: 978-0-12-370523-5.

Banks, J (1999). Introduction to simulation. *Proceedings of the 31st conference on Winter simulation*, **1:** 7–13, doi: 10.1145/324138.324142.

Gupta, D and Benjaafar, S (2004). Make-to-order, make-to-stock, or delay product differentiation? A common framework for modeling and analysis. *IIE Transactions*, **36**(6): 529–546, doi: 10.1080/07408170490438519.

Negahban, A., Smith, J. (2014). Simulation for manufacturing system design and operation: Literature review and analysis. *Journal of Manufacturing Systems*, 33(2): 241-261.

Sargent, R G (2013). Verification and validation of simulation models. *Journal of Simulation*, **7**: 12–24, doi: 10.1057/jos.2012.20.

Seleim, A, Azab, A, and AlGeddawy, T (2012). Simulation Methods for Changeable Manufacturing. *Procedia CIRP*, **3:** 179–184, doi: 10.1016/j.procir.2012.07.032.

Shannon, R E (1998). Introduction to the art and science of simulation. *Proceedings of the 30th conference on Winter simulation*, **1**: 7–14.

Smith, J S (2003). Survey on the Use of Simulation for Manufacturing System Design and Operational. *Journal of Manufacturing Systems,* **22**: 157-171.

Wang, T, Guinet, A, Belaidi, A, and Besombes, B (2009). Modelling and simulation of emergency services with ARIS and Arena. Case study: the emergency department of Saint Joseph and Saint Luc Hospital. *Production Planning & Control*, **20**(6): 484–495, doi: 10.1080/09537280902938605.

AN EXECUTION MODEL FOR EXCEPTION HANDLING IN A MULTI-AGENT SYSTEM

Zina Houhamdi[a], Belkacem Athamena[b]

[a] Software Engineering Department, Al Ain University of Science and Technology
Al Ain Campus, UAE
[b] Management and MIS Department, Al Ain University of Science and Technology
Al Ain Campus, UAE

[a]z_houhamdi@yahoo.fr
[b]athamena@gmail.com

ABSTRACT

Multi-Agent Systems (MASs) are required to exhibit diverse quality attributes like robustness, flexibility, and possibly the ability to accommodate to their agents and context dynamics without external intervention. Exception supervision contributes to the achievement of these goals, and the agent society has proposed many approaches and patterns to supply MASs with exception handling skills. This paper is dedicated to studying this specific subject, particularly in knowledge-based agents systems. The research aim is bilateral: the first aim is to understand the exception concept in MAS. Therefore, we can determine the study objectives to discuss. The second aim is to examine part of these objectives. The suggested methods in this paper describe approaches and outcomes which are estimated to support the agent society and ultimately to support Software Engineering, possibly included within an undergo evolution manner. Previous investigations have concentrated primarily on the systematic perspective of exception handling. Our methods propose to introduce exception utilities in the MAS context.

Keywords: multi-agent system, exception, exception handling

1. INTRODUCTION

MASs consist of multiple autonomous entities called agents, each having different information and/or diverging interests. They are distributed, and complex systems and the agent society try to achieve collaboration and competition between agents to perform their actions in an extremely easy and modular manner. The agent technologies are very widely applied, and we can find in the literature several applications varying from software agents that support people across the network to independent robots in industry. Consequently, MASs are an optimistic approach and technological advancement in artificial intelligence and software engineering (Chopinaud et al., 2006; Houhamdi, 2011).

Since MASs are considered principally as software, and according to computer science history of the past fifty years, the development of reliable systems needs devoted effort, attempts and exercises. Reliability is a system quality measuring system convenience to the user, system accuracy to support the user requirements, and system execution performance and efficiency. Methods for fault detection were proposed and implemented in conventional software engineering to improve the reliability level of software. Contemporary accomplishments assure some of the previously mentioned characteristics in diverse circumstances of non-open and uniform systems. MASs defy existing implementations and focus on complicated applications because they are requested by the users of the software and the organization structure. MASs apply to systems which are heterogeneous, interactive and composed of independent agents.

Among methods to enhance the software reliability, exception handling is reputable and well known as a robust and simple technique (Castelfranchi, 2005; Houhamdi and Athamena, 2011a, 2011b, 2012). Exception handling was included in Programming Languages (PL) since a long time ago to manage unusual situations faced during the code running adequately and methodically. On the other hand, distributed systems have demonstrated that exception handling techniques need particular expansions to apply to these kinds of systems. Simultaneously, achievements in software development have increased the necessity for alternate methods also. MASs also possess characteristics that require re-examining the exception subject.

The purpose of this paper is to understand the concept of exceptions in MASs and to suggest an appropriate architecture for MASs which is open, heterogeneous and features mainly autonomous agents. The agent society has prompted many studies that demonstrated the necessity to handle exceptions in MASs at the system level. This handling includes management and necessary techniques encompassing the management. Solutions proposed up until now apply to a restricted set of MASs only, where usually agents are non-autonomous, and the methods at system level necessitate a perfect collaboration between agents

during exception handling execution (Chopinaud et al., 2006; Houhamdi and Athamena, 2011a). Agent autonomy is an essential quality that must be guaranteed when agents treat exceptions by themselves in the first place; this is considered a requirement of any proposed solution. In this case, exception handling then depends on mechanisms at agent level to manage the weakness of existing solutions and improve them. In an exception situation, the decision is taken by the agent itself to start treating the exception, trust in own expertise or request help from the system level.

The model proposed in this paper guarantees the agents' autonomy by ensuring that the agent maintains control during its processing even when exceptions occur. This approach allows the agent to make an individual decision if a situation is an exception; hence the autonomy is enforced. The approach is specified formally, and its relative architecture is described.

2. BACKGROUND

The actual solutions for exception handling in MASs give the illusion that agents can deal with a situation which isn't Exceptions at Programming level (PE), although they still require taking into account PE situations like unforeseen or uncommon states (Issarny, 2001; Romanovsky, 2001). This illusion is conducive to the notion of Agent Exception (AE) that is described in this paper. To investigate and define an appropriate signification of AE we use the primary definition of PE as starting point. According to function oriented and object oriented PLs, the word "exception" has obtained a specific definition, firmly joined to programming standards, exemplified by the Good enough definition (Goodenough, 1975):

Of the conditions detected while attempting to perform some operation, exception conditions are those brought to the attention of the operation's invoker. The invoker is then permitted (or required) to respond to the condition.

Whenever a program makes a procedure call during its runtime, the procedure needs to evaluate conditions that must be valid before execution. If one condition in the minimum is not approved, the procedure sends a note to the caller declaring that it cannot be performed because of the condition infraction.

Since the MAS constituents (agents, resources, and context) are all programs, we can apply this definition. Nevertheless, the MAS properties and past studies prove that this definition is inappropriate to deal with AEs, due to the autonomy, heterogeneity and openness features. The previous definition of exception constrains the called procedure to assert definitively that a circumstance is unusual. This approach is inadequate to MASs, where ambiguous understanding is likely to arise. An agent is assumed to be autonomous when it can make a decision alone. The interpretation of PE doesn't grant this decision, as represented in Figure 1.

If the method decides that there is an exception, the caller doesn't have an alternative choice to execute. For instance, the method response is an exception object in several object-oriented PLs. However, this solution does not outline the purpose for AEs, as illustrated below in Figure 2.
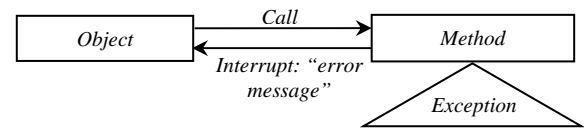
Figure 1: Programming Exceptions
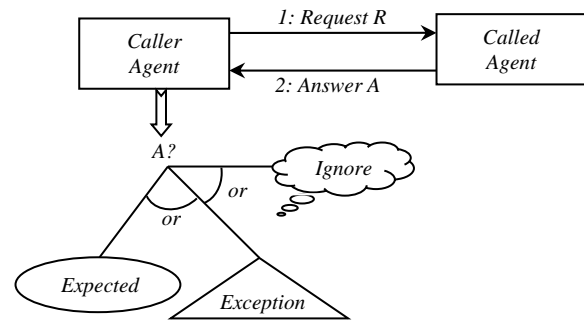
Figure 2: Agent Exceptions

Independent agents must be qualified to determine if a received message from other agents (initially or following a demand) is usual, unusual, or, for example, to be disregarded. Therefore, this declaration is extended to every received message by an agent from agents, the context, or the external environment components.

- *Agent Exception*: By MAS traits, the AE model described in this paper is presented and understood at the agents' level. In other words, the basic element in the exception handling is a whole agent unit, not just the instructions set in its program.
- *Definition*: An AE is the agent's understanding of an observed situation as unusual or not expected.

The previous definition describes the agent's influence in the exceptions situation and, more precisely, during the decision step directly related to the events observed by the agent (Figure 2). During the reception of an event, the agent can determine how to arrange this event. This belief is the main criterion for exception decisions. The agent is a knowledge based entity that performs a protocol. The agent goals and tasks permit the expression of expectations for context modification in the future: agents communicate by sending messages for the purpose of obtaining special outputs that are considered as expectations. Accordingly, the agent is qualified to classify a received message as unexpected if the message does not agree with its expectations.

The PEs are different from AEs. The former concerns the event and the later concerns the event understanding. Agents, which are autonomous, can

control themselves and determine how to deal with events; this is the principle of AE.

Someone can be in dispute with this definition and argue: independent agents are usually assumed to operate in a society setting. A society establishes strong relations. These relations assume that, even with independency, agents act based on received demands. Such context is relevant in a closed systems environment in which a human user supervises all system components. The illustrative supervisor-worker pattern assumes that workers comply with the supervisor. However, in an open-system, the designer of particular agents desires to preserve total control of its agents and wants to decide on how to reply to requests from other, possibly anonymous, agents. Notwithstanding strong relations among two agents, independency is conducive to the previous definition. It is the responsibility of the agent alone to determine how to deal with an event.

This definition does not go against the strong relations decided by societies. Coordination ability is merely considered as an extension of agents' autonomy. When an agent has concluded that the event is unusual to its understanding, the agent can improve its conclusion based on strong relations. For instance, an 'operator agent' can decline to abort when the mandate is from a 'supervisor agent,' for example, if the two agents work in separate organizations but share a virtual space for cooperation.

In MASs, AEs are associated with agent tasks, and they influence the agent level. PEs are related to situations. They influence the agent at the code level. This description is illustrated in Figure 3.
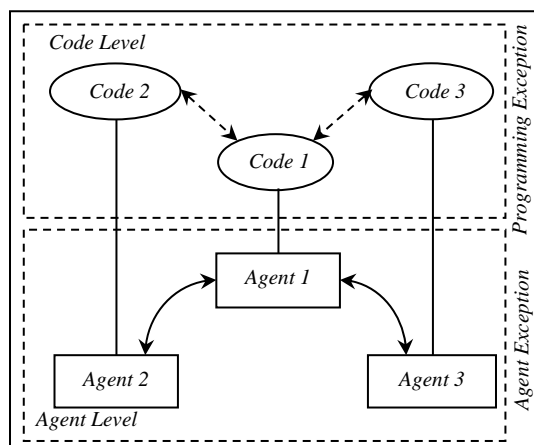
Figure 3: Exception Levels in MAS

The following section's purpose is to determine the exception area of agents and to describe the relationships between the PE and AE.

Note that the PEs can cause AEs. By way of illustration, an unforeseen agent ending as result of a PE, such as a null pointer, impacts the system organization directly. This PE will then cause an AE, 'agent death' (Houhamdi and Athamena, 2011a; Klein et al., 2003). In this situation, the remaining agents require rearranging their tasks to compensate for the agent death. In this manner, the rearrangement is an exception that happens at the agent level.

However, some PEs, occurring in an agent, will not generate an AE. As case in point, network exceptions, where a handler retries the network connection to approach this issue, are commonly controlled at the code level. Accordingly, the agent pursues performing its task.

Nevertheless, AEs do not generate PEs. Particularly, agents are not aborted by the AE's occurrence. That is to say, AEs do not provoke the agent code to contend with a malfunction. AEs do not cause PEs because AEs are discovered in input messages using a particular estimation method. The message is treated as an AE, while the program is correctly performed and no PE is revealed. The agent proceeds its activity by managing the anomaly or disregarding the message and continuing the following iteration. During this method, the agent status and its code are consistent with the typical progression without producing any PE.

The previous characteristics single out one-sided relations among the two exceptions kinds, which are illustrated in Figure 4. It shows the relations between the exception areas that are conceived for a MAS. The PEs can, in some situations, generate an AE, although the opposite is impossible.
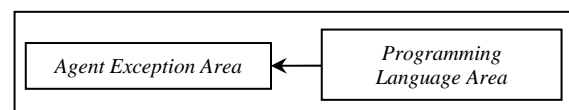
Figure 4: Exception Area Mapping

MAS Exception: The relation between PEs and AEs allows classifying the uncommon circumstances that confront an agent. This section purpose is to identify exception classes to make their investigation easier and also to organize the handlers' classes that will be produced.

- *The Identification Axis*: Mainly, there are two classes of exceptions depending on if the exception is identified or unidentified. If there is a handler to manage the exception then it is classified as identified; otherwise, the exception is unidentified. In PE, unidentified exceptions provoke an early program ending, because it cannot manage the event and possibly risks damage to the physical devices or operating system. However, in AEs, unidentified exceptions signify that the agent lacks the necessary knowledge to manage the event in the present situation. Nevertheless, the agent state is still reliable and can make a decision by its qualifications. The simplest solution is to not pay attention to the event (like Does-Not-Understand in Smalltalk), but the complex solution is to exploit the circumstance, like KGP agent (Antonis et al., 2004).

- *The Coverage Axis*: According to coverage of the agent exception, we define two types: Alone and Team. If the agent can manage the exception without the assistance of other agents, the level is labeled as Alone. If the agent needs to interact with other agents to manage the exception, the level is thus assumed as Team. In an agreement contract, if a customer gets a remarkable bid, for example, inferior 15% of the estimated cost proposed by the customer, in this case, the exception level is Alone and can be managed quickly. The customer updates its status so that this bid will gain the call for proposal. At this point, the customer continues the procedure execution to agree on the bid officially and reject the remainders without additional cooperation needed to manage this special circumstance, so this exception is an example of the Alone type. On the other hand, an exception such as declaration postponement is an example of a Team exception. A supplier declares a deferment to the customer, who replies by allowing a deadline prolongation to every supplier.
- *Handler Description*: In this section, we define the exception classes, and then we categorize the applicable handlers. By default, the death of an agent is presumed as a common situation that is classified as an identified AE (Klein et al., 2003; Miller and Tripathi, 2004). Still, this AE can be dealt with in either an alone or team manner, according to the handler type used by the agent to control the situation.

There are two objectives behind exception organization. The first one is to guide developers to design handlers or methods to elaborate during execution. Based on the MAS application, certain handlers' types are essential and others are unessential. Handlers for the identified exception need particular methods to seek or create them; this is very expensive for some systems. The other objective is to help agents in the decision process. Based on the exception, the agent examines a special handler.

Handlers' classes are described by the abbreviation in Table 1. For example, IAA refers to handlers for identified exceptions at Agent Alone level, while UC represents handlers for an unidentified exception at Code level.

Table 1: Exception Classes

	Agent level		Code level
	Alone	**Team**	
Identified	IAA	IAT	IC
Unidentified	UAA	UAT	UC

- *Handlers Classification*: Since there are two manners to manage exceptions (Alone or Team), agents will confront a difficulty in deciding the availability of the handler of every category. Team handlers are a costly process, particularly in distributed systems, and they overcome the distribution advantages because they extend the computation cost with interaction expenses. Therefore, the agent prefers handlers that handle exceptions in an Alone style. Further, the interaction complexity in MASs accentuates this choice.

The exception handling method emphasizes representing exceptions on the communication protocol because "MAS" principally refers to cooperative agents that perform as reported by the communication protocol. In this paper, we use the word "exception" to denote AE if there is no confusion with PE.

3. AGENT EXECUTION MODEL

The AE definitions have impacted the agents' execution model and framework. The best agent frameworks use the Belief-Desire-Intention model, including the Jason and Jadex architectures or the KGP architecture. However, these models suffer from two weaknesses regarding AEs. Exception handling is not processed explicitly in the agent execution model and no distinction is made between exceptions. The exception is treated as PE and depends upon the languages services. The AE's handling, on the other hand, needs to consider the MAS properties, and the best practices propose to distinguish clearly between the application logic and the exception handling. This work proposes an agent execution model which includes exception handling so that the previous distinction is achieved.

In general, the execution model for MAS is iterative, traditionally a cycle of perception, reasoning, and action. Our model uses the same iteration but extends the perception and action activities to relevantly arrange the reasoning activity in exception situations, considering the agent independency property.

The remaining of this paper defines the proposed agent's framework mainly by describing its protocols, handlers, and knowledge, and after that the execution model.

Protocol and Handler Structure: AUML and allied studies have modeled handlers and protocols using sequence diagrams or graphs. We decided to describe handlers and protocols as diagrams (more specifically directed trees) to set up the description formally. The root represents the initially transmitted message. The tree is organized by using the relation *R,* defined as follows: If *T* is a directed tree; *L* represents the leaves kit $(L \subset T)$ and *M* the edges kit. The edges represent operations such as sending a message in handlers and protocols.

R is a non-symmetric, non-reflexive and transitive binary relationship. *T* verifies the following structural properties:

1) $\forall m \in M \setminus L, \exists m' \in M, m \, R \, m'$
2) $(\forall m \in M \setminus L, suc_T(m) = \{m' \setminus m \, R \, m'\}$

3) $\forall m \in M \setminus \{root\}, \exists m' \in M, m' \, R \, m$

The first definition declares that all sent messages have a successor except leaves. $suc_T(m)$ represents the successors set for a given edge of T in definition two. Definition three states that all sent messages have a predecessor, except the root.

In the case where protocol comprehends a loop in its description, the tree specification utilizes the cycles unrolling over the tree branches. Such unrolling action is usual, e.g. Petri nets.

We describe two sets: the M messages Set, the H histories Set, and $\phi \in H$ (Empty execution). The execution continues based on the acquired message kind and the handler (h) and protocol (p) state which the agent executes. "*Perform*" defines the progress of the agent running the protocol and the handler.

Perform: $M \times H \times H \rightarrow H \times H$

$$(m, H_p, \phi) \quad \begin{cases} (H_p \cup \{m\}, \phi), \text{if } m \neq end \\ (\phi, \phi), \text{if } m = end \end{cases}$$

$$(m, H_p, H_h) \quad \begin{cases} (H_p, H_h \cup \{m\}), \text{if } m \notin \{end_p, end_h\} \\ (H_p, \phi), \text{if } m = end_h \\ (\phi, H_h \cup \{m\}), \text{if } m = end_p \end{cases}$$

(m, H_p, ϕ) describes the protocol p's execution. The execution history evolves during message processing (sent and received), and the processing terminates when the end is obtained; in this case, the protocol history is cleaned out but (m, H_p, H_h) represents a handler execution. Consequently, the handler treatment follows the protocol execution. When m is end_p, the handler starts after the protocol interruption. Finally, when m is end_h, the handler processing is completed with success, and the protocol execution is restarted.

Figure 5 represents a general execution model of agents which contains three layers. We explain them consecutively in the following paragraphs. The description depends on algorithms which are independent of the application domain.

First Layer: This layer contains message reception, pertinence checking and belief comparison, which are the basic phases in the agent processing model. The agent collects the messages from its inbox. They are sent to pertinence checking to discard messages that are not important for the agent, as reported by the relevance table. Pertinent messages are matched to the agent beliefs in the *Beliefs Table*. If an equal entry is located, then the output is an expected message, otherwise *Taking Decision* is started when the message is unexpected, and the *Handler Selection* is activated.

In the *Decision Process*, which is the agent's brain, the message is treated to define the agent action, if any, and

update the agent knowledge. Besides this task, the *Decision Process* performs continually and does not need an input to generate an output. This function is not illustrated in Figure 5 because it is not related to exception management. Nevertheless, it is essential because it is the 'dynamic' part, indispensable for the agent to induce actions.
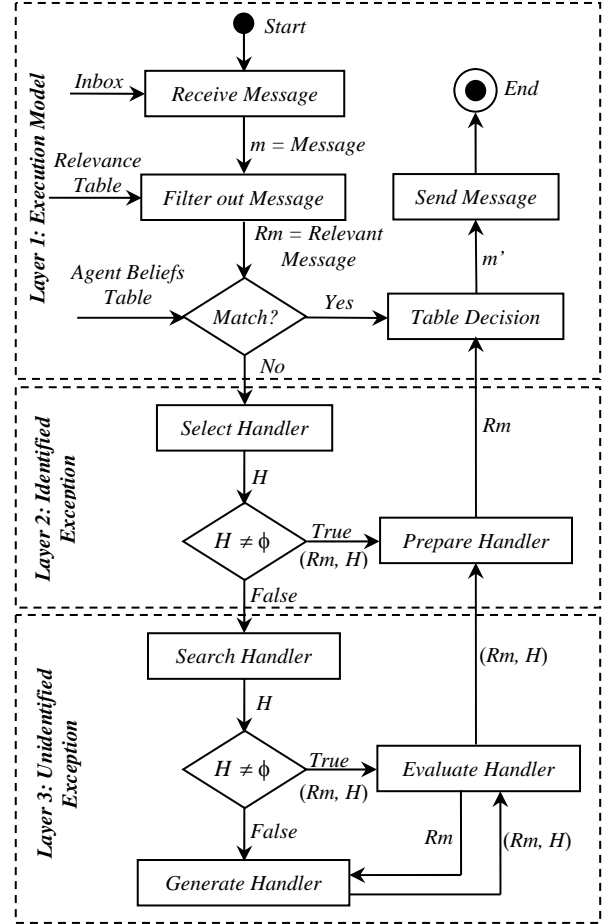
Figure 5: Agent Execution Model

Second Layer: The agent finds an unusual situation when a match is not found in *belief matching* phase:

- *Handler Selection*: concerns *Identified Exceptions*, i.e. the agent possesses a handler to manage the found exception. Unforeseen messages are forwarded to the *Handler Selection*; this later explores the handler table for a relevant handler. If a table entry has a requirement that meets the message, a handler is located. If diverse handlers are located, the *Favorite* method determines which handler is better for the agent, according to its environment and state. The *Favorite* method is thus domain dependent. *Favorite* methods use metrics to appraise handlers (e.g., the handler complexity).

- *Handling Preparation*: when a handler is located, it is sent to the *Handling Preparation* phase which interrupts the protocol affected by the unexpected message, starts the execution of

the handler and specifies that the interrupted protocol needs to be appraised at the termination of the handler by producing a protocol dominion for the handler. Thus, the agent decides whether to continue the interrupted protocol or to abort it. The output of this phase is a message which is sent to the *Decision process*, apt to treat the exception.

Third Layer: If the selection phase does not find a handler, the agent faces an *Unidentified Exception*; in other words, the agent does not possess a handler:

- *Handler Search & Evaluation*: The agent attempts to find a *Handler* by communicating with other agents or with a handler depository. A successful search provides a handler. The *Handler Evaluation* process analyses the handler fitness, keeps the agent autonomy concerning this foreign handler, and saves the exception class and its handler in the handler table for future use. Usually, the evaluation process is complex but we use a simple method by assuming that a handler is considered adequate if it reaches a situation allowing the suspended protocol to continue its processing.

Formally, the handler adequacy, H, is adequate if and only if $H_n = P_{it}$,

where H is a Handler: $H = (H_i), i \leq n$

and P is a Protocol: $P = (P_i), i \leq n$ interrupted at the

statement P_{it}

and $end_p = end$

More precisely, the agent accepts the handler if it directs the processing to the desired situation before the exception occurrence. However, this basic verification does not certify that each statement in the handler is adequate for the agent. This generic approach is domain dependent.

- *Handler Generation*: In case the handler search fails to find a handler (or the found handler is unacceptable), the agent tries to generate a *Handler*. In the proposed approach, this phase unavoidably engenders a default handler H_d if no acceptable one is found; this H_d is important since it ensures the execution progression. The H_d will disregard messages during a period before declaring the failure of the protocol. For example, the handler H_g produced to expect the message m during two times related to the protocol P is described as:

$$H_g = \big((m, end_h)|m, \tau(ignore)\big),$$
$$\big((m, end_h)|m, \tau(ignore)\big),$$
$$\big((m|(m, \tau(update(p)), end_p, end_h)\big)$$

The agent assumes it will get *m* two times, and then sends end_h to the protocol. Each time, the message is ignored by the agent if it does not match its beliefs. After three non-matching messages, the protocol state is updated and a message end_h indicating the protocol annulment is forwarded to all the agents.

Table 2 presents the model complexity in all cases, with the following legend: d_p (decision process), h_s (handler selection), *eval* (handler evaluation) and *p* (protocol).

Table 2: Model Complexity Table

Case	Complexity
No Exception Handling	n_{dp}
With Exception Handling	$n_{dp} = \max(n_{dp}, O(n_p))$
Identified Exception	$\max\big(O(n_{hs}), O(n_d)\big)$
Unidentified Exception	$\max\big(n_{eval}, O(n_{hs}), O(n_d)\big)$
Unidentified Exception Default Handler	$\max\big(n_{eval}, n_{dh}, O(n_{hs}), O(n_d)\big)$

Since agents execute a few protocols concurrently, the cost is reasonable in comparison with MASs without exception handling. However, in the case of heavy agents, we should consider other approaches for fault tolerance.

4. AGENT ARCHITECTURE

Figure 6 represents the agent framework. It is comparable to existing agent architectures and it integrates special components for exception handling. In particular, these components can be taken out from the architecture if the agent does not need this feature or as a result of design choices.
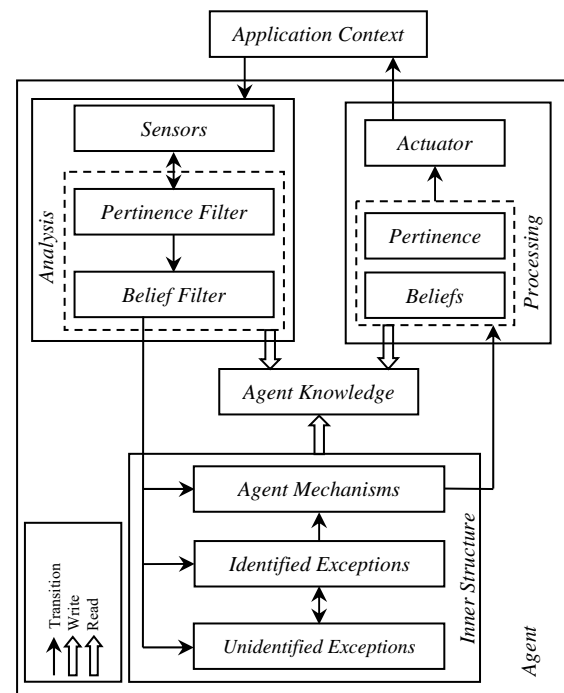
Figure 6: Agent Architecture

The agent framework includes four principal components to be harmonious with the execution model: the internal description, internal processes, perception, and operations.

The correspondence between the execution model and the framework is illustrated in Table 3. The left column names the execution model components defined in the third section. The middle column lists the framework components described in the fourth section. Parts of these components are common architecture components and the others are arranged in different framework components in the right column.

Table 3 shows that the proposed agent architecture includes the required features to implement the whole execution model. The constituents of the architecture support two characteristics of agent models with exception handling abilities:

- The Architecture Components supply the developer with an advanced and general architecture prototype (Athamena and Houhamdi, 2017). The column guides architecture refinement.
- The Architecture Constituents separate the application concerns from the exception concerns.

Table 3: Correspondence between Agent Architecture and Agent Execution Model

Execution Model Constituents	Architecture Constituents	Architecture Components
Receive Message	Sensor	Perception
Filter out Message	Relevance Filter	Perception
Compare with beliefs	Expectation Filter	Perception
Take Decision	Base Mechanism	Internal Processes
Select Handler	Identified Exception	Internal Processes
Prepare Handler	Identified Exception	Internal Processes
Search Handler	Unidentified Exception	Internal Processes
Evaluate Handler	Unidentified Exception	Internal Processes
Generate Handler	Unidentified Exception	Internal Processes
Update State	Generation	Operations
Send Message	Operator	Operations
Tabular Knowledge	Internal Description	Internal Description

These two characteristics are essential for the developer since the refinement and separation of concerns are well-known as good practices in Software Engineering.

5. SIMILAR WORKS

Exception handling research covers investigations in artificial intelligence and software engineering. As MASs are also related to these areas, several tangible implementations are noticed in the MAS theory or their building practices. Nevertheless, these achievements do not satisfy the essential qualifications to approach MAS exceptions. PEs have established theories, but they are not applicable to MASs adequately because of MAS features such as heterogeneity, openness, and autonomy. These approaches can deal with the openness and heterogeneity issues; however, they cannot manage the autonomy characteristic. One remarkable effect is that there is no attempt to provide a precise description of the exception notion in a MAS, particularly in the agent society. Several illustrations are clarified in depth,

e.g. agent death, but the exception notion remains implicit. The most notable works that approach exception handling in MASs are:

- *The Sentinel Architecture*: sentinels are agents inserted in a MAS software to supply the application with a fault tolerance capability level (Athamena and Houhamdi, 2017; Hägg, 1997). The Sentinel supports the agents in their communication. Sentinels are specially designed for fault detection and recovery. The detection of an exception during agents' communication activates the sentinels which try to resume a reliable situation. The problem is that the sentinel violates the agent paradigm assumptions (encapsulation is not respected and, consequently, neither is agent autonomy).
- *Sentinel-Like Agents*: extends the sentinel model with a reliability database (Klein et al., 2003) where the failed agents are stored. The database leads sentinels in recovery functions to reduce the needed time to recover. The Sentinels operate similarly to the Hagg initial model without inspecting agent interiors, as to enhance the agents' autonomy. However, this system suffers from two weaknesses: the agent autonomy is violated (as Sentinels can change agent message), and the exception handling system is fragile in the case where sentinels cannot perform their activities when executing a handler.
- *Commitment Protocols*: consider exception management in the business milieu (Mallya and Singh, 2005). This model uses commitment protocols to describe the agents' communication in an open system. This approach preserves the agents' autonomy. However, it is principally abstract, and it requires validation in real world applications.

SaGE in the Mad-Kit Platform: SaGE is a framework which adds to the Java exception management system services to manage problems related to autonomous agents in the Mad-Kit (Souchon et al., 2004). In Mad-Kit, an agent possesses roles and provides services to other agents. Exceptions can happen at the role, service and agent level.

SaGE follows the agent exception description, but does not ascend to the heterogeneous system level because it uses just benevolent agents. However, SaGE contributes notably to agent-oriented engineering by including exception handling, to wit the exception expansion according to the particular organization model and the cooperative exceptions.

Our approach endows an individual agent with relevant potentialities concerning exception situations and conforming to agent characteristics. Existing systems satisfy part of the agent features, but our model approaches the autonomy issue appropriately. The principal model advantage compared to other systems is

its robustness and reduction of the developer job; in this manner, the developer will be able to focus on more important processing matters.

6. CONCLUSION

MASs should have many features such as robustness, flexibility and automatic adaptation to the agents' dynamics and context. Exception handling is one of the techniques that contribute to the achievement of these features, and the agent society has proposed diverse approaches to supply MASs with exception handling facilities.

Agent exceptions need certain special approaches to assist developers in writing pertinent handling programs. Our model improves the agent framework's ability to analyze the messages and identify the unforeseen ones during collaboration protocol. The presented approach is integrated into the agent framework to allow the developer to concentrate on writing suitable handlers' programs. The model uses these handlers to manage exceptions whenever needed. The framework supplies agents with the model, thus they treat exceptions autonomously.

The proposed approach discusses exception handling at the agent level, which treats agent level and system level exceptions in a decentralized manner (complex and inefficient), however strong and adaptable if the MAS faces exceptions. The agent level hides the problems related to the system robustness because agents are autonomous and the MASs are open and heterogeneous, and the system level improves the system performance.

Finally, a future improvement of the proposed approach is an extra investigation into the handler generation methodologies in different circumstances to make the agents more autonomous when encountering diverse, unusual events. Further, an interesting domain is the agent evaluation of management approaches proposed by the collaborative agents in the MAS.

REFERENCES

Antonis, C., Paolo, M., Fariba, S., Kostas, S., & Francesca, T. (2004). The KGP model of agency. In the 16th European Conference on Artificial Intelligence (pp. 28–32). IOS Press.

Athamena, B., & Houhamdi, Z. (2017). An Exception Management Model in Multi-Agents Systems. Journal of Computer Science, 13(5), 140–152.

Castelfranchi, C. (2005). Mind as an Anticipatory Device: For a Theory of Expectations. In International Symposium on Brain, Vision, and Artificial Intelligence (pp. 258–276). Springer Berlin Heidelberg.

Chopinaud, C., El Fallah-Seghrouchni, A., & Taillibert, P. (2006). Prevention of harmful behaviors within cognitive and autonomous agents. Frontiers in Artificial Intelligence and Applications, 141, 205–209.

Goodenough, J. B. (1975). Exception handling design issues. ACM SIGPLAN Notices, 10(7), 41.

Hägg, S. (1997). A sentinel approach to fault handling in multi-agent systems. In Australian Workshop on Distributed Artificial Intelligence (pp. 181–195). Cairns, Australia: Springer Berlin Heidelberg.

Houhamdi, Z. (2011). Multi-Agent System Testing: A Survey. International Journal of Advanced Computer Science and Applications, 2(6), 135–141.

Houhamdi, Z., & Athamena, B. (2011a). Structured Integration Test Suite Generation Process for Multi-Agent System. Journal of Computer Science, 7(5), 690–697.

Houhamdi, Z., & Athamena, B. (2011b). Structured System Test Suite Generation Process for Multi-Agent System. International Journal on Computer Science and Engineering, 3(4), 1681–1688.

Houhamdi, Z., & Athamena, B. (2012). Monitoring and Diagnosis of Multi-Agent Plan : Centralized Approach. European Journal of Scientific Research, 87(4), 541–551.

Issarny, V. (2001). Concurrent Exception Handling. In Advances in Exception Handling Techniques (Vol. 2022, pp. 111–127).

Klein, M., Rodriguez-Aguilar, J. A., & Dellarocas, C. (2003). Using Domain-Independent Exception Handling Services to Enable Robust Open Multi-Agent Systems: The Case of Agent Death. Autonomous Agents and Multi-Agent Systems, 7(1/2), 179–189.

Mallya, A. U., & Singh, M. P. (2005). Modeling exceptions via commitment protocols. In Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems - AAMAS '05 (p. 122). New York, New York, USA: ACM Press.

Miller, R., & Tripathi, A. (2004). The guardian model and primitives for exception handling in distributed systems. IEEE Transactions on Software Engineering, 30(12), 1008–1022.

Romanovsky, A. (2001). Exception handling in component-based system development. In 25th Annual International Computer Software and Applications Conference. COMPSAC 2001 (pp. 580–586).

Souchon, F. F., Dony, C., Urtado, C., & Vauttier, S. (2004). Improving Exception Handling in Multi-agent Systems. In Lecture Notes in Computer Science (pp. 167–188). Springer Berlin Heidelberg.

AUTHORS BIOGRAPHY

Zina Houhamdi is an Associate Professor at the Department of Software Engineering, Al Ain University of Science and Technology, UAE. Her research work has been published in several academic journals and has been presented to scientific conferences. Her research areas of interest are data quality, agent-oriented software engineering, software testing, goal-oriented

methodology, software modelling and analysis, Petri nets, and formal methods.

Belkacem Athamena is an Associate Professor at the Department of Management and MIS, Al Ain University of Science and Technology, UAE. His research focuses on information system, data quality, data integration, software testing, software modelling and analysis, Petri nets, and formal methods. He has published in various recognized international journals and conference proceedings.

DESIGN BY CONTRACT OF CYBER-PHYSICAL SYSTEMS DRIVEN BY SIMULATION AND BASED ON PROPERTIES MODELING

Andrea Tundis, Max Mühlhäuser

Telecooperation Lab, Department of Computer Science, Technische Universität Darmstadt
Hochschulstrasse 10, 64289, Darmstadt, Germany

tundis@tk.tu-darmstadt.de, max@tk.tu-darmstadt.de

ABSTRACT

Requirements elicitation and analysis is the basis for the successful development of a Cyber-Physical Systems (CPS). The misunderstanding of one or more requirements, due to different skills and knowledge between stakeholders and engineers, could com-promise the success of an entire project with harmful consequences. Usually, agreements on the system to be delivered and related expected results are based on textual requirements with a big lack of not being computationally verifiable and difficult to trace. To this purpose, the employment of innovative engineering tools for supporting the modeling and the verification of system requirements represent a viable solution. In this context, the pa-per proposes the exploitation of a Properties Modeling (PM) approach combined with Simulation techniques as Design-by-Contract method for CPS. In particular, PM is adopted for sup-porting the definition and the representation of system requirements and constrains as computable entities, whereas a Simulator is developed and exploited for enabling their automatic verification. Such combination is used as tool for defining requirements and conditions and verify their fulfillment before the sys-tem deployment. The results gathered from the simulation represent the contract on which the parties can agree for the realization of the actual system. The approach is exemplified in the Smart Grid domain.

Keywords: Properties Modeling, Requirements Specification, Simulation-based Verification, Design-by-Contract, Cyber Physical Systems, Smart Grids

1. INTRODUCTION

According to the International Council on Systems Engineering (INCOSE 2017), the causes that determine the success or the failure of a product, a service or an entire project, usually rely in the bad management of those factors related to its life cycle (INCOSE Book 2015). Since cyber-physical systems (CPS) (Danda, Rawat, and Rodrigues 2015) become more and more complex, the requirements to be fulfilled, both in terms of functionality and performance, are of primary interest. As a consequence, there is need to include them among the operational constraints from the beginning of the design stage. Especially, in application domains such as power plants, medical appliances, aerospace, and automotive, some non-functional requirements (such as reliability, availability, maintainability, safety and security) have to be guaranteed and comply to standard specifications and regulations (Lahtinen, Johansson, Ranta, Harju, and Nevalainen 2010; Rierson 2013; Furfaro, Garro, and Tundis 2014; Furfaro, Gallo, Garro, Sacca, and Tundis 2016). Indeed, the violation of some requirements can generate the failure of a project whose impact can be measured in terms of: (i) economic and temporal; (ii) motivational; (iii) individual and organizational stress; (iv) the destruction of value and corporate reputation, and even worse as (v) loss of human lives.

Unfortunately, because of the high heterogeneity of CPS in terms of system components and functionalities to be provided, the management and the manual checking of the requirements is a challenging task to be performed. Thus, maintaining the compliance between the requirements and the actual system becomes increasingly difficult and unproductive to be performed. So, there is the need, from one hand, to clearly define constraints and requirements, and from the other hand to be able to verify them, possibly before the realization of the system or even before an advanced stage of its development is reached. This, in turn, implies to address some important challenges ranging from (Garro and Tundis 2015; Seshia, Hu, Li, and Zhu 2016; Falcone, Garro, and Tundis 2014) (i) identification of concepts and notations for modeling requirements; (ii) approaches for integrating design and requirements; (iii) automatic mechanisms that provide indications on the level of fulfillment of requirements during the system development.

In this panorama, the use of engineering tools in terms of innovative methods and techniques represent a profitably solution. Particularly interesting is the Design by Contract (DbC) method, typical of software engineering (Ozkaya and Kloukinas 2013). According to the DbC, the involved entities in the design have obligations towards other entities on the basis of well-formalized rules. A functional specification, called *contract* is created for each software module before it is

implemented. Finally, the overall program execution is seen as the interaction among the various modules bounded according to these contracts (Klaeren, Pulvermüller, Rashid, and Speck 2001).

In this paper, the main purpose is to exploit the DbC method (a) as a contracts–based validation tool to support the development process of cyber-physical systems by simulation; and (b) to reuse the results gathered from the simulation as *contract* to verify the actual system operation after its deployment. To this aim, a possible solution is represented by Properties Modeling (PM) approach (Otter et. Al 2015; Nguyen 2014; MODRIO 2016) for modeling *contracts*, combined with the use of Simulation techniques to automatically verify them. Specifically, PM is recently considered among the major research fields of SE (INCOSE 2017). It deals with system requirements and constraints from the early stages of the system development process, passing through its release, up to its operational maintenance (Garro and Tundis 2015; Nguyen 2014; Rubio-Medrano, Ahn, and Sohr 2013). The combination of system properties along with simulation techniques allow to support and evaluate the goodness of the system through the virtual verification of the requirements. In this case, the modeling of *requirements* in terms of *properties* represent the *contract* among stakeholders and designers. The results gathered from the simulation represent "formalized requirements", that the actual system have to comply.

The rest of the paper is organized as in the following. Section 2 provides related work and motivation; Section 3 describe the proposal and how to combine the Properties Modeling and Simulation to enable the Design by Contract method for supporting the cyber physical system development. A case study along with a Simulator is described in Section 4, whereas a discussion is reported in Section 5. Conclusions are drawn in Section 6.

2. RELATED WORK AND MOTIVATIONS

2.1. Background on Design by Contract
Design by Contract is a typical methodology for supporting software design and development (Meyer 1992). Usually, designers and stakeholders collaborate to define the software specifications in terms of abstract data, routines and functions clearly and testable. Generally, the contract is defined through a set of rules called *pre-conditions* and *post-conditions*. In particular, *pre-conditions* state what a data, a function or method has to satisfies before it is used (e.g. the expected input); whereas the *post-conditions* state the status that the output (the data or the routine) will be comply after it has been used. For example, a contract takes the following general form:

if *pre-conditions* (A) == true

then *post-conditions* (routine(A)) == true;

So, if the caller satisfies the initial conditions then a "correct" output is guaranteed, but if the initial conditions are violated then the correctness of the output cannot be guaranteed. So using a composite approach, the correctness of each software module can rely, in theory, on the correctness of the software modules that are used, as long as their preconditions are satisfied (Jazequel and Meyer 1997).

A lot of interest is shown towards this approach as evidenced by the different research efforts already available in literature. As an example, in many application domains where the software engineering development process is involved, the DbC is strictly related to the programming language. A popular example is represented by the DbC in Java language (Zimmerman and Kiniry 2009), where introduction of specific keywords, based on the concept of *assertion*, are introduced into the language. The main advantage in this case is that the developer and the designer share the same notation, so the level of ambiguity and misunderstanding is reduced. From the other side, the code to build the software and the one to check its correctness results mixed and highly coupled. As a consequence, software readability and maintainability become harder.

Another contribution is represented by the DbC with JML (Leavens, Baker, and Clyde 2006; Leavens and Y. Cheon 2013), a Behavioral Interface Specification Language (BISL). Besides pre- and post-conditions, it also allows assertions to be intermixed with Java code, and it is designed to be used by working software engineers. However, whereas Java expressions lack some expressiveness that makes more specialized assertion languages convenient for writing behavioral specifications, JML solves this problem by extending Java's expressions with various specification constructs, such as quantifiers.

Another approach to enable DbC is based on a Temporal Logics (Cimatti and Tonetta 2016). An implementation of it is available in OCRA (Cimatti, Dorigatti, and Tonetta 2013) that is very suitable to represent temporal relationships among events. It provides a support to contract-based design, ranging from the formal specification of the architecture and contracts to the automatic analysis of the refinement, implementations, and safety of the contract specification (Nuzzo, Sangiovanni-Vincentelli, Bresolin, Geretti, and Villa 2015).

In (Heckel and Lohmann 2005) is provided another contribution regarding the employment of the DbC method, by representing contracts in terms of graph transformation rules, for testing web services against their description. In (Seshia, Hu, Li, and Zhu 2016) is discussed an important research effort based on optimization techniques; whereas a more recent discussion is provided in (Murthy 2016; Ozkaya and Kloukinas 2013), where the authors highlight possible perspectives and argue the exploitation of the DbC as

method to increase the dependability, in terms of correctness and robustness, of a software by support different phases in its development process.

2.2. Main Objectives

As explained, the DbC is a very powerful and wide adopted method in Software Engineering. Indeed, all the approaches, languages and methods mentioned in the previous Section are meant to be employed for supporting the development process and verification of software.

However, particular interest is shown nowadays for using the DbC also in CPS domain due to its own intrinsic nature (Sreram, Buonopane, Srinivasan, Subathra, and Ayyagari 2015; Derler, Lee, and Törngren 2013). Indeed, according to NIST perspective (NIST 2017), Cyber-Physical Systems are co-engineered interacting networks of physical and computational components (such as Smart Grids, Internet of Things, Smart Cities etc.). These systems represent the basic of critical infrastructures, for building smart services, and improve the quality of the society. Furthermore, most of their part are physical components (e.g. Electrical, Mechanical and Hydraulic components), whose dynamics in terms of inputs, outputs and evolution is well-know. This means that:

- their behavior can be represented without ambiguity through algorithms, functions and mathematical expressions;
- their output values, in terms of data and physical flows, can be combined and evaluated quantitatively.

As a consequence, these characteristics can be employed to define formal and computable *contracts*, on which the parties can agree, before the actual realization of the system.

In fact, a preliminary evaluation and approval of its functionalities through virtual tool, represents an important step not only to identify gap in design but also to distribute "responsibility" as well as to clarify the objectives by reducing ambiguities due to the natural language and different background among the involved parties (stakeholders, designers and developers). Specifically, the use of simulation allows to support domain experts and engineers during the design of the CPS, as a testing tool, and for the evaluation of design choices in terms of constraints violation, according to the stakeholders' expectation, agreements and regulations. The output of the overall process represents the *contract* between stakeholders and system engineers, that is the basis on which the actual system has to be build. To this aim, next Section illustrates how to exploit Properties Modeling combined with Simulation as enabling approach.

3. ENABLING DESIGN BY CONTRACT THROUGH PROPERTIES MODELING

This Section describes how to extend the concept of DbC as method to support the development process of cyber-physical systems. In particular, it wants to be

used as a preliminary assessment tool, based on the definition, evaluation and virtual validation of requirements, by defining and suitably calibrating system parameters before the actual implementation of the system. In fact, in this context, a model-based and simulation-driven verification approach can be adopted not only to model requirements in a formal way but also to compute them in order to discover emergent system behavior that is not typically identifiable through the classic requirements analysis.

3.1. Combining the Properties Modeling approach and Simulation Techniques

The Design by Contract method for CPS is based on the combination of the Properties Modeling (PM) approach and Simulations Techniques (ST). Specifically, the aim of the PM approach is not only to allow in a more formal way the requirement's representation, but also to enable their monitoring. From the other side ST and related tools enable not only to run and emulate the system under consideration but also to trace every requirement and to be notified *where*, *how* and *when* one of them (requirements) is violated.

The Design by Contract of a CPS can be defined as $DbC(CPS)=<C, Rc, S, P, Rp, Rpc>$, where:

- C represents a set of Components*;
- Rc is a set or Relationships among the Components C.
- By using C and Rc a System Design model D of the Cyber Physical System can be defined.
- S represents a set verification Scenario, that is, the flow of actions that can be triggered manually or timed to be carried out, in order to stress and stimulate the system;
- P is a set of (System) Properties that are used to validate the design D against the scenario S;
- Rp is a set or Relationships among the Properties P.
- Rcp is a set or Relationships among the Components C and Properties P.

As a consequence, the *contract* on a *CPS* is accepted by the parties when $DbC(CPS)=true$, that is, when none of the P properties is violated against the scenario S.
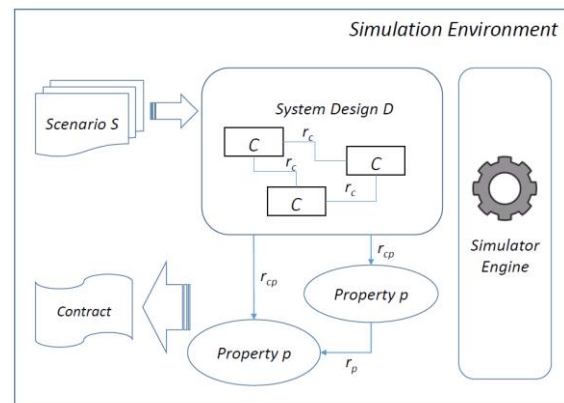
Figure 1: Simulation Model

Figure 1 shows the proposed *Simulation Model (SM)* where the involved *C components, P properties, Rc, Rp and Rpc relationships* and *S scenarios* are used for modeling the overall system. This is, in turn, wrapped into a *Simulation Environment* for enabling the simulation-based design validation of the system by properties evaluation against inputs defined in terms of verification scenarios.

In particular, beside the *System Design* model, system requirements are formalized as computable components in terms of *System Properties* that can be verified. It is worth noticing that the *properties* are fed only with the values coming from the *System Design* model, without interfere with the behavior of the system; whereas the *Scenario* is exploited as input to the system and a consequence as a verification scenario to be checked. The output of the simulation represents the *contract*, that is generated from the automatic execution and evaluation of the *system properties*, through the virtual environment. The next Section describes how the process takes place step by step.

3.2. Reference Methodology

In this Section, a proposed methodology for supporting the Design by Contract method in cyber-physical systems domain is described. It is based on the involvement of different types of Actors, as represented in Figure 2 and described in Table 1, who are exploited throughout the overall process: *Stakeholder, Property Designer (Designer), System Designer (Designer)* and *System Developer*.
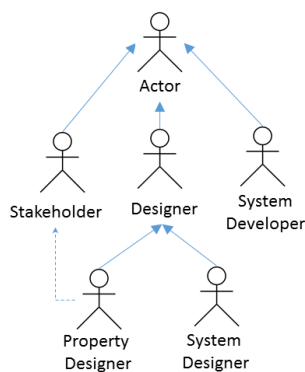
Figure 2: Hierarchy of the involved actors

The proposed methodology is centered on a 6-phases process: *Requirement Specification, Properties Modeling, System Modeling, Virtual Design Modeling, Results Assessment, System Realization and Deployment*, as shown in Figure 3.

Specifically, in the *Requirement Specification* phase, the *Stakeholders* provide all the necessary details and describe textually the system according to their needs and expectations, in terms of functionalities, performances and expected results. The output produced at the end of this phase are two work-products: textual *System Requirements* and the *System Conditions & Constrains*.

Table 1: Involved Actors and related description

Involved Actor	Description
Stakeholder	Domain expert who has interest in building the system under consideration. Typically, with limited technical knowledge. Stakeholder express they requirements and constrains in their own language. Different stakeholders can express conflicting requirements.
Property Designer (Designer)	It is a Designer. He has both: (i) enough expertise in the stakeholder domain, in the context of the system under development; and (ii) technical competences about the system development environ-ment. He is the responsible to translate and provide a formal definition of requirements from the textual version provided by the stakeholder to computable component expressed in terms of System Properties.
System Designer (Designer)	It is a Designer. He is an actor with high level multi-disciplinary competences (e.g. electrical, informatics, hydraulic, etc.). He is in charge to define and design the overall system in terms of components their relationships, functions and input/output values.
System Developer	He is the technical expert with low level of knowledge about the domain. Based on the system design he is in charge the development and/or the integration of system components and system property in the simulation environment, in order to analyze the system behavior and get results from the Simulation.

Starting from such outputs, two phases can get started and proceed in parallel. In particular, the *Properties Modeling* phase takes in input only the *System Requirements*. It is performed by the *Property Designers*, who have similar competencies of the *Stakeholders* and who are able to interpreter correctly the *System Requirements* produced by the *Stakeholders*. In this phase the *System Requirements* are transformed into formalized and computable components (*Properties*) by using mathematical notation, algorithms or functions. The work-products produced in this phase are represented by a set of *System Properties* that able to (i) read and elaborate values originated from the system under consideration and (ii) provide in output a quantitative evaluation according to the requirements.

In parallel can take place the *System Modeling* phase, which is performed from different actors called *System Designers*, who, by using both *System Requirements* and *System Conditions & Constrains*, are able to (i) define the *System Design* in terms of system components, internal behavior and functionalities provided by each component as well as the interaction among them in order to build the overall system; (ii) derive the verification *Scenario* to be used as input for the system.

It is worth noticing that these two phases are completely decoupled and they do not have to interfere to each other, as well as the *Properties Designers* and the *System Designers* are not supposed to communicate and influence each other.

Starting from the outputs produced from the previous phases, during the *Virtual Design Modeling* phase, technical actors called *System Developers* are in charge to build virtually the system simulation environment. In particular, in this phase the *System Design* and the *Properties* are mapped in order to enable their validation against the defined *Scenario*. The output of this phase is represented by an integrated design model. By executing the simulation, different results are automatically generated against the input *Scenario*.

Such results are used and analyzed in the *Results Assessment* phase by the *Stakeholders*, who can approve them completely or partially by sending feedback backward to the *System Designers* in the *System Modeling* phase, based on the output of the *Properties* that codify the requirements. Here, the *System Designer* can use such feedback to improve the *System Design* in order to meet the requirements encoded into the *Properties* that will be again evaluated. The process iterates as long as all the *Properties* are not violated. This means that all the requirements specified by the *Stakeholders* are fulfilled and as a consequence the *System Design* is validated. Once the *Result Assessment* phase ends the *contract* takes place by relying on the fulfillment of the *Properties,* then the last phase can start.

At this point, the *System Realization* and *Deployment* phase of the real CPS can be done. In this phase real system is built, according to the specification clearly identified, and specified preliminary in terms of *Properties* and whose expected outputs have been already analyzed through the Simulation.
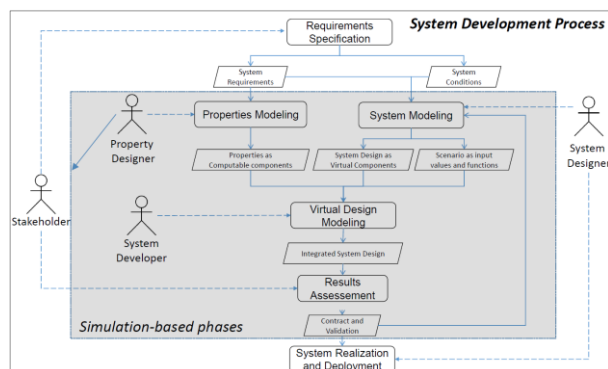
Figure 3: Reference Process and Work-products

In the next Section the proposal is exemplified through a case study in Smart Grid.

4. EXPERIMENTING THE DESIGN BY CONTRACT IN SMART GRID DOMAIN

In this Section, a case study in Smart Grid environment, that illustrates the exploitation of the Design by Contract approach, driven by Simulation and based on Properties Modeling, is presented. In particular, after a brief overview on Smart Grids, the Property Modelling approach is adopted for representing computable requirements; whereas, a Smart Grid Simulator, that implements the properties model, has been ad-hoc

developed for supporting their validation through simulation.

4.1. Smart Grid description and Main Objectives
The CPS under consideration consists of a Smart Grid (SG) (Karnouskos 2011), a modern electric power grid infrastructure. It relies on smooth integration of renewable and alternative energy sources through automated control and modern IT technologies for improving its management. A SG integrates both a cyber-part that encompasses computing and networking resources, and a physical part consisting of physical processes such as mechanical and electrical.

Some of the SG components are *Energy Providers (EPs)*, such as Power Plants, Wind Turbines, Solar Panels, whereas other are *Energy Consumers (ECs)* such as Houses, Cars, Hospitals and so on. In particular, *EPs* are responsible to provide the required amount of energy, according to specific agreements, in order to satisfy the needs of the *ECs*, as efficient as possible.

Furthermore, a SG has to be able to (i) find an optimal balance of energy production for a dynamic demand, (ii) collect data from devices within the grid to manage and discover information, (iii) organize either small micro grids or continental-scale grids, and (iv) integrate heterogeneous devices ranging from big transformers and power plants to smart household appliances (Seo, Lee, and Perrig 2011).

4.2. Requirements Specification
This case study has been defined by cooperating with domain experts in the context of a German research project called PolyEnergyNet (PEN 2017), whose main objective is to support the increase of resilience in Smart Grid environments.

It is worth noting that, there are different types of requirements. Some of them are perceived by the end user in terms of provided services and functionalities; whereas, other requirements are more transparent to users to whom the services are provided.

Nevertheless, they are essential to guarantee a certain level quality and performance. In this phase, the specifications provided by the *Stakeholders* are analyzed and the textual *Requirements* are extracted. Table 2 reports 9 identified requirements, some of which comply the National Electrical Manufacturers Association (NEMA 2017) specifications.

According to the proposed method, some of the above mentioned requirements, are used by the *System Designers* in order to define the *System Design* in terms of Smart Grid structure (e.g. component involved in the design, connections, input and output) and behavior (e.g. actions, process, and events intra- and inter-components) as well as the verification *Scenario*, as described in Section 4.3. Other requirements, instead, are exploited by the *Property Designers* who are in charge to identify and define SG *System Properties*.

Table 2: Smart Grid System Requirements

Requirement	Description
Requirement_01	During operation, the SG has to ensure the functioning of a minimum number of energy producers (Solar Panels, Wind Turbines, etc.).
Requirement_02	The SG has to ensure, for each typology of involved component, that the sum of their output (i.e. the sum of the energy produced) do not drop below a certain threshold.
Requirement_03	When an energy producer is turned on (or turned off), it has to reach its maximum production level (or stop producing) within a certain time (otherwise there could be the risk of burn some SG component and, in general, affect the operation of the overall system).
Requirement_04	The output produced by each energy producer can fluctuate within specific and limited boundaries. For example, the temperature t or the pressure p must fluctuate/oscillate within a certain range.
Requirement_05	After an energy producer is activated in the grid, its operation (and as a consequence its service) has to be guaranteed for at least a specific period of time (e.g. hours, days, years).
Requirement_06	In case of the disservice of the SG, the time of undersupply of each energy consumer must not exceed a certain time interval.
Requirement_07	The total duration of undersupply for each specific consumer, must not exceed a certain limit of annual hours.
Requirement_08	The Smart Grid has to include at least one Power Plant and one House.
Requirement_09	The Smart Grid could include Solar Panels, Wind Turbines, Hospital, and Vehicles.

4.3. Design Modeling

As mentioned in the previous Section, in this phase, the *System Designers* extract the requirements that they need for defining both the structure of the SG as well as for modeling its functional- and non-functional behavior in terms of (i) type of components, (ii) components cardinality, (iii) connections and interactions (iv) constrains such as level of priority of being supplied with energy, and so on.
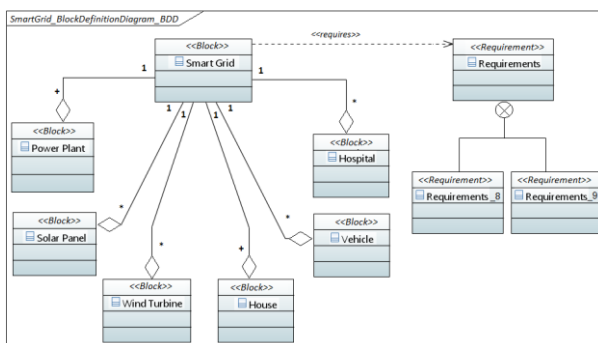
Figure 4: Block Definition Diagram of the main Smart Grid components

As an example, Figure 4 shows a possible high level structural of a Smart Grid, derived from *Requirement_8* and *Requirement_9* that provides information about the involved type Energy Producers and Energy Consumers and their cardinality. The SG components are

represented through SysML Block by using a SysML Bock Definition Diagram (BDD) (OMG 2017) in the left side, whereas the related requirements are reported in the right side.

It is worth notice that, typically, requirements are defined textually without verifiable relationships. Indeed, it is not trivial to establish verifiable relationships in order to proof the fulfillment of textual requirements on the basis of the behavior of system components.

Moreover, specific verification *Scenario* are also identified in this phase. As described before, a *Scenario* model catches a specific sequence of actions that are used to stimulate the *System Design* model in order to induce a particular reaction. So each verification scenario is defined, based on requirements, with the purpose to test the *System Design*. In the following, some of the *Scenarios*, defined starting from the *Requirements Specifications* are reported:

- *Scenario_1*: decreasing or increasing of the level of energy of a Power Plant under or over the allowed thresholds;

- *Scenario_2*: switching-off time a Power Plant. To verify the time that a Power Plant takes to move from the Max level of energy production to 0;

- *Scenario_3*: switching-on time of a Power Plant. To verify the time that a Power Plant takes to move from 0 to the Max Level of energy production.

4.4. Properties Modeling

According to the proposed methodology, this phase aims to define the P*roperties*. Specifically, the *Property Designers* use specific requirements, or extract specific information from them, to define a middle layer of computable *properties*, that allows to bind the *System Design* with the *Requirements*. A graphical representation of the binding between the design of the Smart Grid under consideration and the *Requirements* through *System Properties* is shown in Figure 5.
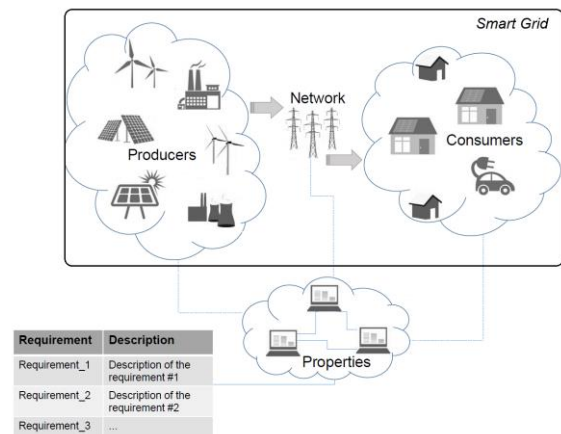
Figure 5: Binding Smart Grid with Requirements through System Properties

Example of *properties*, identified starting from the *requirements* specified in Section 4.2 and defined by using a mathematical notation, are the following:

• *Property_1 (derived from Requirement_1)*: It is related to the *Components Cardinality* of a specific type of Component C. In particular, given a set of *n* components of type C $\{c_1,...,c_m,...,c_i,...,c_n\}$. Given the function $W(c_i)=1$ if the component c_i is working and $W(c_i)=0$ if c_i is not working. Then the property states that at least *m* out *n* components must be working as specified by the equation (1).

$$\sum W(c_i) \geq m \qquad i=1...n \qquad (1)$$

• *Property_2 (derived from Requirement_2)*: It is related to the *Cumulative Value* CV_C provided by a specific set of Energy Producer component of type C. It states that given a set of *n* producer components of type C $\{c_1, c_2,...,c_i,...,c_n\}$ and given the function $P(c_i, t)$ that provides information on the current amount of energy provieded from c_i at the time *t*, then the total amount of the energy produced from the components CV_C has to be, at any time, at least more than the minimum value P_{min} specified according to the equation (2).

$$CV_C = P(C) = \sum P(c_i, t) \geq P_{min} \qquad i=1...n, \forall\, t \in \mathrm{T} \qquad (2)$$

• *Property_3 (derived from Requirement_3)*: It is related to the *Variation Time* of energy production of Energy Producer component.
Given an initial value of energy $P_{Initial}$, that is roduced from a component *c* at time t_0, $P(c,t_0)= P_{Initial}$. Then the final value of energy to be produced from c, at time t_f, $P(c,t_f)= P_{Final}$ has to be reached within a specific time Δt, $|\,t_f - t_0| \leq \Delta t$. As a consequence, the variation of energy $\Delta P(c)$ produced by a component c must be reached within Δt according to equation (3).

$$|P(c,t_0)- P(c,t_f)|_= \Delta P(c), \qquad \text{with } |t_f - t_0| \leq \Delta t \qquad (3)$$

• *Property_4 (derived from Requirement_4)*: It is related to *Threshold Value*. It state that given *k* Energy Producer component C $\{c_1, ...c_j,...,c_k\}$. The value of energy $P(c_j)$ produced in output from each producer c_j has to be kept between an Upper Bound (V_{max}) and a Lower Bound (V_{min}), according to equation (4).

$$V_{min} \leq P(c_j) \leq V_{max} \quad \forall c_j \in C\; j=1...k \qquad (4)$$

• *Property_5 (derived from Requirement_5)*: It is related to the *Service Life Time* of a component. The property state that if a component *c* is ON at time t_0, the service S provided must be hold for at least a *period* of time t_{period}. So no variation of the service has to occur in t_{period} according to equation (5).

$$|S_{ON}(c, t_{i+1}) - S_{ON}(c, t_i)| = 0, \forall\, t_i,\; t_{period} > t_i > t_0 \qquad (5)$$

• *Property_6 (derived from Requirement_6)*: It is related to *Grid Resilience*. If a Energy Consumer component c is undersupplied at time t_0, the system must be able to recover within a time Δt. Given a function $U(c,t_i)$ that is equal to 1 when a consumer *c* is undersupplied at time t_i, and $U(c, t_i)$ is equal to 0 when *c* is supplied. Then the duration of the disservice has to be no longer than Δt according to (6).

$$\forall\, U(c,t_i) = 1 \rightarrow t_{i+1}- t_i \leq \Delta t \qquad \forall t_i \in T\; i= 0..k \qquad (6)$$

Starting from the above produced work-products, defined in terms of models (e.g. *System Design*, *Properties* and *Scenarios*), the *Simulation Environment* can be defined.

4.5. Virtual Design Modeling

In this phase the Smart Grid is virtually represented by the *System Developers*, by modeling first the SG components and connections according to the design provided by the *System Designer*, and then by integrating the *Properties model* provided by the *Properties Designers*. An example of smart grid configuration is shown in Figure 6 through the user interface of the Smart Grid Simulator (SGS) that has been implemented according to the defined models (System Design, Properties and Scenarios).
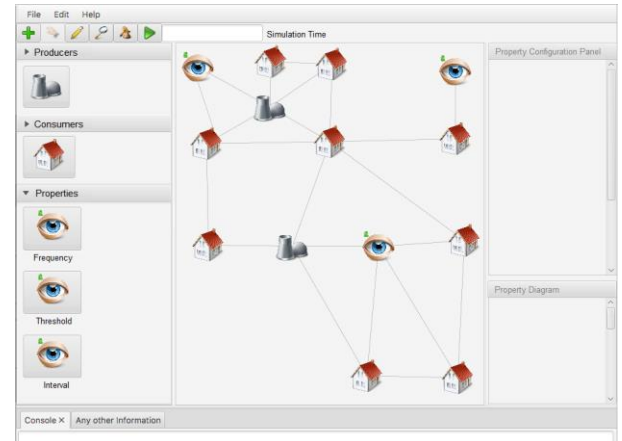
Figure 6: Smart Grid design extended with Properties

In the left side of the Graphic User Interface (GUI), different kind of components to represent SG configuration are provided and grouped as *Producers* (e.g. Power Plant, Solar Panel and Wind Turbine) and *Consumers* (e.g. House, Hospital, and so on). Each component can be configured to send and receive physical data flows as well as digital signals. Their configuration and the way to interact are enabled via data lines to form a network, that determine the actual behavior of the overall grid.

Furthermore, the *Property model* is integrated into the *Simulation Environement* by implementing the *Properties* described in Section 4.4 as additional network components for validating the SG configuration.

As we can see, in the end, the SG *System Design is* extended by introducing verifiable *requirements* defined in terms of *properties* that create the matching between textual requirements and the SG System Design. Once the overall configuration is defined, it is possible the execute the simulation in order to perform different experiments. This allows to evaluate quantitatively the fulfillment of the requirements on the basis of the behavior of the smart grid.

4.6. Results Assessment

One of the experiment has been conducted by validating the SG System Design against the *Scenario_1*, introduced in Section 4.3. To this aim the *Property_4*, called *Threshold Value,* has been employed. By using the zooming-in feature of the SGS simulator, a closer view on each specific property can be provided. As example, Figure 7 shows a *Threshold Value* property defined on two components: a Power Plant as Energy Producer and a House as Energy Consumer.
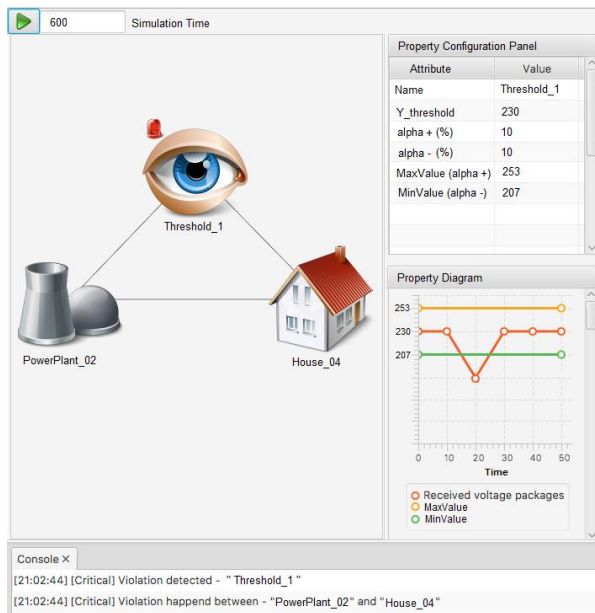
Figure 7: Threshold Value Property

The Property Configuration Panel shows how the *Threshold Property* is configured. In particular, the Y_threshold = 230V represents the desired value of Voltage to be transmitted, whereas the deviation parameters "alpha +" and "alpha -" are configured with the same percentage ± 10%. These initial values generate automatically the MinValue (V_{min}) and MaxValue (V_{max}). The property is fulfilled if, for the overall transmission time, the amount of energy is between MinValue = 207V and MaxValue = 253V. As shown from the Property Diagram in Figure 7, a violation of the property is detected at the simulation time around t=200, because the amount of transmitted Voltage is less than the required MinValue. The property is then highlighted in red color, whereas the components responsible of violation are printed out in the console. This implies the existence of anomaly in the SG System Design and in particular because the

Energy Producer behavior do not respect the required energy that the Energy Consumer needs receive.

This means that the implementation of some functionalities of the SG do not comply the requirements, as a consequence the contract is not respected. As a consequence, as long as exists a property violated, at least a requirement is not fulfilled. This implies that it is necessary to re-iterate the modeling process of the Smart Grid in order to improve its System Design.

Further information is shown in Table 3 which reports the properties used in the verification *Scenario_1* and their related results. In particular, some properties are Not Violated, whereas others are Violated. Moreover, Table 3 shows, also, when a property was Violated by reporting its Violation Timestamps, and duration of each violation (Violation Length).

Table 3: Simulation Results based on Properties as Contracts

Property	Property Status	Violation Timestamps	Violation Length
Threshold_1	*Violated*	*VT_[20]*	*VD_[1]*
Grid Resilience_1	*Not Violated*	*VT_[:]*	*VD_[:]*
Threshold_2	*Violated*	*VT_[3,7,26]*	*VD_[4, 2, 5]*
…	…	…	…

When all properties are not violated that means that all the contracts are fulfilled and the process can move towards the *System Realization and Deployment* phase.

According to the proposed methodology, the inputs of this phase are not anymore represented by only textual requirements, but they are also accompanied by a well-defined model of the Smart Grid that represents how the System Design should look like in reality and how services have to work; whereas the *System Properties* can be reused as monitoring tool that observe the SG during its operation and to compare the deviations among real values against the simulated ones.

So, the specifications formalized in terms of software (Design and Properties) and the simulation results represent quantitative indicators, to be used as *contracts,* for evaluating the level of fulfillment of requirements starting from the early stages of the development process of a CPS.

5. DISCUSSION AND ADVANTAGES

As above described, the proposal extends the traditional system development process by introducing virtual modeling and evaluation phases based on simulation where requirements become computable system entities. This allows to obtain quantitative evaluations of the system design by creating formal matching based on variables as well effective automatic way to trace and verify the fulfillment of requirements using *Properties*. Moreover, thanks to the decoupled level of development between the *System Design* and the *System Properties*

the fulfillment of requirements through properties is not biased.

Furthermore, from the point of view of the actors involved, such approach is beneficial both for the *System Designers* and for the *Stakeholders*. In fact, *System Designers* know concretely what to develop. They know not only the inputs and outputs of the system, but also their relationships and how the inputs need to be combined and processed in order to obtain the outputs according to the *Stakeholder* expectations; *System Designers* also have a way to prove that the real system behaves as the virtual one, on which the agreement (the *contract*) was established. On the other hand, the approach is also beneficial for the *Stakeholders* who have the certainty that the requirements have been correctly understood. In addition, this allows *Stakeholders* to claim the development of capabilities without ambiguity, but rather with "evidence", in case of the real system behaves differently from the virtual one thanks to the virtual model and related gathered results, on which the agreement (the *contract*) was established.

Finally, once the system is realized and deployed, properties may be used, as a control tool, for observing its real behavior during its operation in order to detect anomalies and generate notifications and alerts.

6. CONCLUSION

The paper has proposed a Systems Engineering approach for enabling the Design by Contract method in cyber-physical systems domain. It is focused on the combination of a Properties Modeling approach and Simulation techniques, in which the former is used for enabling in a more formal way the representation of system requirements and constrains, whereas the latter is exploited as tool to objectively check their correctness.

The approach is based on a methodology centered on 6 phases (*Requirement Specification, Properties Modeling, System Modeling, Virtual Design Modeling, Results Assessment, System Realization and Deployment*). It aims at supporting the process for the definition of *contracts* by formalizing the requirements in terms of *System Properties* which in turn are exploited to validate the System Design before actual realization of a system.

A first experimentation has been conducted on a case study in the field of Smart Grid. In particular, a Smart Grid Simulator has been developed to enable the automatic verification of system requirements. The simulator implements not only the main SG components and their related behavior but also it allows to include computable requirements defined as system properties among the operational constraints as well as automatically evaluate them against specific verification scenarios.

Based on the work-products(models) gathered from the overall process and from the simulation results, the approach allows *Stakeholders* and *Engineers* to obtain a preliminary evaluation and agree on specific *contracts* that rely on the validation of the requirements driven by simulation.

Ongoing works are devoted to (i) enrich the approach by defining and integrating a specific Attack Scenario Models based on a Faults and Failures Generator Model able to support the injection of external attacks and faults in the Smart Grid, and (ii) experiment the approach in other application domains.

REFERENCES
INCOSE Book, 2015. Systems Engineering Handbook A Guide for Systems Life Cycle Processes and Activities, 4th Edition, Wiley.

Jazequel J. M. and Meyer B., 1997. Design by contract: the lessons of Ari-ane, in Computer, vol. 30(1), January, pp. 129-130.

Meyer, 1992. Applying design by contract, in Computer, vol. 25(10), pp. 40-51.

Garro A. and Tundis A., 2015. Modeling of system properties: Research challenges and promising solutions, Proceedings of the first IEEE International Symposium on Systems Engineering (ISSE), Rome, Italy, pp. 324-331.

Zimmerman D. M. and Kiniry J. R., 2009. "A Verification-Centric Software Development Process for Java," Proceedings of the Ninth International Conference on Quality Software, pp. 76-85.

Leavens T., Baker L., and Clyde Ruby. 2006. Preliminary design of JML: a behavioral interface specification language for java. SIGSOFT Software. Engineering. 1-38.

NIST (National Institute of Standards and Technology) Cyber Physical Systems, 2017 - available online at https://www.nist.gov/el/cyber-physical-systems.

Leavens G. and Y. Cheon, 2003. Design by contract with JML.

Cimatti A. and Tonetta S., 2016. A Temporal Logics Approach to Contract-Based Design. In: Architecture-Centric Virtual Integration (ACVI), Venice, pp. 1-3.

Cimatti A., Dorigatti M., and Tonetta S., 2013. OCRA: A Tool for Checking the Refinement of Temporal Contracts, in ASE, pp. 702–705.

Nuzzo P., Sangiovanni-Vincentelli A., Bresolin D., Geretti L., and Villa T., 2015. A Platform-Based Design Methodology With Contracts and Related Tools for the Design of Cyber-Physical Systems, Proceedings of the IEEE, vol. 103(11), pp. 2104–2132.

Heckel R. and Lohmann M., 2005. Towards Contract-based Testing of Web Services, In Electronic Notes in Theoretical Computer Science, vol. 116, 19, pp. 145-156.

Murthy PVR., 2016. Reliability by Construction using Design by Contract Methodology. In Proceedings of the 9th ACM India Software Engineering Conference (ISEC '16), New York, USA.

NEMA (National Electrical Manufacturers Association) 2017. [Online]. Available: https://www.nema.org/.

PEN (PolyEnergyNet) 2017. [Online]. Available: http://www.polyenergynet.de/.

OMG (System Modeling Language), 2017. [Online]. Available: http://www.omg.org/spec/SysML/.

Otter M., Thuy N., Bouskela D., Buffoni L., Elmqvist H., Fritzson P., Garro A., Jardin A., Olsson H., Payelleville M., Schamai W., Thomas E., and Tundis A., 2015. Formal Requirements Modeling for Simulation-Based Verification, Proceedings of the 11th International Modelica Conference, Versailles, France.

Nguyen T., 2014. FORM-L: A MODELICA Extension for Properties Modelling Illustrated on a Practical Example, Proceedings of the 10th International Modelica Conference, Lund, Sweden.

MODRIO (ITEA 3 Model Driven Physical Systems Operation) Project, 2016. Available: https://itea3.org/project/modrio.html.

Rubio-Medrano C. E., Ahn G., and Sohr K, 2013. Verifying Access Control Properties with Design by Contract: Framework and Lessons Learned, Proceedings of the IEEE 37th Annual Computer Software and Applications Conference, Kyoto,Japan, July 22-26.

Ozkaya M. and Kloukinas C., 2013. Towards Design-by-Contract Based Software Architecture Design, Proceedings of the 12th IEEE International Conference on Intelligent Software Methodologies, Tools and Techniques. Budapest (Hungary) September 22-24.

Lahtinen J., Johansson M., Ranta J., Harju H. and Nevalainen R., 2010. Comparison between IEC 60880 and IEC 61508 for certification purposes in the nuclear domain, Proceedings of the 29th International Conference on Computer Safety, Reliability and Security. Vienna, Austria, September 14-17.

Rierson L., 2013. Developing Safety-Critical Software: A Practical Guide for Aviation Software and DO-178C Compliance.

Sreram B., Buonopane F., Srinivasan S., Subathra B., and Ayyagari R. 2015. Verification of Design Contracts for Cyber-Physical System Design Using Evolutionary Optimization. Proceedings of the IEEE International Conference on Circuit, Power and Computing Technologies.

Seshia S.A., Hu S., Li W., Zhu Q., 2016. Design Automation of Cyber-Physical Systems: Challenges, Advances, and Opportunities, Transaction on Computer-Aided Design of Integrated Circuits and Systems.

Seo D., Lee H., and Perrig A., 2011. Secure and Efficient Capability-Based Power Management in the Smart Grid, Proceedings in the Parallel and Distributed Processing with Applications Workshops (ISPAW), pp. 119–126.

Karnouskos S., 2011. Cyber-physical systems in the SmartGrid, Proceedings of the IEEE International Conference on Industrial Informatics (INDIN), pp. 20–23.

Danda I.S., Rawat B. and Rodrigues Joel J.P.C., 2015. Cyber-Physical Systems: From Theory to Practice.

Derler P., Lee E.A., and Törngren M., 2013. Cyber-Physical System Design Contracts, Proceedings of the International Conference of Cyber Physical System, Philadelphia, PA, USA, April 8-11.

INCOSE (International Council on Systems Engineering), 2017 - [Online]. Available: http://www.incose.org/.

Klaeren H., Pulvermüller E., Rashid A. and Speck A. 2001. Aspect Composition Applying the Design by Contract Principle, In: Butler G., Jarzabek S. (eds) Generative and Component-Based Software Engineering. LNCS, vol. 2177. Springer, Berlin, Heidelberg.

Furfaro A., Garro A., and Tundis, A. 2014. Towards Security as a Service (SecaaS): On the modeling of Security Services for Cloud Computing. Proceedings of the 48th Annual IEEE International Carnahan Conference on Security Technology (ICCST 2014), Rome, Italy, October 13 – 16.

Furfaro A., Gallo T., Garro A., Sacca D., and Tundis, A., 2016. Requirements specification of a cloud service for Cyber Security compliance analysis. Proceedings of the International Conference on Cloud Computing Technologies and Applications, (CloudTech 2016) pp. 205-212. doi: 10.1109/CloudTech.2016.7847700.

Falcone A., Garro A., Tundis A., 2014. Modeling and simulation for the performance evaluation of the on-board communication system of a metro train. Proc. Of the 13th International Conference on Modeling and Applied Simulation, (MAS 2014), Bordeaux, France, September 10 – 12, pp. 20-29.

MULTI-CLASS MULTI-SERVER QUEUEING NETWORKS FOR PRODUCTION SYSTEMS DESIGN

Pasquale Legato[a], Rina Mary Mazza[b]

[a],[b]Department of Informatics, Modelling, Electronics and Systems Engineering
University of Calabria, 87036 Rende (CS), ITALY

[a]legato@dimes.unical.it, [b]rmazza@dimes.unical.it

ABSTRACT

Product form queueing networks with multiple customer classes and multiple server stations arise in the design and performance evaluation of stochastic job-shop models of manufacturing, warehousing and logistic systems. Real-size models of this type cannot be solved by the classical Mean Value Analysis (MVA) algorithm, due to its exponential computational complexity in the number of classes. Consolidated (pseudo) polynomial approximation methods have been proposed in literature since some decades. They are based on the transformation of the recursive MVA equations in a system of nonlinear equations to be solved iteratively. Unfortunately these contributions do not cover the case of stations with multiple servers. A new technique based on the idea of class aggregation to cope with the latter case, under a first-come-first-served policy is presented. Preliminary numerical experiments are encouraging upon comparison against the exact MVA algorithm.

Keywords: production systems, decision making, queueing networks, mean value analysis

1. INTRODUCTION

Queueing networks are well consolidated as performance models of systems where congestion phenomena need to be quantified since the earlier stage of system design. This allows to address any successive choice of design by focusing on a restricted number of alternatives. Most consolidated examples of the above systems are assembly, machining, warehousing and other modern flexible manufacturing systems (FMS) in industrial engineering, (Buzacott and Shantikumar 1993), (Zijm, Adan, Buitenhek, and van Houtum 2000), (Fukunari and Malmborg 2009), (McGinnis and Wu 2012). This paper deals with the approximate analytical solution of multi-class queueing network models of the BCMP type (Baskett, Chandy, Muntz, and Palacios 1975) as models of dynamic job-shop systems under stochastic routing among service stations and stochastic duration of services. In this modeling framework, the "processor sharing" service discipline also covered by BCMP networks is not useful and customers (jobs) belonging to different classes are usually serviced according to a first-come-first-served (FCFS) rule. Moreover, service stations are typically equipped with multiple identical servers (Stecke and Kim 1989) (i.e. manufacturing machines and/or automated guided vehicles) each having a fixed-service rate (i.e. not load-dependent). Whatever be the number of servers at each station, to apply the BCMP modeling framework under a FCFS discipline, one has to assume that the average service time is the same for all the job classes and it is referred to an exponential probability distribution function. In the multi-class modeling framework, the exponential assumption for service times over all classes, at any given station, corresponds to a mixture of many short service durations for some classes of jobs with a few long service durations for other job classes. A sample of such a mixture with an average duration equal to one time unit is shown in Figure 1. The exponential assumption is considered suitably "safe" for a first-order estimation of queueing phenomena in automated manufacturing systems where no interruptions during service are expected (Hopp and Spearman 2000), (Manitz 2015). In such a case, a less than exponential variability in service durations is typically estimated (Tempelmeier 2003) and this leads to a (safe) overestimation of the actual queue lengths.
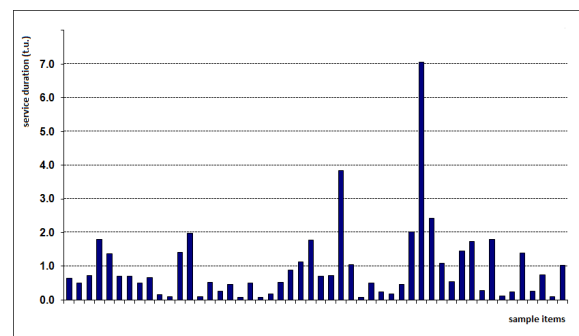
Figure 1: A sample of many short service durations mixed with a few long ones

It is well recognized since a long time that the solution of a BCMP model of a manufacturing system (Solot and Bastos 1988), besides its direct adequateness (Zhuang and Hindi 1990), can also be required as a nested step in approximate iterative methods for solving more

realistic, but non-BCMP models (Jain, Maheshwari, and Baghel 2008). One such pioneering example can be found in (Legato 1993), where a BCMP network (under a FCFS policy) must be solved repeatedly, to estimate the time spent by jobs waiting for admission to the machine-shop section of the whole job-shop model. This reasoning is at the basis of the problem focused in this paper.

In principle, the exact solution of a BCMP model could be obtained by the classical MVA algorithm (Reiser and Lavenberg 1980), but in practice, computational costs become prohibitive for models with more than three or four classes and not restricted to a few customers per class. This occurs because the recursive MVA algorithm is characterized by a dependency between some variables being updated at the current iteration and some others that have been calculated at the previous iteration and whose number rapidly increases with the number of classes. Furthermore, let N_c be the number of class c customers and $\mathbf{N} = (N_1,\ldots, N_c,\ldots,N_C)$ the network population, then the MVA iteration must be repeated from $(0_1,\ldots,0_C)$ to $(N_1,\ldots,N_C)$. Therefore, in a multi-class queueing network the computational complexity of MVA is exponential in both the number of iterations and the number of variables.

For networks with a few service stations and many customer classes, there is an MVA variant proposed ten years after the classic MVA. It is called the Mean Value Analysis by Class (MVAC) algorithm (Conway, de Souza e Silva, and Lavenberg 1989). It exhibits computational costs much lower than MVA, but, as the number of stations increases, MVAC becomes more costly than MVA. On the other hand, since the 1980's, several computer scientists have focused on developing heuristic approximations to the MVA algorithm for BCMP networks aimed at reducing the computational complexity. We have critically examined their work and, in particular, we show in the appendix that one such heuristic algorithm, i.e. the Bard-Schweitzer heuristics (Bard 1979, Schweitzer 1979), may be also derived by a probabilistic argument, i.e. from the assumption that customers circulate within the queueing network according to a semi-Markov process. However, the main contribution of this paper is the proposal of a new approximate, but (pseudo) polynomial MVA algorithm covering the multi-server case under FCFS.

The paper is organized as follows. Background and previous work are provided in section 2. The new approximate algorithm is presented in section 3. Numerical experiments for accuracy validation are given in section 4. Conclusions and issues for further work are in section 5.

2. BACKGROUND AND PREVIOUS WORK

To make the paper self-contained, the MVA algorithm is first resumed.

2.1. Multi-class multi-server MVA algorithm

The following notation is used from now on.

M = number of stations; $j = 1,\ldots,M$ as station index.

m_j = number of servers at station j.

C = number of classes; $c = 1,\ldots,C$ as class index.

N_c = population of class c customers.

n_c = current class c population, from 0 to N_c.

$\mathbf{N}$ = population vector, i.e. $\mathbf{N} = (N_1,\ldots,N_c,\ldots,NC)$.

V_{jc} = expected number of passages through station j by a class c customer in its trip through the network.

R_j = expected service duration at station j.

$\mathbf{n}$ = current population vector, $n = (n_1,\ldots,n_c,\ldots,n_C)$.

$\mathbf{n}$-1_c = $\mathbf{n}$ with one less customer of class c.

$D_{jc}(\mathbf{n})$ = expected sojourn time per visit at station j by a class c customer, under $\mathbf{n}$.

$Q_{jc}(\mathbf{n})$ = expected number of customers at station j, under $\mathbf{n}$.

$T_c(\mathbf{n})$ = expected network throughput for class c customers, under $\mathbf{n}$.

$T_{jc}(\mathbf{n})$ = expected station j throughput for class c customers, under $\mathbf{n}$.

$P_j(l|\mathbf{n})$ = long-run (marginal) probability that l customers are present at station j, under $\mathbf{n}$.

MVA algorithm under FCFS discipline:

For j=1 to M **Do**

  $Q_j(0_1,\ldots,0_C)=0$, $P_j(0_1,\ldots,0_C)=1$

  For l=1 to m_j-1 **Do**

    $P_j(l|0_1,\ldots,0_C)=0$

For $\mathbf{n}=(0_1,\ldots,0_C)$ to $\mathbf{N}=(N_1,\ldots,N_C)$ **Do**

  For c=1 to C **Do**

    For j=1 to M **Do**

(1) $\quad D_{jc}(\mathbf{n}) = \dfrac{R_j}{m_j}\left[1 + \sum_{d=1}^{C} Q_{jd}(\mathbf{n}-1_c) + \sum_{l=1}^{m_j-1}(m_j-l)P_j(l-1|\mathbf{n}-1_c)\right]$

  For c=1 to C **Do**

(2) $\quad T_c(\mathbf{n}) = n_c \,/\, \sum_{j=1}^{M} V_{jc} D_{jc}(\mathbf{n})$

  For j=1 to M **Do**

    For c=1 to C **Do**

(3) $\quad Q_{jc}(\mathbf{n}) = V_{jc} T_c(\mathbf{n}) D_{jc}(\mathbf{n})$

      For l=1 to m_j-1 **Do**

(4) $\quad P_j(l|\mathbf{n}) = (1/l)\cdot \sum_{c=1}^{C} R_{jc} V_{jc} T_c(\mathbf{n}) P_j(l-1|\mathbf{n}-1_c)$

(5) $\quad P_j(0|\mathbf{n}) = 1 - \dfrac{1}{m_j}\left[\sum_{c=1}^{C} R_{jc} V_{jc} T_c(\mathbf{n}) + \sum_{j=1}^{m_j-1}(m_j-l)P_j(l|\mathbf{n})\right]$

It is easy to recognize that both the sojourn time equation (1) and the marginal probability equation (4) are responsible for the exponential computational complexity of MVA. In fact, due to the recursive dependence of both $Q_j(\mathbf{n})$ and $P_j(l|\mathbf{n})$ from the corresponding measures related to the population $\mathbf{n}$ with one class c customer removed, the complexity of the algorithm under just a few servers per station is:

$$O(M \cdot C \cdot \prod_{c=1}^{C}(N_c + 1)), \qquad\qquad (6)$$

both in time and space requirements.

Observe that the marginal probabilities are needed for computing the expected values of the customer sojourn time at each station only in the multi-server case. In the

particular case of just one server per station, the sojourn time (fundamental) equation (1) of the MVA algorithm reduces to the following:

$$D_{jc}(\mathbf{n}) = R_j \left[1 + \sum_{d=1}^{c} Q_{jd}(\mathbf{n} - 1_c) \right] \tag{7}$$

whenever the user is not interested in computing marginal probabilities, but expected performance measures only (e.g. sojourn times, queue lengths and throughputs for each station).

2.2 Heuristic relationships

If we had an approximate relationship between $Q_{jc}(\mathbf{n}-1_c)$ and $Q_{jc}(\mathbf{n})$ and, furthermore, a second relationship between $P_j(l-1|\mathbf{n}-1_c)$ and $P_j(l-1|\mathbf{n})$, then we would be able to circumvent the recursive nature of the MVA algorithm. This recursion requires the "**For n=(0_1,…,0_C) to N=(N_1,…,N_C) Do**" loop which, in turn, is responsible of the exponential complexity in computation. Hence, at the price of introducing an approximation, one may collect equations (1), (2), (3) and (4) into a set of nonlinear equations to be solved simultaneously (Ortega and Rheinboldt 1970) under the (fixed) population $\mathbf{N}$. The previous work that appeared in literature immediately after the earlier publication of the MVA algorithm has been guided by this idea and is resumed in a unifying view. Let us start with the identity:

$$\frac{Q_{jc}(\mathbf{n}-1_c)}{n_c-1} - \frac{Q_{jc}(\mathbf{n})}{n_c} = \frac{Q_{jc}(\mathbf{n}-1_c)}{n_c-1} - \frac{Q_{jc}(\mathbf{n})}{n_c} \tag{8}$$

valid for whatever station-class couple. By introducing a measure (δ) of the change in the fraction of the total number of customers found in station j resulting from the removal of one class c customer out of the (current) network population $\mathbf{n}$:

$$\delta_{jc}(\mathbf{n}) \triangleq \frac{Q_{jc}(\mathbf{n}-1_c)}{n_c-1} - \frac{Q_{jc}(\mathbf{n})}{n_c} \tag{9}$$

a relationship amenable to heuristic particularizations of the δ measure is obtained:

$$Q_{jc}(\mathbf{n}-1_c) = (n_c-1)[(Q_{jc}(\mathbf{n})/n_c) + \delta_{jc}(\mathbf{n})] \tag{10}$$

Clearly, the first option consists in setting:

$$\delta_{jc}(\mathbf{n}) = 0 \tag{11}$$

thus obtaining a proportionality assumption ((n_c-1/n_c) as proportionality factor) on the relationship between $Q_{jc}(\mathbf{n}-1_c)$ and $Q_{jc}(\mathbf{n})$. This is known as the Bard-Schweitzer (BS) heuristic relationship (Bard 1979, Schweitzer 1979). According to these authors, one has to further assume that removing one class c customer does not affect the expected queue length of customers belonging to different classes, hence:

$$Q_{jd}(\mathbf{n}-1_c) = Q_{jd}(\mathbf{n}), \quad d=1,...,C; d \neq c \tag{12}$$

Hence, the BS approximation to equation (7) of the exact MVA algorithm for (fixed rate) single server stations is obtained:

$$D_{jc}(\mathbf{n}) = R_j \left[1 + \frac{n_c-1}{n_c}Q_{jc}(\mathbf{n}) + \sum_{d=1,d\neq c}^{c} Q_{jd}(\mathbf{n}) \right] \tag{13}$$

Provided that marginal probabilities are not required but only expected performance measures are required, formula (13) is all the user needs to define a (fixed-point) iterative heuristic algorithm (called BS_MVA) for queueing networks with fixed-rate single-server stations only. This because the expected system throughput equation (2) can be inserted in the formula for the expected station queue length:

$$Q_{jc}(\mathbf{n}) = V_{jc}T_c(\mathbf{n})D_{jc}(\mathbf{n}) \tag{14}$$

thus obtaining the following equation:

$$Q_{jc}(\mathbf{n}) = n_c \cdot V_{jc}D_{jc}(\mathbf{n}) / \sum_{i=1}^{M} V_{ic}D_{ic}(\mathbf{n}) \tag{15}$$

Equations (13) and (15) are at the basis of a fixed-point iteration algorithm.
The computational complexity of BS_MVA is O(MC) in both space and time (per iteration) requirements. An extensive theoretical study on the existence and uniqueness of the solution of BS_MVA, as well as on convergence, has been carried out in (Pattipati, Kostreva, and Teele 1990). In particular, the existence of the solution is established for monotonic, but single-class networks, i.e. networks where the service rates are monotonically non decreasing functions of the number of customers at the stations. Uniqueness and convergence results have been obtained only under the limiting condition that the number of customers of each class grows to infinity.
An improvement over BS_MVA may be based on the assumption that the change in the fraction of class r customers present at station j resulting from the removal of one class c customer is constant around the current population vector $\mathbf{n}$. This corresponds to replacing the null setting (11) by the constant setting:

$$\delta_{jc}(\mathbf{n}) = \delta_{jc}(\mathbf{n}-1_c) \tag{16}$$

Hence, at the price of introducing a nested fixed-point iteration within the BS_MVA (to estimate the constant δ measure), the so-called Linearizer heuristics, first proposed by Chandy and Neuse (1982), has been obtained. The new assumption on the $\delta_{jc}(\mathbf{n})$ yields an improvement over the BS_MVA algorithm (called Lin_MVA) at the expense of an acceptable increase of the space complexity O(MC^2) and the time (per

iteration) complexity $O(MC^3)$. Some years later, Zahorjan, Eager and Sweillam (1988) proposed working with cumulative (i.e. over all classes rather than per class) queue lengths in formula (10). This allows reducing the space complexity of Lin_MVA from $O(MC^2)$ to $O(MC)$ and time/iteration complexity from $O(MC^3)$ to $O(MC^2)$, without significantly affecting the accuracy of the algorithm. They were not able to establish formal convergence results for the original Lin_MVA algorithm, nor for their own variant to it. Since then, some other variants to both BS_MVA and Lin_MVA and more sophisticated implementations have been proposed (Wang and Sevcick 2000), (Wang, Sevcick, Serazzi, and Wang 2008), but always restricted to fixed-rate services in single-server stations.

To the focus of this paper, (fixed-rate services in multi-server stations) it is worth mentioning that using the following further assumption:

$$P_j(l-1\,|\,\mathbf{N}-1_c) = P_j(l-1\,|\,\mathbf{N}) \qquad (17)$$

within equation (4) allows to define a straightforward version of Lin_MVA capable of solving networks with load-dependent (thus multi-server) service rate stations. Unfortunately, this choice has been exploited by Krzesinski and Greyling (1983), but indicated as the source of several failures for Lin_MVA (i.e. convergence failure, convergence errors and numerical instabilities). Among more successful proposals, it is worth recalling that, before presenting the Linearizer heuristics, Neuse and Chandy (1981) introduced the SCAT algorithm for the approximate solution of multi-class queueing networks with load-dependent service rate stations. SCAT was based on the idea of reconstructing the profile of the distribution of the marginal probability of having one customer at station j from the estimate of the mean queue length at the same station. To this purpose, Chandy and Neuse were assigning the whole marginal probability "mass" to the first two integer values neighboring the estimated (generally fractional) value of the mean queue length. This idea was later refined by Akyildiz and Bolch (1988), who proposed to scatter the assignment of probability mass to a wider range of neighboring integer values. Their scattering was carried out according to a (pseudo) normal distribution function. Akyildiz and Bolch (1988) were able to achieve significant improvements upon the original SCAT in some numerical instances. Unfortunately, other numerical evidence for not heavy-loaded networks states that the profile of marginal queue length probability strongly departs from a normal-like one.

More recently Suri, Sahu and Vernon (2007) have pursued the idea of multiplying $Q_{jc}(\mathbf{n}-1_c)$ by a correction factor within the BS formula (13), in order to cope with the multi-server case. Their factor is aimed to capture the reduction of the expected queue length resulting from the presence of many servers instead of one at the same station. The above factor is defined as an empirical function of both the number of servers and their utilization factor for each station at hand. Unfortunately, this is not appealing for the applications of generic queueing networks in real practice.

3. A NEW APPROXIMATION

Our approach for reducing the computational complexity of the multi-class multi-server MVA from exponential to (pseudo) polynomial when computing both queue lengths and marginal probability distributions in multi-server networks under the FCFS discipline is presented in this section.

Let us introduce $P_j(l\,|\,\mathbf{N})$, $l=1,\dots,N$ as the probability that l customers are present at station j under a single class population of $N=N_1+N_2\dots+N_C$ customers circulating within the queueing network. The idea is that of using $P_j(l-1\,|\,\mathbf{N}-1)$ in place of $P_j(l-1\,|\,\mathbf{N}-1_c)$ $l=1,\dots,N$ to get an approximate sojourn time equation for an MVA algorithm which adopts the BS approximation for queue lengths. To this purpose, the concept of representative class is now introduced.

We associate a single-class network to a multi-class queueing network characterized by a population vector $\mathbf{N}=[N_1,\dots,N_c]$, a matrix of visits $\mathbf{V}=[V_{jc}, j=1,\dots,M; c=1,\dots,C]$ and a vector of service requests $\mathbf{R}=[R_j, j=1,\dots,M]$. The single class, here called "representative class", is defined by the following formulas:

$$N = \sum_{c=1}^{C} N_c \qquad (18a)$$

$$V_j = \sum_{c=1}^{C} V_{jc} T_c(\mathbf{N}) / \sum_{c=1}^{C} T_c(\mathbf{N}) \qquad j=1,\dots,M \qquad (18b)$$

The expected throughput values per class, $T_c(\mathbf{N})$, required by the (aggregation) formulas (18) to define the representative class are computed iteratively by *Large*, a fixed-point procedure based on the BS approximation for queue lengths coupled with the marginal probabilities w.r.t. the representative class:

Procedure: *Large*

Use $P_j(l\,|\,N-1)$ $N=1,\dots,N; j=1,\dots,M$
Initialize $D_{jc}(\mathbf{N}) = R_j$ $c=1,\dots,C; j=1,\dots,M$
Repeat
For c=1 to C **Do**
(19) $T_c(\mathbf{N}) = N_c / \sum_{i=1}^{M} V_{ic} D_{ic}(\mathbf{N})$
For j=1 to M **Do**
(20) $Q_{jc}(\mathbf{N}) = V_{jc} T_c(\mathbf{N}) D_{jc}(\mathbf{N})$
(21) $D_{jc}(\mathbf{N}) = \dfrac{R_j}{m_j}[1 + \dfrac{N_c-1}{N_c}Q_{jc}(\mathbf{N}) + \sum_{d=1,d\neq c}^{C} Q_{jc}(\mathbf{N}) + \sum_{l=1}^{m_j-1}(m_j-l)P_j(l-1\,|\,N-1)]$
Until convergence upon $D_{jc}(\mathbf{N})$, $c=1,\dots,C; j=1,\dots,M$

The representative class and the *Large* procedure are used in the iterative algorithm called *Approx*, presented in the following.

Approx is based on the following reasoning. If we had the exact throughput values $T_c(\mathbf{N})$ c=1,…,C then the expected number of visits to define the representative class would be determined from (18b) once and for all. Instead, since this is not the case and moreover the above throughputs depend in turn from the marginal probabilities of the representative class, we have to resort to a nested fixed-point algorithm where both the expected throughputs and the marginal probabilities are iteratively refined. Thus, *Approx* requires an initial estimate of the marginal probabilities of the representative class at each station and then it iteratively updates the initial estimate until convergence is achieved. *Large* operates as a nested "repeat until" statement. Given the current estimate of the marginal probabilities, *Large* computes the corresponding expected throughput values per class which are needed to update the current parameters defining the representative class.

Procedure: *Approx*

Initialize $P_i(l-1|N-1)$,   l=1,…,m_i-1; j=1,…,M
Repeat
1. Compute class throughput by Large
2. Define the representative class
3. Solve the multi-server network under the representative class, by polynomial single class (polynomial) MVA
4.Update $P_i(l-1|N-1)$,   l=1,…,m_i-1; j=1,…,M
Until *convergence upon* $P_i(l-1|N-1)$,   l=1,…,m_i-1; j=1,…,M
Return *expected performance measures per class and station*

Returned performance measures for each (original) class of customers, e.g. expected throughputs and queue lengths at each station, are computed by the following formulas:

$$T_{jc}(\mathbf{N}) = V_{jc}N_c \Big/ \sum_{i=1}^{M} V_{ic}D_{ic}(\mathbf{N}) \tag{22}$$

$$Q_{jc}(\mathbf{N}) = T_{jc}(\mathbf{N}) \cdot D_{jc}(\mathbf{N}) \tag{23}$$

where $D_{jc}(\mathbf{N})$ values, for j=1,…,M and c=1,…,C, are those returned by *Large* after convergence.

Computational complexity of *Approx* is that of an exact single-class MVA, i.e. O(m_jMN), because step 3. is more costly than step 1. under the BS approximation.

3.1 Convergence

In a large number of numerical experiments we have always observed a very robust behavior. Just a few iterations (outer iterations) are usually needed to determine the marginal probabilities for the representative class at each station. Vice versa, for the nested procedure (*Large*), we have observed that at most some tens of (inner) iterations are needed, in the worst case, to achieve convergence with the initial estimate of the marginal probabilities. But, after the first updates of the marginal probabilities (i.e. the first outer

iterations) the number of inner iterations decreases significantly. Typically, it ranges from almost ten to no more than twenty and it remains constant or decreases further during the next few outer iterations. Nevertheless, in principle, we cannot exclude the existence of some (pathological) cases under heavy-loaded networks in which inner, outer or both convergence conditions are not achieved.

4. NUMERICAL EXAMPLES

In this section *Approx* is applied to the solution of the two queueing network models of the job-shop systems shown in Figure 2 and Figure 3. Each station of the job shop is equipped with a pool of identical facilities (see the numbered circles) bearing a common input buffer whose size is big enough to prevent the occurrence of a full buffer. Services are provided according to a FCFS discipline and their individual duration cumulated over all classes well fits an exponential probability distribution. Multiple types of jobs circulate within the network according to a routing path described by a different Markov chain for each different type of jobs. The resulting network configuration is known as Central Server Model (CSM) to highlight the role of the central station within the network topology.
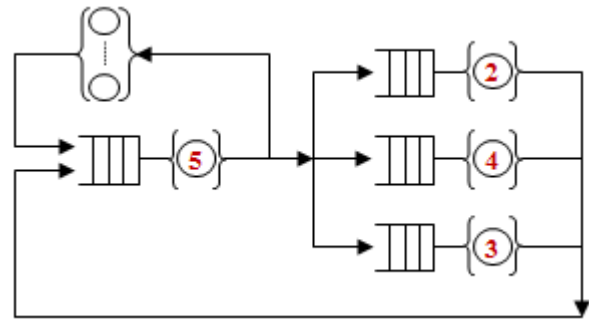
Figure 2: CSM with a pure delay station modelling the delay between finished jobs and new jobs
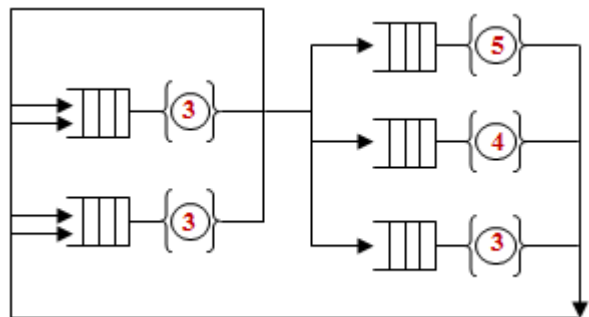
Figure 3: CSM with two central stations and no delay between finished jobs and new jobs

A concrete example of a CSM for an FMS is now given according to the authors' past experience with a German automotive industry in Salzgitter. The material handling system is represented as the central station whose multiple servers are automated guided vehicles (carts) which transfer the workpieces (parts) among the

manufacturing centers. Each manufacturing center is a peripheral station whose servers are the flexible machine tools and the queue is the (local) input storage area. The (local) output storage areas of the manufacturing centers are coalesced into the unique queue of the central (transport) station. A multi-class model must be used when different types of parts have non-compatible geometrical properties and, thus, require different types of support devices (pallets). The parts are mounted on pallets and carried by carts. They follow different routes in the network, according to the class to which they belong. Finally, a fixed number of circulating pallets is assumed and modeled as the fixed multi-class population of customers of the queueing network. The model is closed since no pallet enters the system from outside, nor leaves it; rather, a finished part is removed from the pallet and immediately replaced by a new (raw) part or after some delay. In the latter case, the related average delay is captured by the pure delay station in the CSM (see Figure 2); in the former, the pure delay station is short-circuited.

The numerical results presented in the next two subsections refer to CSM configurations under a few stations and a few classes with a limited number of customers per class. This allows a relatively fast exact solution by the MVA algorithm. Moreover, from the probabilistic argument at the basis of the derivation of our approximation in the Appendix, it should be clear that our experiments under a low-moderate load condition at the central station coupled with a heavy load at one or more peripheral stations should correspond to a worst-case validation.

4.1 Results from the first model

The first set of numerical results presented here is referred to the CSM with a finite source of customers (Figure 1). There are 3 peripheral stations (station 3, station 4 and station 5), 1 central station (station 2) and 3 classes of customers. The population of class 1 is 12, 15 for class 2 and 18 for class 3; therefore, 45 servers are at the pure delay station (station 1), which is visited only once per customer passage through the system. The average delay time differs per class (40 t.u. for class 1, 60 t.u. for class 2 and 80 t.u. for class 3). The central station is equipped with 5 servers, the first peripheral with 2, the second peripheral with 4 and the last peripheral with 3. Using the average number of visits and the average service time per visit shown in Table 1, we tune (by output results) a set of (server) utilization factors ranging from 67% at the first peripheral station (station 3) to 95% at the second peripheral station (station 4). The expected queue lengths per class and the expected throughput per class at each station are reported in Tables 3a-3c. The same performance indices computed by the exact solution of the queueing network returned by the classical MVA algorithm are in these tables as well. The comparison between approximate and exact results for the average queue lengths shows a level of accuracy which is

widely acceptable in practice; this accuracy is even higher for the throughput results.

It is well known that queueing approximations are less accurate in the case of a specific station acting as a strong system bottleneck within an unbalanced and heavily loaded system. So, we show in Tables 4a-4c a second set of results obtained under the condition that the second peripheral station is utilized at 99% while the first is around 83%. The new condition has been obtained by reducing to a half the expected delay per class at the pure delay station. From Tables 4a-4c one may recognize that the only significant loss of accuracy (about 10%) arises in the evaluation of the average queue length at station 4. This is the station that has been forced, by our input data setting, to become a very strong bottleneck (utilization level = 99%) for the entire network. Fortunately, we believe that in this case a more accurate analytical evaluation of the average queue length is unnecessary under a similar system bottleneck.

Table 1: Remaining data of the first model

Class c_i	Service time per station s_i case 1/case 2				
	s_1	s_2	s_3	s_4	s_5
c_1	40/20	0.5	1.0	2.0	1.5
c_2	60/30	0.5	1.0	2.0	1.5
c_3	80/40	0.5	1.0	2.0	1.5
Class c_i	Visits per station				
c_1	1	9	1	3	4
c_2	1	7	3	2	1
c_3	1	5	1	2	1

Table 2: Remaining data of the second model

Station	Visits per class c_i				Service times	
	c_1	c_2	c_3	c_4	case 1	case 2
1	5	7	0	0	0.10	0.35
2	0	0	12	11	0.25	0.25
3	1	3	4	2	0.70	0.50
4	2	1	4	3	0.50	0.30
5	1	2	3	5	0.20	0.20

Table 3a: Class 1 results of the first model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.1982	0.1977	7.9294	7.9105
2	1.7841	1.7798	0.9079	0.9026
3	0.1982	0.1977	0.2575	0.2583
4	0.5947	0.5932	1.4780	1.4910
5	0.7929	0.7910	1.4269	1.4374

Table 3b: Class 2 results of the first model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.2019	0.2016	12.115	12.098
2	1.4134	1.4115	0.7199	0.7197
3	0.6057	0.6049	0.7757	0.7810
4	0.4038	0.4032	1.0161	1.0247
5	0.2019	0.2016	0.3726	0.3759

Table 3c: Class 3 results of the first model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.1983	0.1981	15.868	15.853
2	0.9918	0.9908	0.5055	0.5071
3	0.1983	0.1981	0.2582	0.2596
4	0.3967	0.3963	1.0008	1.0094
5	0.1983	0.1981	0.3663	0.3699

Table 4a: Class 1 results of the first model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.2334	0.2266	4.6688	4.5325
2	2.1009	2.0396	1.1349	1.0640
3	0.2334	0.2266	0.4368	0.4162
4	0.7003	0.6798	3.5357	3.8191
5	0.9337	0.9065	2.2236	2.1680

Table 4b: Class 2 results of the first model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.2886	0.2837	8.6592	8.5112
2	2.0204	1.9859	1.0918	1.0399
3	0.8659	0.8511	1.5433	1.5344
4	0.5772	0.5674	2.9824	3.2166
5	0.2886	0.2837	0.7231	0.6977

Table 4c: Class 3 results of the first model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.3135	0.3088	12.543	12.352
2	1.5679	1.5440	0.8490	0.8123
3	0.3135	0.3088	0.5817	0.5677
4	0.6271	0.6176	3.2411	3.5072
5	0.3135	0.3088	0.7847	0.7601

4.2 Results from the second model

In the second model (Figure 3) the pure delay station has been eliminated (i.e. short-circuited), two central stations (station 1 and station 2) are coupled with three peripheral stations (stations 3-5) and the number of customer classes has been increased from 3 to 4. Class 1 population is 12, class 2 is 10, class 3 is 6 and class 4 is 8. There are 3 servers in each of the two central stations, 4 at the first peripheral, 3 at the second and 2 at the last one. All the customers visit all the peripheral stations, but only class 1 and class 2 customers visit the first central station and, in turn, only class 3 and class 4 customers visit the second central station. In terms of the FMS of reference, this means that some carts are dedicated to some types of parts and other carts to other types of parts.

Using the data in Table 2, a double strong bottleneck condition in a heavily loaded system is obtained. In fact, the first and second peripheral stations result to be utilized both at 99% and the less loaded station (central station 1) is at 72%. The results in Tables 5a-5d confirm the good accuracy achieved in practice by our approximate algorithm. Finally, we eliminate one server

from each of the two central stations and modify service times as shown in Table 2 (case 2 column). This leads to moving the double bottleneck condition from the two peripheral stations to the two central stations (the first central at 99% and the second at 97%, the first peripheral at 83% and the second at 69%). Results are in Tables 6a-6d. As expected, the accuracy does not depend from the position of the bottleneck stations but, more importantly, no further loss of accuracy is observed w.r.t. the previously discussed results from the first model (where only one bottleneck station was forced to exist by our input data tuning).

Table 5a: Class 1 results of the second model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	7.0665	6.9051	0.7665	0.7288
2	0.0000	0.0000	0.0000	0.0000
3	1.4133	1.3810	4.4237	4.2038
4	2.8266	2.7620	6.3748	6.6603
5	1.4133	1.3810	0.4348	0.4069

Table 5b: Class 2 results of the second model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	5.7237	5.4921	0.6218	0.5802
2	0.0000	0.0000	0.0000	0.0000
3	2.4530	2.3537	6.7243	7.0202
4	0.8176	0.7845	2.1540	1.9390
5	1.6353	1.5691	0.4996	0.4605

Table 5c: Class 3 results of the second model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.0000	0.0000	0.0000	0.0000
2	2.7426	2.7246	0.8472	0.8336
3	0.9142	0.9082	2.7259	2.7475
4	0.9142	0.9082	2.2157	2.2180
5	0.6856	0.6811	0.2111	0.2007

Table 5d: Class 4 results of the second model – case 1

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.0000	0.0000	0.0000	0.0000
2	4.8636	4.8405	1.4846	1.4632
3	0.8843	0.8801	2.7220	2.6814
4	1.3264	1.3201	3.1335	3.2170
5	2.2107	2.2002	0.6599	0.6382

Table 6a: Class 1 results of the second model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	3.1715	3.1429	10.474	10.567
2	0.0000	0.0000	0.0000	0.0000
3	0.6343	0.6285	0.3374	0.3306
4	1.2686	1.2571	0.4091	0.4006
5	0.6343	0.6285	0.7787	0.7008

Table 6b: Class 2 results of the second model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	2.5427	2.5287	8.4283	8.5155
2	0.0000	0.0000	0.0000	0.0000
3	1.0897	1.0837	0.5776	0.5661
4	0.3632	0.3612	0.1176	0.1159
5	0.7264	0.7224	0.8764	0.8023

Table 6c: Class 3 results of the second model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.0000	0.0000	0.0000	0.0000
2	3.5358	3.4282	3.9019	4.1091
3	1.1786	1.1427	0.6233	0.5910
4	1.1786	1.1427	0.3799	0.3610
5	0.8839	0.8570	1.0947	0.9387

Table 6d: Class 4 results of the second model – case 2

Station	Throughput		Queue length	
	exact	approx	exact	approx
1	0.0000	0.0000	0.0000	0.0000
2	4.3407	4.2689	4.9401	5.1395
3	0.7892	0.7761	0.4199	0.4060
4	1.1838	1.1642	0.3826	0.3695
5	1.9730	1.9404	2.2572	2.0848

5. CONCLUSIONS

The proposed fixed-point algorithm seems to be a practical alternative to the classical MVA algorithm for large multi-class queueing networks with multi-server stations. It applies under a FCFS discipline for the exponentially distributed service durations. It does not require restrictive assumptions on the probabilistic shape of the marginal queue length at each station, nor complicated tuning of empirical functions. This should encourage using it in real practice.

Preliminary validation experiments against the exact solution provided by the MVA algorithm state a good accuracy on queueing network models of job-shop systems. The complete exact results of our instances highlight that accuracy lies in the slight variability of the customer sojourn time per class, under a FCFS discipline. The different number of visits to the same station by customers belonging to different classes is responsible for the variability of the waiting time per class. The different average number of visits per class produces different sampling patterns according to which the queue length at that station is observed and suffered by arriving customers.

Polynomial complexity per iteration has been obtained by aggregating all the customer classes into a unique representative class. Marginal queue length probabilities are computed for the representative class at each station by using the (polynomial) single-class MVA algorithm. After a disaggregation step, the expected performance measures per customer class are obtained. A formal proof of convergence seems difficult and even unlikely,

considering the past efforts put into this type of iterative schema. Failure in convergence cannot be excluded for pathological instances yet to be classified. Refinements of the algorithm and extensive numerical experiments are also the subject of a future work.

APPENDIX

Here a probabilistic argumentation on the BS heuristics for queue lengths is provided.

Introducing $PF_{jc}(\mathbf{N})$ as the probability of finding one class c customer at station j (i.e. one out of class c population) given that it circulates in a network together with other $N_1, N_2, \ldots N_{c-1}, \ldots N_C$ other customers, we assume that:

$$Q_{jc}(\mathbf{N}) = Q_{jc}(\mathbf{N}-1_c) + PF_{jc}(\mathbf{N}), \quad c=1,\ldots,C; \quad j=1,\ldots,M.$$

(24)

Relationship (24) lies on the hypothesis that adding a unique class c customer to a network population $\mathbf{N}-1_c$ does not produce the effect of redistributing preexisting customers in station j among other stations and, therefore, it does not affect the preexisting average queue lengths. Rather, on the long run we expect a very limited and purely additive marginal contribution to queue lengths. Precisely, this contribution is uniquely determined by the average proportion of the time spent at each queue with respect to the average time spent within the entire network by the (marginal) customer just added.

At this point, we estimate $PF_{jc}(N)$ as follows:

$$PF_{jc}(\mathbf{N}) = V_{jc}D_{jc}(\mathbf{N}) / \sum_{i=1}^{M} V_{ic}D_{ic}(\mathbf{N}), \quad c=1,\ldots,C; j=1,\ldots,M.$$

(25)

The approximation underlying relationship (25) consists in evaluating the average time spent over all visits at station *j* by a class *c* customer as the product of the average number of visits multiplied by the average time spent per visit. This is not exact, unless the customer circulates within the network according to a semi-Markov process (Cinlar 1975) and, as a consequence, the sojourn time per visit by any customer at any given station is independent from the specific visit (e.g. the first, the second…or the last). Only under this independence assumption the numerator of formula (25) corresponds to the exact evaluation of the cumulative (i.e. over all visits) sojourn time spent by a class c customer at station j during its lifetime within the queueing network, while the denominator corresponds to the average sojourn time within the entire network. Otherwise the correlation among the sojourn time per visit arises, thus reducing formula (25) to an approximation formula. Actually, we have verified by simulation experiments that a significant correlation may exist among the sojourn times experienced by a class *c* customer at station *j*. This qualifies our semi-Markov assumption as an approximation, but at the same time, it indicates the way to improve our approximation. To give some details, by simulating the CSM model in Figure 2, we have observed point

estimates of the correlation factor among sojourn times experienced by a specific (sample) customer at the second peripheral station (station 4). They typically range from 0.15 to 0.25 under a very high-loaded central station coupled with a moderate-loaded station 4. On the other hand, the same correlation factor typically ranges from 0.55 to 0.80 under a moderate-loaded central station coupled with a high-loaded station 4.

After this discussion, the BS heuristics is derived here using some algebra.

By our approximate relationship:

$$Q_{jc}(\mathbf{N}-1_c) = Q_{jc}(\mathbf{N}) - V_{jc}D_{jc}(\mathbf{N})/\sum_{i=1}^{M}V_{ic}D_{ic}(\mathbf{N}), \quad \forall(j,c)$$

(26)

recalling the sojourn time equation of MVA for a (fixed rate) single-server station under FCFS:

$$D_{jc}(\mathbf{N}) = R_J[1 + Q_{jc}(\mathbf{N}-1_c) + \sum_{d=1,d\neq c}^{C}Q_{jd}(\mathbf{N}-1_c)], \quad \forall(j,c)$$

(27)

we assume the following:

$$\sum_{d=1,d\neq c}^{C}Q_{jd}(\mathbf{N}-1_c) = \sum_{d=1,d\neq c}^{C}Q_{jd}(\mathbf{N})$$

(28)

Hence, it suffices inserting first (26) in (27):

$$D_{jc}(\mathbf{N}) = R_J\left[1 + Q_{jc}(\mathbf{N}) - \frac{\mathbb{V}_{jc}D_{jc}(\mathbf{N})}{\sum_{i=1}^{M}V_{ic}D_{ic}(\mathbf{N})} + \sum_{d=1,d\neq c}^{C}Q_{jd}(\mathbf{N}-1_c)\right]$$

(29)

and then repeating equation (15), with reference to population $\mathbf{N}$:

$$Q_{jc}(\mathbf{N}) = N_c \cdot V_{jc}D_{jc}(\mathbf{N})/\sum_{i=1}^{M}V_{ic}D_{ic}(\mathbf{N}), \quad \forall(j,c)$$

(30)

in equation (29) to obtain the BS heuristic approximation:

$$D_{jc}(\mathbf{N}) =$$

$$= R_j\left[1 + N_c \cdot \frac{V_{jc}D_{jc}(\mathbf{N})}{\sum_{i=1}^{M}V_{ic}D_{ic}(\mathbf{N})} - \frac{\mathbb{V}_{jc}D_{jc}(\mathbf{N})}{\sum_{i=1}^{M}V_{ic}D_{ic}(\mathbf{N})} + \sum_{d=1,d\neq c}^{C}Q_{jd}(\mathbf{N}-1_c)\right]$$

$$= \left[1 + \frac{N_c-1}{N_c}Q_{jc}(\mathbf{N}) + \sum_{d=1,d\neq c}^{C}Q_{jd}(\mathbf{N}-1_c)\right]$$

(31)

This concludes our alternative derivation of the BS heuristics based on the core assumption that customer travel through any sequence of network stations is described by a semi-Markov process.

REFERENCES

Akyildiz I.F., Bolch G., 1988. Mean value analysis approximation for multiple server queueing networks. Performance Evaluation 8: 77-91.

Bard J., 1979. Some extensions to multi-class queueing network analysis. In: M. Arato, A. Butrimenko, E. Gelenbe, eds. Performance of Computer Systems, Amsterdam: North Holland, 51-62.

Baskett F., Chandy K.M., Muntz R.R., Palacios F., 1975. Open, closed and mixed networks of queues with different classes of customers. Journal of the ACM 22 (2): 248-260.

Buzacott J.A., Shanthikumar G.J., 1993. Stochastic models of manufacturing systems. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Chandy K.M., Neuse D., 1982. Linearizer: a heuristic algorithm for queueing network models of computing systems. Communications of the ACM 25 (2): 126-134.

Cinlar E., 1975. Introduction to stochastic processes. Engl. Cliffs NJ: Prentice-Hall.

Conway A.E., de Souza e Silva E., Lavenberg S.S. 1989. Mean value analysis by chain of product form queueing networks. IEEE Transactions On Computers 38 (3): 432-442.

Fukunari M., Malmborg C. J., 2009. A network queuing approach for evaluation of performance measures in autonomous vehicle storage and retrieval systems. European Journal of Operational Research 193: 152–167.

Hopp W.J., Spearman M.L., 2000. Factory physics—Foundations of manufacturing management (2nd ed.). New York: Irwin/McGraw-Hill.

Jain H., Maheshwari S., Baghel K.P.S., 2008. Queueing network modelling of flexible manufacturing systems using mean value analysis. Applied Mathematical Modelling 32: 700–711.

Krzesinski A., Greyling J., 1984. Improved linearizer methods for queueing networks with queue dependent centres. ACM SIGMETRICS Performance Evaluation Review 12 (3): 41-51.

Manitz M., 2015. Analysis of assembly/disassembly queueing networks with blocking after service and general service times. Annals of Operations Research 226 (1): 417-441.

McGinnis L., Wu K., 2012. Performance evaluation for general queueing networks in manufacturing systems: characterizing the trade-off between queue time and utilization. European Journal of Operational Research 221: 328–339.

Ortega J.M., Rheinboldt W.C. 1970. Iterative solution of nonlinear equations in several variables, New York: Academic Press.

Neuse D., Chandy K.M., 1981. Scat: a heuristic algorithm for queueing network models of computing systems. ACM SIGMETRICS Performance Evaluation Review 10 (3): 59-79.

Pattipati K.R., Kostreva M.M., Teele J.L., 1990. Approximate mean value analysis algorithms for queueing networks: existence, uniqueness and

convergence results. Journal of the ACM 37 (3): 643-673.

Reiser M., Lavenberg S.S., 1980. Mean value analysis of closed multichain queueing networks. Journal of the ACM 27(2): 313-322.

Schweitzer P.J., 1979. Approximate analysis of multi-class closed networks of queues. Proceedings of the International Conference on Stochastic Control and Optimization, 25-29. April 5-6, Amsterdam.

Solot P., Bastos J.M., 1988. MULTIQ: a queueing model for FMSs with several pallet types. Journal of the Operational Research Society, 39(9):811-821.

Stecke K.E., Kim I., 1989. Performance evaluation for systems of pooled machines of unequal sizes: unbalancing versus balancing. European Journal of Operational Research 42: 22-38.

Suri R., Sahu S., Vernon M., 2007. Approximate mean value analysis for closed queuing networks with multiple-server stations. Proceedings of the 2007 Industrial Engineering Research Conference, 1-6. December 2-4, Nashville (Tennessee, USA)

Tempelmeier H., 2003. Practical considerations in the optimization of flow production systems. International Journal of Production Research, 41 (1): 149–170.

Wang H., Sevcick K.C., 2000. Experiments with improved approximate mean value analysis algorithms. Performance Evaluation 39: 189–206.

Wang H., Sevcik K.C., Serazzi G., Wang S., 2008. The general form linearizer algorithms: a new family of approximate mean value analysis algorithms. Performance Evaluation 65: 129–151.

Zahorjan J., Eager J.D., Sweillam H:M., 1988. Accuracy, speed and convergence of approximate mean value analysis. Performance Evaluation 8: 255-270.

Zhuang L., Hindi K.S., 1990. Mean value analysis for multiclass closed queueing network models of FMS with limited buffers. European Journal of Operational Research 46: 366-379.

Zijm W.H.M., Adan I.J.B.F., Buitenhek, R., Houtum, van, G.J., 2000. Capacity analysis of an automated kit transportation system. Annals of Operations Research 93 (1): 423-446.

AUTHORS BIOGRAPHY

PASQUALE LEGATO is an Associate Professor of Operations Research in the Department of Informatics, Modeling, Electronics and System Engineering (DIMES) at the University of Calabria, Rende (CS, Italy). He has been a member of the Executive Board of the University of Calabria as well as university delegate for the supervision of associations and spin-offs from the University of Calabria. He has been involved in several EEC funded research projects aimed at the technological transfer of SO procedures and frameworks in logistics. He is a member of the INFORMS Simulation Society. His research activities focus on queuing network models, stochastic simulation and the integration of simulation techniques with combinatorial optimization algorithms. His e-mail address is legato@dimes.unical.it and his web-page can be found at http://wwwinfo.dimes.unical.it/legato.

RINA MARY MAZZA is the Research Manager of the Department of Informatics, Modeling, Electronics and System Engineering (DIMES) at the University of Calabria, Rende (CS, Italy). She graduated in Management Engineering and received a PhD in Operations Research from the above university. She has been a consultant for operations modeling and simulation in container terminals. Her current research interests include discrete-event simulation and optimum-seeking by simulation in complex systems. Her e-mail address is rmazza@dimes.unical.it.

ENERGY OPTIMIZATION OF A WIND SYSTEM WITH SMALL POWER BY FUZZY MPPT ALGORITHM

Jean Claude Rakotoarisoa[(a)], Sonia Moussa[(b)], Jean N. Razafinjaka[(c)]

[(a),(c)] Electrical Engineering Department, Automation Laboratory, University of Antsiranana, Madagascar.
[(b)]Electronic Systems Laboratory LR11ES15, Engineering School of Tunis, University El Manar, Tunis, Tunisia.

[(a)]rajeanclaude@gmail.com, [(b)]moussainos@gmail.com, [(c)]razafinjaka@yahoo.fr

ABSTRACT

This work is devoted to the energy optimization of a wind power system based on synchronous permanent magnet machine. A strategy MPPT (Maximum Power Point Tracking) with fuzzy logic is used in order to obtain a maximum energetic output at each speed of wind. The control of the generator is based on a vector control with structure PWM. A speed regulation whose reference is defined according to fuzzy algorithm MPPT is added with this structure. A sample of the wind measured on real site (Antsiranana Madagascar), during a given time is used for simulation. The obtained results show the effectiveness of the adopted strategy. Indeed, one could collect 97% of the maximum energy which can be extracted from the profile of the considered wind. The energy gain obtained by the MPPT device is 13% compared to the operation with velocity of constant reference.

Keywords: energy optimization, wind power system, MPPT, fuzzy logic, synchronous permanent magnet machine, vector control.

1. INTRODUCTION

Renewable energies systems or RES, (hydraulic, solar, wind, geothermic…) constitute only 20% of electrical production in the world. Excluding the hydroelectric one, this rate falls to 2% (B. Multon *and al* 2004; C. Alonso 2010).
At the horizon of 2020, the European Economic Community has triple objectives to increase the share of the renewable energies to a total value of 20%, to decrease pollution of 20% and to save 20% of energy. In Madagascar, the use of these RES is now necessary to resolve pollution problems and to bring economic solutions taking into account problems of fuels whose coasts do not cease increasing.
Currently, wind power systems take more and more place in the production of electricity. Parallel to the system with great power, those with small one begin to increase for isolated site. This case is well adapted for more countries like Madagascar. It is the reason which justifies the choice of this technology for the present paper.
More chains of wind power production use synchronous permanent magnet machine. In order to maximize the efficiency of the aero generator, various solutions are examined.
For the wind turbine, a characteristic, the power coefficient according to the reduced speed, Cp (λ), which depends on the parameters of construction, makes it possible to obtain the curve of power. This one requires an adaptation of the mechanical load in order to ensure a good efficiency: it is the maximization of the power or MPPT.
The generator is connected electrically to a static inverter. This converter can have a structure which depends on the strategies of research of the maximum point of power. Two families of these strategies exist for this purpose (Mirecki 2005):

1. An indirect control of the wind power using a chain of conversion containing bridge of diodes and a chopper. This structure requires the knowledge of the Cp (λ)
2. A speed or torque control starting from the structure of rectifier MLI and a mechanical sensor (velocity and position). In this way, various strategies for optimal point of power are proposed, specially the one that the characteristic Cp (λ) is not needed.

In this paper, it is assumed that the characteristic Cp (λ) is unknown. An algorithm, based on fuzzy logic, is used to control the machine in order to extract the maximum of the wind.

2. WIND GENERATOR MODELLING

The wind generator includes the turbine with variable speed which is coupled directly with a synchronous permanent magnet connected to a continuous bus by a converter MLI. The system is represented in figure 1.
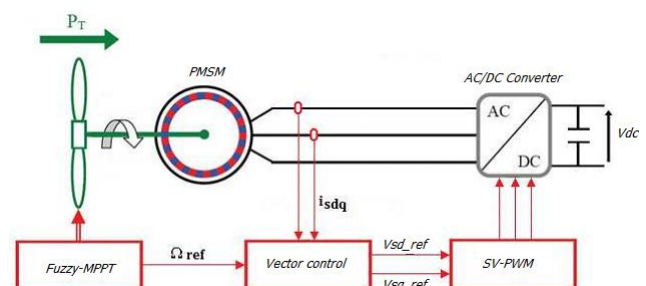
Figure 1: Structure of the wind generator

2.1. Mechanical equations

The differential equation which characterizes the mechanical behaviour of the unit harnesses – generator is given by (Cardenas and Dobson 1996; Gerqud 2002):

$$(J_T + J_m)\frac{d\Omega}{dt} = C_{eol} - C_{em} - (f_m - f_t).\Omega \qquad (1)$$

Where J_T et J_m are respectively the inertia of the harnesses and the generator, f_m friction coefficient of the machine, f_t friction coefficient of the pales and C_{eol} , the static torque given by the wind generator.

The torque given by the wind generator is:

$$C_{eol} = \frac{C_p(\lambda).\rho.R^2.H.V_v^2}{\lambda} \qquad (2)$$

Where ρ is the density of the air, V_v the wind speed, R radius of the aerofoil and H its height.

2.2. Synchrounous machine modelling

The modelling of the PMSM is already the subject of many works. A model in the Park reference is used in this work.

The stator dq-winding Voltages can be expressed as follows (Sturtzer and E. Smichel 2000; Abdessemed and M. Kadjoudj 1997):

$$V_{sd} = R_s.i_{sd} + L_s\frac{di_{sd}}{dt} - L_s.\omega.i_{sq} \qquad (3)$$

$$V_{sq} = R_s.i_{sq} + L_s\frac{di_{sq}}{dt} + L_s.\omega.i_{sd} + K_A.\omega \qquad (4)$$

Where, i_{sd}, i_{sq} are the stator currents, V_{sd} and V_{sq}, the stator voltages, R_s and L_s denote respectively stator and inductance cyclic stator, p and ω are the number of pair of poles and the speed of the equivalent dq-windings (in electrical rad/s) in order to keep the d-axis always aligned with the stator magnetic axis (Kimbark 1995), K_A is a coefficient characterizing the machine(maximum flux megnet).

The speed ω is related to the actual rotor speed Ω as:

$$\omega = p.\Omega \qquad (5)$$

In the normal speed range below the rated speed, the reference for the d-winding current is kept zero (ids = 0). The electromagnetic torque can be expressed as follows:

$$C_{em} = p.K_A.i_{sq} \qquad (6)$$

The simulation block diagram for the PMSM is shown in Figure 2.
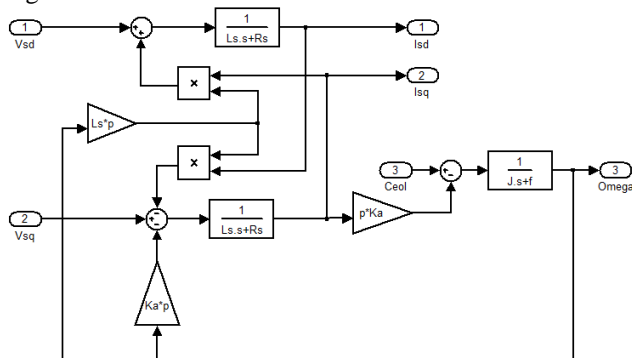
Figure 2: PMSM block diagram

2.3. Model of the rectifier with MLI

For the study of system generator-rectifier MLI, ideals switches are used for the rectifier one models the rectifier.

The function of the switches is to establish a connection between alternate side and continuous bus. The states of these switches are defined like follow (A.S. Toledo 2000; Abdessemed 2011):

$$S = \begin{cases} +1 \Rightarrow \overline{S} = -1 \\ -1 \Rightarrow \overline{S} = +1 \end{cases} \qquad (7)$$

The coupling equation between of the alternative and continuous sides is given by (Belakehal and al 2010):

$$C\frac{dU_{dc}}{dt} = S_a i_a + S_b i_b + S_c i_c - i_L \qquad (8)$$

3. ENERGY OPTIMIZATION OF WIND SYSTEM

3.1. Power maximization method

The characteristic of the optimum capacity of a wind system is strongly nonlinear and has a shape like a bell. For each speed of wind, it is necessary that the system finds the power maximum which is equivalent to the optimal speed revolutions so corresponds to a defined load torque. The diagram of the Figure 2 gives an example of characteristic curves of a wind generator in the power-rotation speed plan.
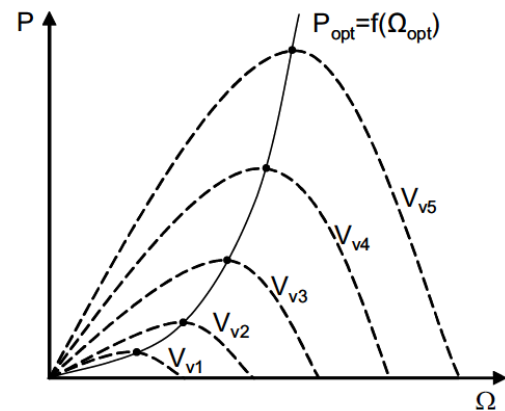
Figure 3: Characteristics in power-speed plan

The whole of the tops of these curves, which are the required optimal points, defines a curve (cubic form) known as of optimum power. Ideally, this curve must be followed constantly during the operation of the wind system.

If the knowledge of these characteristics is missing, rules of behavior to ensure the convergence to the optimal point are relatively simple to establish. These rules depend on the variations of power ΔP and speed $\Delta\Omega$.

Example:

IF $\Delta P > 0$ AND $\Delta\Omega > 0$ THEN $\Delta\Omega ref > 0$

From the existence of these linguistic rules, the use of a MPPT device based on fuzzy set can be applied

3.2. The MPPT with Fuzzy Logic

The MPPT device based on measurement of electrical power (ΔP) in the DC bus and speed rotation ($\Delta\Omega$) variations gives the reference speed $\Delta\Omega_{ref}$ according the equations:

$$\Delta P = P_k - P_{k-1} \qquad (9)$$

$$\Delta\Omega = \Omega_k - \Omega_{k-1} \qquad (10)$$

$$\Omega_{ref} = \Omega_{k-1} + \Delta\Omega_{ref_k} \qquad (11)$$

Figure 3 gives an application of a research in (P- Ω) plane.
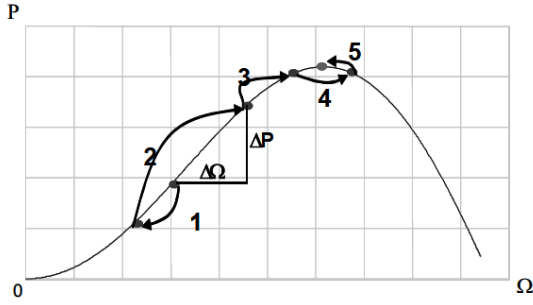
Figure 4: Principle of the MPPT operation at constant wind speed

Example:

IF ($\Delta P > 0$ AND $\Delta\Omega > 0$) THEN $\Delta\Omega ref > 0$

IF ($\Delta P < 0$ AND $\Delta\Omega > 0$) THEN $\Delta\Omega ref < 0$

So, the change of ΔP resulting from the variation of the speed is either in the positive direction or in the negative direction. The value of ΔP can also be small or on the contrary large. From this judgement, the set point speed is increased or decreased in a small or respectively large way in the direction which makes it possible to increase the power. This command allows the research of the optimum point.

If the wind speed is not constant, the research of the maximum point of power is carried out in the way presented on Figure 5: it is noted that the same type of rules like with constant wind speed can be applied.
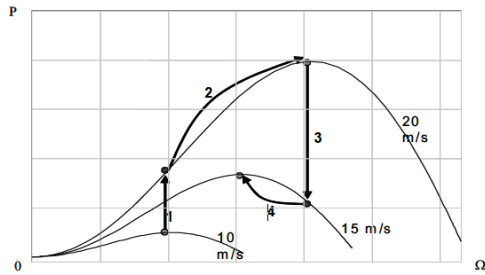
Figure 5: Principle of the MPPT at changing wind speed.

Control by fuzzy logic does not require the system modelling. However, a global knowledge of its behaviour, in closed loop for example, is needed. This knowledge must be transcribed to a form of rules.

The common scheme uses fuzzy logic method is shown in Figure 6:
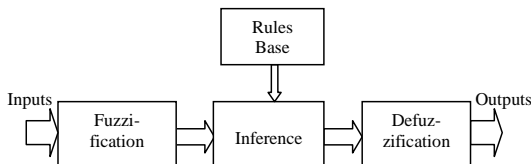
Figure 6: Basic scheme of fuzzy logic

In this case, ΔP and $\Delta\Omega$ are the inputs and $\Delta\Omega_{ref}$ is the output. The membership functions of the inputs are chosen as triangular and trapezoidal and a singleton one is used for the output. The expression of the rule is then as follows:

(If x_1 is A) AND (x_2 is B) THEN $S_k = C_k$ (16)

Here, C_k is a constant.

Figure 7 shows the inputs membership functions and Figure 8 gives the output membership one.
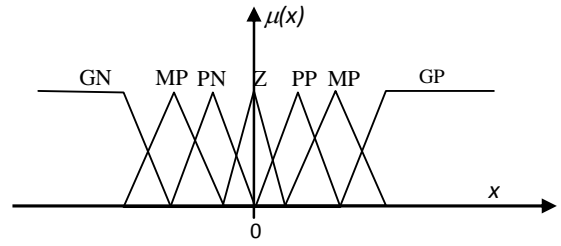
Figure 7: Inputs membership functions
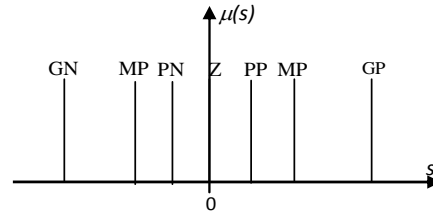
Figure 8: Output membership function

Table 1 resumes the rules base with a number of membership functions N = 7.

Table 1: Rules base

$\Delta\Omega_{(k-1)}$ \ ΔP	GN	MN	PN	ZE	PP	MP	GP
GN	GP	GP	MP	ZE	MN	GN	GN
MN	GP	MP	PP	ZE	PN	MN	GN
PN	MP	PP	PP	ZE	PN	PN	MN
ZE	GN	MN	PN	ZE	PP	MP	GP
PP	MN	PN	PN	ZE	PP	PP	MP
MP	GN	MN	PN	ZE	PP	MP	GP
GP	GN	GN	MN	ZE	MP	GP	GP

1. If a great increase speed involves a great increase in the power, the speed must be strongly increased.
2. If a great increase speed involves a great reduction in the power, the speed must be strongly decreased strongly to obtain a fast increase in the power.
3. If a great increase speed involves a weak increase in the power (near the optimal speed), the speed needs to be fairly increased.
4. If a null speed variation involves an increase in the power, one deduces that the wind speed is increased, it is thus necessary to increase the speed to approach the new optimal point.
5. If a null speed involves a reduction in the power, it means that the wind speed is decreased, it is thus necessary to decrease speed to approach the new optimal speed.

The central line ($\Delta\Omega[k-1]=ZE$) means that a variation of the power is due to a modification of the wind speed rather than to the variation of machine speed Ω. It breaks the horizontal axial symmetry of the Table 1.

In this paper, the Sugeno's methods are chosen: a singleton is used as the membership function of the rule consequent combined by max-min method for the rule evaluation. The Sugeno defuzzification is then weighted average method:

$$s = \sum \frac{\mu(s_k).s_k}{\mu(s_k)} \qquad (12)$$

4. SIMULATION RESULTS

The method consists in measuring, with a sampling time T_{MPPT}, the speed Ω, the voltage Uc and the current Ic on the continuous side and to calculate the power P. The power and speed variations are then deduced. Figure 9 shows the method.
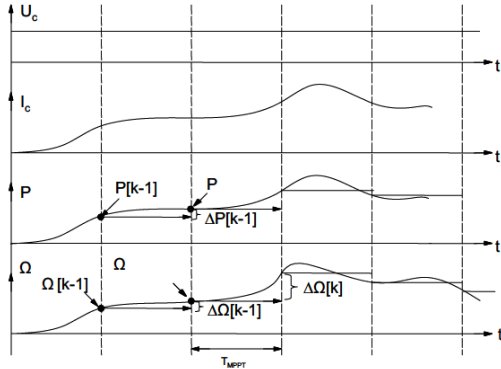
Figure 9 : Sampled inputs for the algorithm.

It may be highlighted here that measuring or calculating the power side receptor (for example battery) permits to obtain an optimization for the global system. Figure 10 shows the complete system.
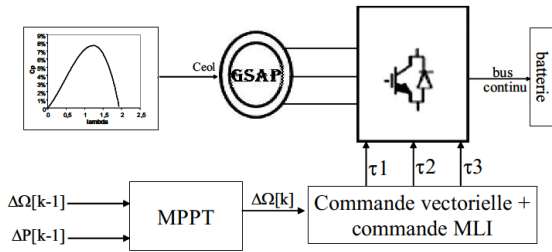
Figure 10: Complete scheme for the fuzzy -MPPT

Figure 11 gives a sample of wind speed measured on real site and used for the simulation.
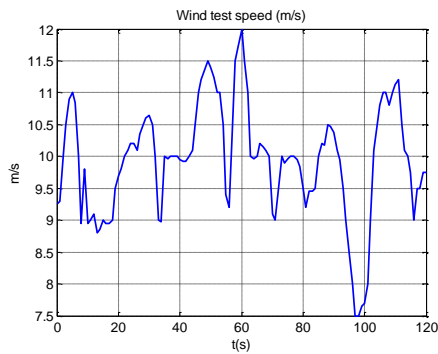
Figure 11 : Sample of the wind speed

Following figures resume the simulation results.
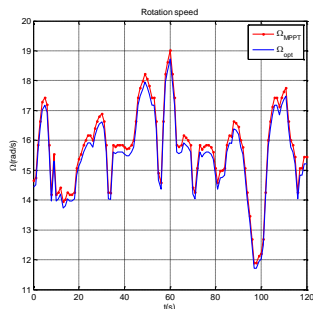
Figure 12: Speed obtained with MPPT and with Ω
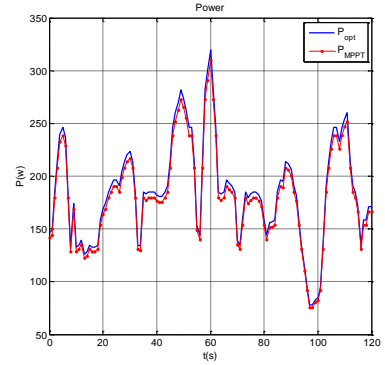
Figure 13: Power curves

The wind turbine speed determines the operation point compared to the point of maximum power. The difference between speed resulting from the MPPT and optimal speed results then in a loss into power presented on Figure13. The optimal power curve is calculated using the expression (17) and the maximized power curve corresponds at the speed resulting from the MPPT shown on Figure 12.

$$P_{opt} = \frac{1}{2} C_p^{opt} . \rho . S . V_v^3 \qquad (17)$$

Figures 13 and 14 give respectively the efficiency calculated according to the equation (18) and energies, calculated according to the equation (19).

$$\eta = \frac{P_{MPPT}}{P_{opt}} \qquad (18)$$

$$E = \int_{t0}^{t1} P(V_v) dt \qquad (19)$$

$$\Delta E = E(t_1) - E(t_0) \qquad (20)$$

$$\varepsilon_E^\% = \frac{\Delta E_{opt} - \Delta E_{MPPT}}{\Delta E_{opt}} 100 \qquad (21)$$
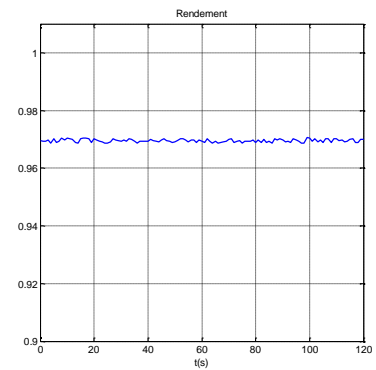
Figure 14: Effiency behaviour
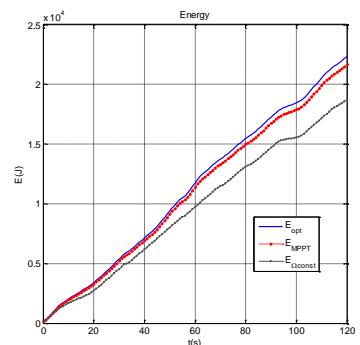
Figure 15: Energies obtained by different methods

According to the energy curves (see Figure15), it can be noted a profit approximately 13% on the energy recovered with MPPT method compared to the equivalent system operating with constant speed.

The representation of the energy "collected" by the system is a relevant criterion of appreciation of the energy effectiveness. The variation calculated according to the equation (16), between optimal energy resulting from optimal power P_{opt} and the obtained using MPPT method one gives a quantification of the energy quality taken on a given lapse of time. Here, this variation is 3,15%.

All these simulation results show the reliability of the adopted energy strategy optimization. It should be noted that the power is measured on electrical side.

5. CONCLUSION

An optimization method of energetic efficiency of a wind system based on a synchronous permanent magnet machine provided with a MLI rectifier converter is implemented in this article.The objective is to maximize the power collected for each wind speed.

The optimization approach adopted in this work makes it possible to move the operation point of the wind system by adjusting its speed at any wind speed, so that the collected power is always maximum.

The proposal offers a very interesting alternative compared to the traditional methods, which require the knowledge of the curve characteristic of the windmilln turbine.

The comparison between the energetic efficiency obtained by fuzzy MPPT with the one by constant speed makes it possible to validate the method. Indeed the average value of the output obtained MPPT, during the time of test is approximately 97%.

However, it may be noted that the algorithm is not able to distinguish if the power variation is due to the wind speed variation or to the variation of Ω.

REFERENCES

R. Abdessemed et M. Kadjoudj, 1997. Modélisation des Machines Electriques. Presses de l'Université de Batna, Algérie.

R. Abdessemed, 2011. Modélisation et Simulation des machines électriques. TechnoSup, Edition Ellipses, Paris.

C. Alonso, 2010. Contribution à l'optimisation, gestion et traitement de l'énergie, Mémoire en vue de l'obtention du HDR, Université Paul Sabatier, Toulouse III.

S. Belakehal, A. Bentounsi, M. Merzoug et H. Benalla, 2010.Modélisation et commande d'une génératrice, Synchrone à aimants permanents dédiée à la conversion de l'énergie éolienne. Revue des Energies Renovelables, Vol. 13, N°1, 149-161.

R. Cardenas-Dobson, 1996. Control of Wind Turbine Using a Switched Reluctance Generator. PhD Thesis, University of Nottingham.

O. Gergqud, 2002. Modélisation Energétique et Optimisation Economique d'un Système de Production Eolien et Photovoltaïque Couplé au Réseau et Associé à un Accumulateur. Thèse de Doctorat, ENS Cachan.

A. Mirecki, 2005. Etude comparative des chaines de conversion d'énergie dédiées à une éolienne de petite puissance. Thèse de doctorat, Institut National Polytechnique, Toulouse.

B. Multon, X. Roboam, B. Dakyo, C. Nichitat, O. Gergaud and H. Ben Ahmed, 2004. Aérogénérateurs électriques. Technique de l'ingénieur, Traité de génie électrique, D3960.

G. Sturtzer et E. Smichel, 2000. Modélisation et Commande des Moteurs Triphasés., Edition Ellipses, Paris.

E. W Kimbark, 1995. Power system stability : Synchronous Machines, ISBN 0-7803-1135-3, IEEE Press.

A.S. Toledo, 2000. Commande Directe et Observation des Convertisseur de Puissance: Application à l'Onduleur de Tension Triphasé. Thèse de Doctorat, Ecole Doctorale de l'Institut National Polytechnique, Grenoble.

AUTHORS BIOGRAPHY

Rakotoarisoa A. Jean Claude. Born in 1976. Doctor Engineer, He works in the laboratory of Automatic . His theme consists of optimization and identification, applied to electrical, mechanical systems and renewable energies

Razafinjaka N. Jean. Born in 1956. Doctor Engineer HDR. He works now in the Laboratory of Automatic, Polytechnic Higher School of Diego Suarez, Madagascar. His field of research relates to the advanced commands applied to electric machines, electromechanical systems and renewable energies.

Sonia Moussa. Born in 1989. Holder of Electronics Engineer Diploma of the Polytechnic Higher School of Diego Suarez. She is currently doing his PhD at the University of El Manar, Tunis, Tunisia. His theme is about micro-grid side DC bus.

MODELING SUPPLY CHAIN QUALITY MANAGEMENT PERFORMANCE

Juan Miguel Cogollo Flórez[a], Alexander Alberto Correa Espinal[b]

[a]Instituto Tecnológico Metropolitano-ITM, Medellín, Colombia
[b]Universidad Nacional de Colombia, Medellín, Colombia

[a]juancogollo@itm.edu.co, [b]alcorrea@unal.edu.co

ABSTRACT

The body of knowledge of Supply Chain Management has evolved by integrating new approaches and concepts, such as Quality Management. Quality Management research in individual companies is usual, but there is a lack of research on how to incorporate quality requirements in supply chains design and planning. This paper describes the design of a rule-based model for measuring Supply Chain Quality Management performance. The proposed model integrates Triple Bottom Line (3BL) approach with the fuzzy set theory, considering environmental, economic and social performances and the imprecision in quantification of Supply Chain Quality Management performance. The application of the model allowed obtaining a crisp Sustainable Supply Chain Quality Management Index incorporating imprecision and vagueness on these calculations through a fuzzy rule-based system.

Keywords: Modeling, Supply Chain Quality Management, Triple Bottom Line, Fuzzy Logic.

1. INTRODUCTION

Quality Management approach has evolved from the traditional scenario focused on the company to complex systems of supply chains. This change in focus has led to a change in competitive priorities of many companies: from product quality to overall supply chain quality (Kuei and Madu 2001). Moreover, research in Supply Chain Management shifted from an operational and tactical level focused on cost, delivery and risk to a more complex and demanding strategic level (Melnyk *et al*. 2009).

A model is a simplified representation of a system and may be conceptual, verbal, diagrammatic, physical or formal (analytical or mathematical). Sayama (2015) highlights the following two approaches for the construction of scientific models:

- Descriptive modeling: it consists of describing the real state of a system at any time using quantitative methods (e.g., regression analysis and pattern recognition).
- Rule-based modeling: it consists of formulating dynamic rules that explain the behavior observed in a system. This type of modeling allows making predictions of possible states of the system, using dynamic equations, rules and principles, among other quantitative methods.

Both descriptive and rule-based approaches are widely applied in quantitative research in engineering and sciences. However, rule-based modeling plays a prominent role in the research of complex systems (Sayama 2015): the development of a rule-based model at microscopic scales and the study of its macroscopic behavior through computational simulation and/or mathematical analysis is almost a fundamental requirement for the analysis of complex systems, such as Supply Chain Quality Management.

Quality modeling is a little studied issue in Supply Chain Management (Batson and Mcgough 2007). Theory and practices about quality in individual companies are usual, but there is a lack of research on how to incorporate quality requirements in designing and planning global supply chains (Carmignani 2009; Dellana and Kros 2014; Mota et al. 2015), and how can be linked these practices and management systems with partners in the supply chain context (Bayo-Moriones et al. 2011; Gylling *et al*. 2015; Truong *et al*. 2016).

Although some researches have been carried about alignment between the type of product, strategy and network management systems quality (Mendes Dos Reis 2011) and models for designing and evaluating supply chains with focus on quality (Bayo-Moriones *et al*. 2011; Das and Sengupta 2010; Rashid and Aslam 2012; Truong *et al*. 2016), these ones have a descriptive approach and there is little use of rule-based modeling methods.

Therefore, it is necessary development and implementation of rule-based models for measuring Supply Chain Quality Management performance, considering environmental, economic and social performances and imprecision of data. In order to contribute to generation and dissemination of knowledge in this area, this paper describes a fuzzy rule-based model based on 3BL approach for measuring Supply Chain Quality Management performance.

This paper is structured as follows: first, there are backgrounds about Supply Chain Quality Management and Triple Bottom Line approach. Then, we describe the stages of proposed fuzzy rule-based model for measuring Sustainable Supply Chain Quality Management performance, its theoretical aspects and the application way. Lastly, we show the results obtained by applying

the model and the analysis is made along with the conclusions.

2. SUPPLY CHAIN QUALITY MANAGEMENT

The theoretical foundation of Supply Chain Quality Management (SCQM) has been developed and its relevance in academic and industrial practice has been evaluated by comprehensively reviewing prior literature in major journals and inductively identifying the themes that emerge within it (Foster *et al*. 2011; Jraisat and Sawalha 2013; Kannan and Tan 2007).

The SCQM is the integration of Supply Chain Management (SCM) and Quality Management (QM) concepts. SCQM concept has been studied using different perspectives: systems approach, process approach, organizational learning, among others.

Kuei and Madu (2001) used a three-equations approach to outline a SCQM definition:

- SC = production-distribution network.
- Q = to meet market demands and to achieve customer satisfaction in a fast and cost-effective way, and
- M = to improve the conditions and the confidence for supply chain quality.

Robinson and Malhotra (2005) stated that "*SCQM is the formal coordination and integration of business processes involving all partner organizations in the supply chain to measure, analyze and continually improve products, services and processes in order to create value and achieve satisfaction of intermediate and final customers in the marketplace*".

Foster (2008) stated that SCQM is a system-based approach to improve performance, taking advantage of opportunities given by linkages between suppliers and customers. Mellat-Parasat (2013) stated that SCQM is the coordination and integration of inter-company processes involving all supply chain members through continuous improvement of inter-organizational processes in order to improve the performance and achieve customer satisfaction by emphasizing cooperative learning.

3. THE TRIPLE BOTTOM LINE

The Triple Bottom Line is a strategic issue and represents the three areas of sustainability: economic, environmental and social (Wilson 2015). Sustainability has become an important issue for most of the organizations and supply chains and it has created the need for developing non-financial measures of performance in addition to traditional measures (Agrawal *et al*. 2016; Hacking and Guthrie 2008).

Implementation of social responsibility policies and practices is not only the work of a company but of its supply chain, through the interrelation of partners to become a social responsible supply chain (Cruz-Trejos, Correa-Espinal and Cogollo-Flórez 2012).

Hubbard (2009) developed a Sustainable Balanced Scorecard conceptual framework and proposed an organizational sustainable performance index, calculated as the average performance in each of the six perspectives. This is a limitation of the model, since it assigns the same relative weight to each perspective.

Longo (2012) developed a simulation model for integrating sustainability aspects (technical, economic and environmental) in supply chain redesign and optimization. The implementation of metaheuristic techniques allowed to obtain positive impacts on the economic and environmental sustainability of the supply chain studied.

Montemanni *et al* (2013) designed a web-based software application for the sustainable design of a textile and apparel supply chain, in order to compute the environmental impact of the processes and material flows and to obtain alternative designs for the implementation of the supply chain, integrating cost and time performance indicators.

Rizzoli *et al* (2015) proposed a methodology for including the sustainability assessment in supply chain design and management. The tool is based in collection and organization of sustainability-related data of all stages of the product life cycle and then optimisation to choice in real time the better performing solution in sustainability.

Schulz and Flanigan (2016) developed a framework for a sustainability model in order to establishing a competitive advantage. The model integrates the concepts and roles of competitiveness with the 3BL theory and proposes some performance indicators for calculating Supply Chain Sustainability Index.

Fallahpour *et al* (2017) developed a model for supplier selection in sustainable supply chain management through a questionnaire-based survey. According to them, the economic aspect was the most essential, followed by environmental aspect and finally social aspect.

4. DEVELOPMENT OF A SUSTAINABLE SUPPLY CHAIN QUALITY MANAGEMENT INDEX

The development of the fuzzy rule-based model for measuring Supply Chain Quality Management performance was carried out through six stages: Development of performance indicators and setting fuzzy subsets parameters, Definition of fuzzy inference method, Elaboration of fuzzy rule-based system, Selection of defuzzification method and Calculation of Sustainable Supply Chain Quality Management Index.

4.1. Development of performance indicators and setting fuzzy subsets parameters

The first stage is the development of the indicators to measure performance in each one of the 3BL perspectives and thus to measure overall performance in SCQM. Table 1 shows the indicators in every 3BL perspective (Economic, Environmental and Social).

Table 1: Performance indicators and fuzzy subsets parameters of 3BL perspectives performance input variables

3BL	INDICATORS	FUZZY SUBSETS PARAMETERS		
		Low (trapezoidal)	Medium (triangular)	High (trapezoidal)
Economic	Percentage of logistics costs (%)	(0, 0, 6, 10)	(6, 10, 14)	(10, 14, 50, 50)
	Percentage of quality costs (%)	(0, 0, 15, 25)	(15, 25, 35)	(25, 35, 50, 50)
	Return on Assets (%)	(0, 0, 4, 7)	(4, 7, 10)	(7, 10, 50, 50)
Environ-mental	Average percentage of defective product (%)	(0, 0, 1, 2.5)	(1, 2.5, 4)	(2.5, 4, 20, 20)
	Emissions, effluent and waste as a percentage of total resources used (%)	(0, 0, 10, 20)	(10, 20, 30)	(20, 30, 40, 40)
	Average Percentage of Rejections and Returns (%)	(0, 0, 2, 5)	(2, 5, 8)	(5, 8, 30, 30)
Social	Customer Satisfaction measured as On Time In Full (OTIF) deliveries (%)	(0, 0, 70, 80)	(70, 80, 90)	(80, 90, 100, 100)
	Employee Satisfaction (%)	(0, 0, 40, 60)	(40, 60, 80)	(60, 80, 100, 100)
	Suppliers Development (%)	(0, 0, 30, 50)	(30, 50, 70)	(50, 70, 100, 100)

The parameters of input and output variables fuzzy sets for the calculations of performance by 3BL perspective and the SCQM Index were established by analyzing historical information (frequency histograms and segmentation into quartiles or quintiles according to the number of fuzzy sets established). The model was applied in the supply chain of a plastics sector company, whose name is kept in reserve due to commitments of confidentiality of the information supplied.

4.2. Definition of fuzzy inference method

A fuzzy inference method is used to obtain conclusions from "IF-THEN" rules and input values to the system by applying composition relations. Because of the outputs of system are continuous values, the Mandani type inference system was used in this model (Figures 1 and 2).

4.3. Elaboration of fuzzy rule-based system

The elaboration of rules is the most important stage in the design of fuzzy-rule based model inasmuch as it combines the expert opinion and the analysis of historical information. In this case, the Senior Management of the company used multiple comparison matrices for the elaboration of fuzzy rules, considering than the main objective of measuring Sustainable Supply Chain Quality Management is not financial or quality

performance but jointly the environmental, economic and social performances.
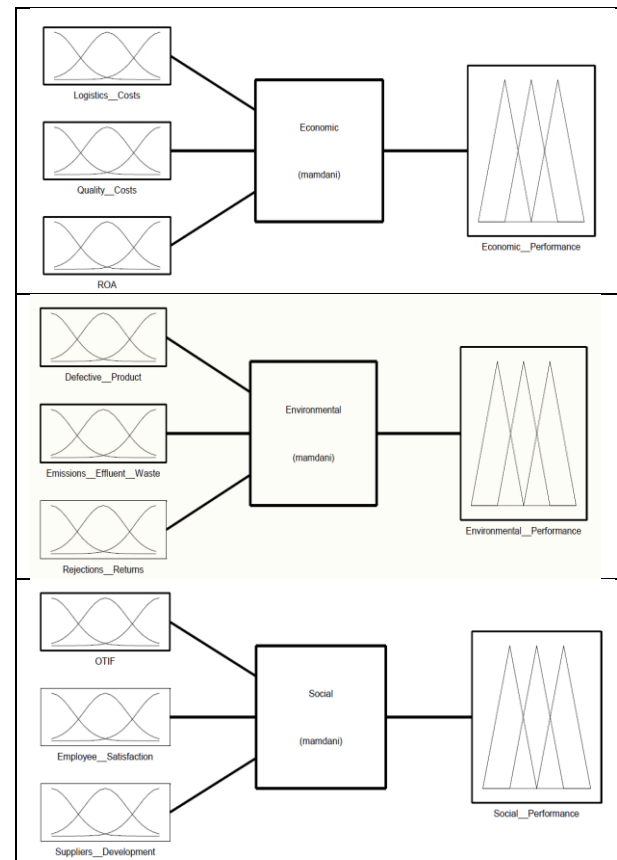
Figure 1: Fuzzy Inference Systems for Economic, Environmental and Social performance measurement.
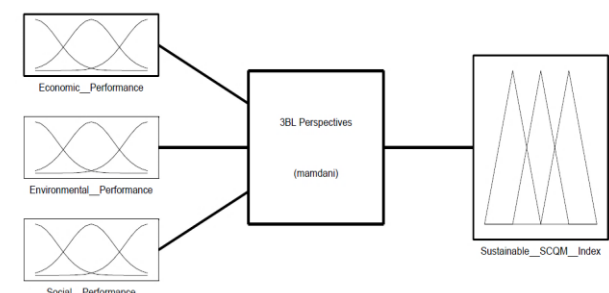
Figure 2: Fuzzy Inference System for measuring Sustainable Supply Chain Quality Management Index.

In this model are used "IF-THEN" rules, composed of the antecedent "IF" and the consequent "THEN", using connectors "AND" to develop necessary decision rules. This model contains three rules sets for 3BL perspectives and one rule set for Sustainable SCQM Index. The construction of fuzzy rule-based system was made by developing a conclusions matrix by considering all possible combinations of inputs and assigning a conclusion to each (Tables 2, 3, 4 and 5).

Table 2: Fuzzy rules matrix of Economic perspective

ROA		Logistics Costs	Quality Costs		
			Low	Medium	High
	Low	Low	M	L	L
		Medium	L	L	L
		High	L	L	L
		Logistics Costs	Quality Costs		
			Low	Medium	High
	Medium	Low	M	M	M
		Medium	M	M	L
		High	M	L	L
		Logistics Costs	Quality Costs		
			Low	Medium	High
	High	Low	H	M	M
		Medium	M	M	M
		High	M	L	L

Table 3: Fuzzy rules matrix of Environmental perspective

Defective product		Emissions, Effluent & waste	Rejections & Returns		
			Low	Medium	High
	Low	Low	H	H	M
		Medium	H	M	M
		High	M	M	L
		Emissions, Effluent & waste	Rejections & Returns		
			Low	Medium	High
	Medium	Low	H	M	M
		Medium	M	M	M
		High	L	L	L
		Emissions, Effluent & waste	Rejections & Returns		
			Low	Medium	High
	High	Low	M	M	L
		Medium	M	L	L
		High	L	L	L

Table 4: Fuzzy rules matrix of Social perspective

OTIF		Employee satisfaction	Suppliers development		
			Low	Medium	High
	Low	Low	L	L	L
		Medium	L	M	M
		High	L	M	M
		Employee satisfaction	Suppliers development		
			Low	Medium	High
	Medium	Low	L	M	M
		Medium	M	M	M
		High	M	M	M
		Employee satisfaction	Suppliers development		
			Low	Medium	High
	High	Low	M	M	M
		Medium	M	M	H
		High	M	H	H

Table 5: Fuzzy rules matrix of Sustainable SCQM Index

Economic Performance		Environmental Performance	Social Performance		
			Low	Medium	High
	Low	Low	L	L	L
		Medium	L	L	M
		High	L	M	M
		Environmental Performance	Social Performance		
			Low	Medium	High
	Medium	Low	L	M	M
		Medium	M	M	M
		High	M	M	M
		Environmental Performance	Social Performance		
			Low	Medium	High
	High	Low	M	M	M
		Medium	M	M	M
		High	M	H	H

Economic, Environmental and Social perspectives and Sustainable SCQM Index were evaluated on three input variables, which have three fuzzy categories (Table 1). Therefore, there are $3^3 = 27$ fuzzy rules in each of their systems.

Values in cells of Tables 2-5 represent the consequent describing each combination and correspond to linguistic labels of output variable fuzzy subsets, and "L" corresponds to low, "M" is medium and "H" is high. For example, in the Table 5, shaded cell corresponds to the following rule: IF "Economic performance" is High AND "Environmental performance" is Medium AND "Social Performance" is High THEN Sustainable SCQM Index is Medium.

The application of the Mandani type fuzzy inference system based on the rules systems previously described also allowed to obtain response surface 3D graphs for modeling the relation between the input and output variables in each 3BL perspective and in the Sustainable SCQM Index (Figures 3,4,5 and 6).

For example, Figure 6 shows that both Environmental and Economic performance contribute to the Sustainable SCQM Index, but the impact of Economic performance is greater at lower values than Environmental performance.
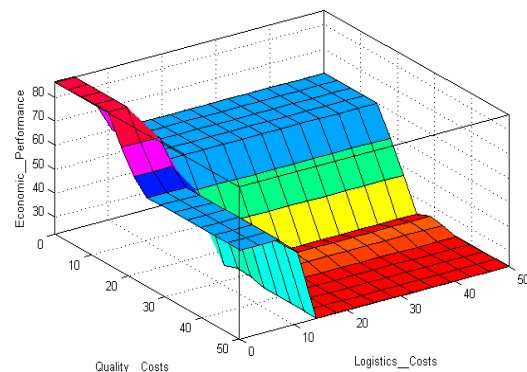
Figure 3: The 3D relation of input indicators (quality costs and logistics costs) and output variable (Economic Performance)
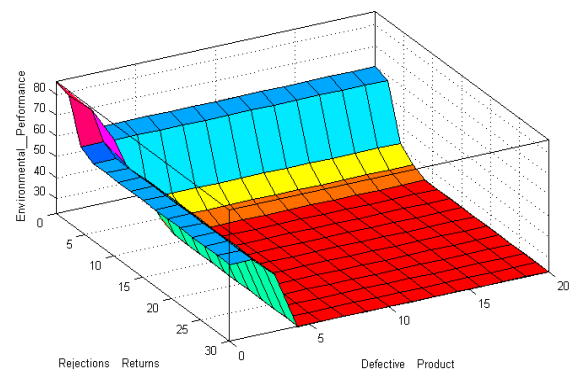
Figure 4: The 3D relation of input indicators (rejections and returns and defective product) and output variable (Environmental Performance)
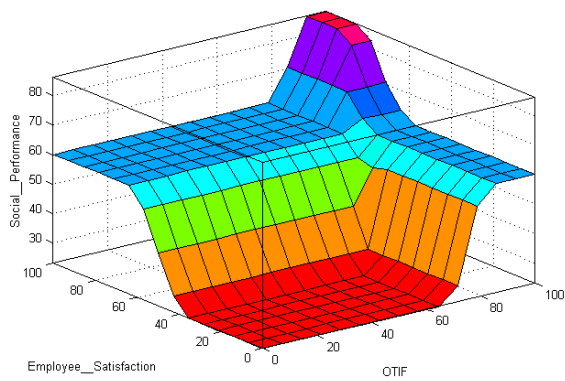
Figure 5: The 3D relation of input indicators (employee satisfaction and OTIF) and output variable (Social Performance)
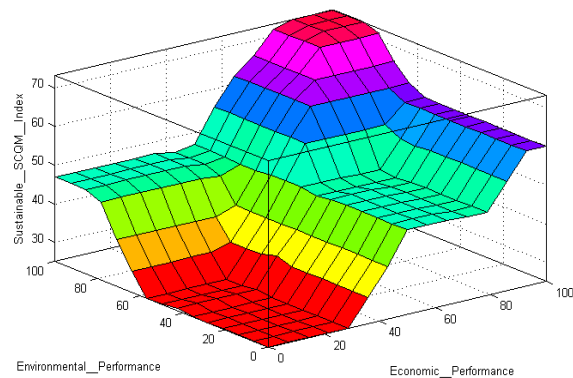
Figure 6: The 3D relation of input performance (Environmental and Economic) and output variable (Sustainable SCQM Index)

4.4. Selection of Defuzzification Method

Since the result of inference process is a set with a fuzzy distribution in response and for decision making it is necessary to use specific responses, the fuzziness must be removed and to obtain a crisp number.

The center of area method was used in this model for defuzzification, because of its continuity and that calculates the overlap area only once. The crisp value of the performance indicators for each 3BL perspective and the Sustainable SCQM Index were generated by the search of gravity center of the membership function of respective fuzzy outputs (Figures 7, 8, 9 and 10). The Fuzzy Logic Designer Application of Matlab© software was used as support for development of every stage of fuzzy rule-based model.
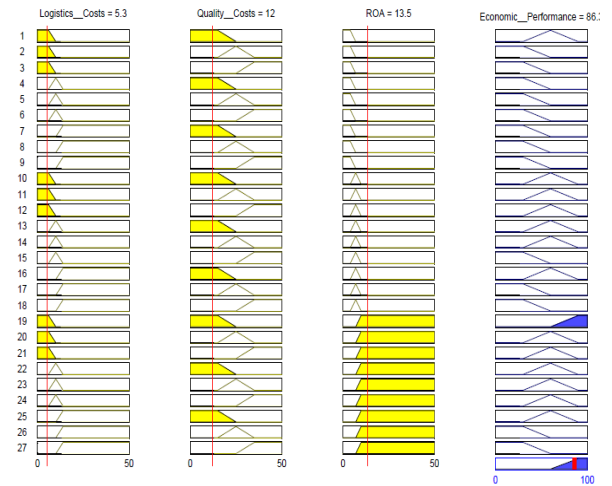
Figure 7: Scheme results of the fuzzy inference procedure to calculate Economic Performance
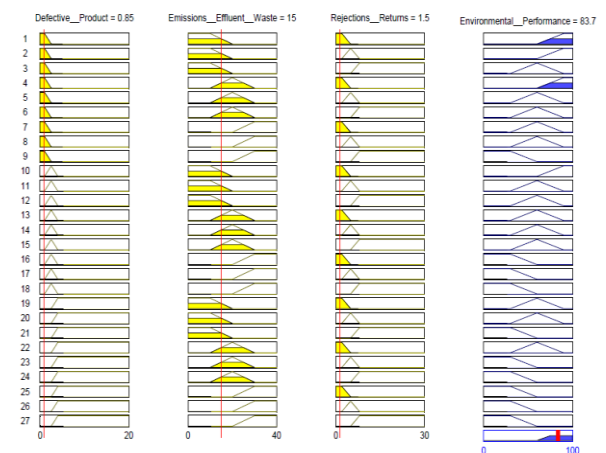
Figure 8: Scheme results of the fuzzy inference procedure to calculate Environmental Performance
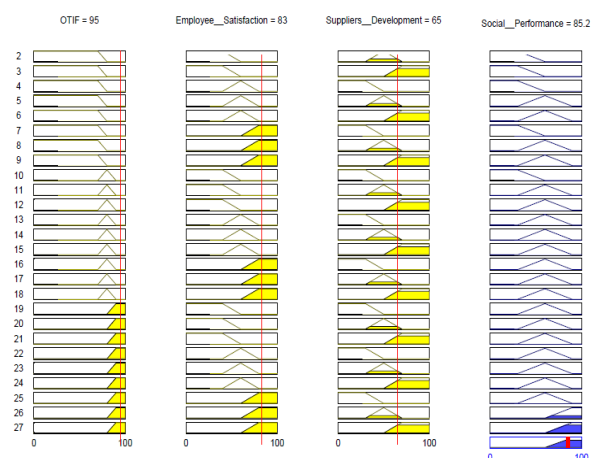
Figure 9: Scheme results of the fuzzy inference procedure to calculate Social Performance
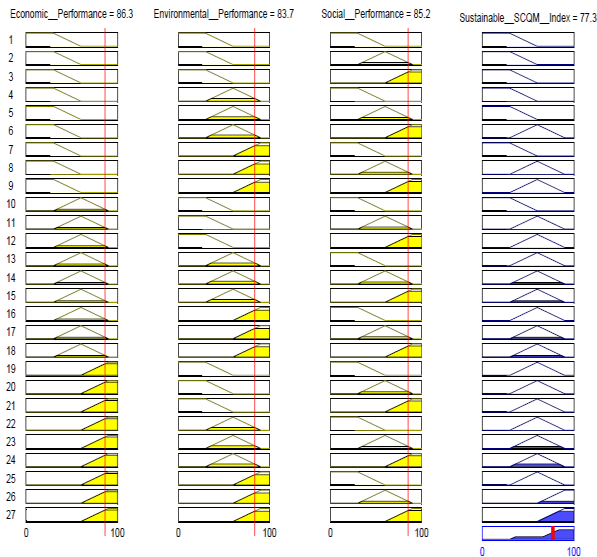
Figure 10: Scheme results of the fuzzy inference procedure to calculate Sustainable SCQM Index

4.5. Calculation of Sustainable Supply Chain Quality Management Index

The crisp numbers results of applying the fuzzy rule-based model for measuring Sustainable SCQM Index are shown in top of Figures 7 to 10. The results may be expressed in terms of the fuzzy rule-based systems, as follows:

- Economic Performance: IF Percentage of logistics costs is 5.3% AND Percentage of quality costs is 12% AND Return of Assets is 13.5% THEN Economic Performance is 86.3% (Figure 7).
- Environmental Performance: IF Percentage of defective product is 0.85% AND percentage of emissions, effluent and waste is 15% AND Percentage of rejections and returns is 1.5% THEN Environmental Performance is 83.7% (Figure 8).
- Social Performance: IF OTIF is 95% AND Employee satisfaction is 83% AND percentage of suppliers development is 65% THEN Social Performance is 85.2% (Figure 9).
- Sustainable SCQM Index: IF Economic Performance is 86.3% AND Environmental Performance is 83.7% AND Social Performance is 85.2% THEN Sustainable SCQM Index is 77.3% (Figure 10).

These numerical results may be interpreted as the compliance percentage of the goals and objectives of sustainability of the quality management practices in the supply chain context.

Moreover, it is important to highlight this model allows to make sensitivity analysis by varying the input variables values and then seeing the impact of the changes on the final performance. For example, if the company achieves to improve the performance of each 3BL perspective by 10% (thus, Economic Performance is 94.9%, Environmental Performance is 92.1% and Social Performance is 93.7%), the Sustainable SCQM Index would be 86.3% (Figure 11), which means that it

would be improved by 11.6%, compared to the initial result of 77.3%.
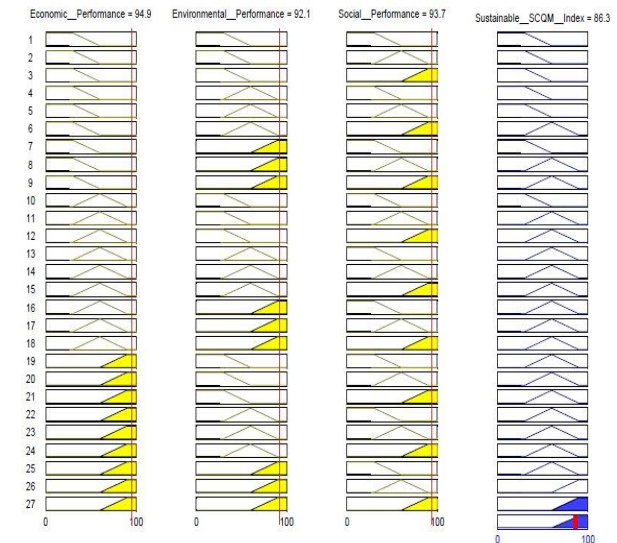
Figure 11: Results of the Sustainable SCQM Index obtained if the performance of each 3BL perspective would improve by 10%

5. CONCLUSIONS

The application of the fuzzy rule-based model allowed obtaining a crisp number as Sustainable SCQM Index that is important for decisions making. The model is based on a set of fuzzy rules easily interpretable and apprehended by the people involved in quality management in the supply chain.

The fuzzy rule-based model has a systematic structure that allows easy adaptation to others supply chains or other business management problems. The joint implementation of fuzzy logic and 3BL provides a new approach for modeling imprecision and interrelation of Economic, Environmental and Social variables in SCQM context.

REFERENCES

Agrawal S., Singh R.K. and Murtaza Q., 2016. Triple Bottom Line Performance Evaluation of Reverse Logistics. Competitiveness Review, 26 (3), 289–310.

Batson R.G. and Mcgough K.D., 2007. New direction in quality engineering: supply chain quality modelling. International Journal of Production Research, 45 (23), 455–5464.

Bayo-Moriones A., Bello-Pintado A. and Merino-Díaz-de-Cerio J., 2011. Quality assurance practices in the global supply chain: the effect of supplier localisation. International Journal of Production Research, 49 (1), 255–268.

Carmignani G., 2009. Supply chain and quality management: The definition of a standard to implement a process management system in a supply chain. Business Process Management Journal, 15 (3), 395–407.

Cruz-Trejos E., Correa-Espinal A.A. and Cogollo-Florez, J.M., 2012. Supply Chain Social Responsibility. Gestión y Región, 13, 89–106.

Das K. and Sengupta S., 2010. Modelling supply chain network: a quality-oriented approach. International Journal of Quality & Reliability Management, 27 (5), 506–526.

Dellana S. and Kros J., 2014. An exploration of quality management practices, perceptions and program maturity in the supply chain. International Journal of Operations & Production Management, 34 (6), 786–806.

Fallahpour A., Olugu E., Musa S., Wong K. and Noori S., 2017. A decision support model for sustainable supplier selection in sustainable supply chain management. Computers & Industrial Engineering, 105, 391–410.

Foster S.T., 2008. Towards an understanding of supply chain quality management. Journal of Operations Management, 26 (4), 461–467.

Foster S.T., Wallin C. and Ogden J., 2011. Towards a better understanding of supply chain quality management practices. International Journal of Production Research, 49 (8), 2285–2300.

Gylling M., Heikkila J., Jussila K. and Saarinen M., 2015. Making decisions on offshore outsourcing and backshoring: A case study in the bicycle industry. International Journal of Production Economics, 162, 92–100.

Hacking T. and Guthrie P., 2008. A framework for clarifying the meaning of Triple Bottom-Line, Integrated, and Sustainability Assessment. Environmental Impact Assessment Review, 28 (2), 73–89.

Hubbard G., 2009. Measuring organizational performance: Beyond the triple bottom line. Business Strategy and the Environment, 19, 177–191.

Jraisat L.E. and Sawalha I.H., 2013. Quality control and supply chain management: a contextual perspective and a case study. Supply Chain Management: An International Journal, 18 (2), 194–207.

Kannan V.R. and Tan K.C., 2007. The impact of operational quality: a supply chain view. Supply Chain Management: An International Journal, 12 (1), 14–19.

Kuei C.-H. and Madu C.N., 2001. Identifying critical success factors for supply chain quality management (SCQM). Asia Pacific Management Review, 6 (4), 409–423.

Longo F., 2012. Sustainable supply chain design: an application example in local business retail. Simulation, 88 (12), 1484–1498.

Mellat-Parast M., 2013. Supply chain quality management: An inter-organizational learning perspective. International Journal of Quality & Reliability Management, 30 (5), 511–529.

Melnyk S., Lummus R., Vokurka R., Burns L. and Sandor J., 2009. Mapping the future of supply chain management: a Delphi study. International Journal of Production Research, 47 (16), 4629–4653.

Mendes Dos Reis J.G., 2011. Modelo de Avaliação da Qualidade para Redes de Suprimentos. Universidade Paulista: Tese de Doutoramento em Engenharia de Produção.

Montemanni R., Valeri C., Nesic S., Gambardella L.M., Gioacchini M., Fumagalli T. Zeller H., Meyer K., Faist M. and Rizzoli A.E., 2013. Supply chain design and sustainability in the textile sector. Proceedings of 5th International Conference on Applied Operational Research, pp. 67-73. July 29-31, Lisbon (Portugal).

Mota B., Gomes M., Carvalho A. and Barbosa-Povoa A.P., 2015. Towards supply chain sustainability: Economic, environmental and social design and planning. Journal of Cleaner Production, 105, 14–27.

Rashid K. and Aslam M.M.H., 2012. Business excellence through total supply chain quality management. Asian Journal on Quality, 13 (3), 309–324.

Rizzoli A.E., Montemanni R., Bettoni A. and Canetta L., 2015. Software Support for Sustainable Supply Chain Configuration and Management. In: L. Hilty and B. Aebischer, eds. Advances in Intelligent Systems and Computing 310, ICT Innovations for Sustainability. Zurich: Springer, 271–283.

Robinson C.J. and Malhotra M.K., 2005. Defining the concept of supply chain quality management and its relevance to academic and industrial practice. International Journal of Production Economics, 96 (3), 315–337.

Sayama H., 2015. Introduction to the Modeling and Analysis of Complex Systems. New York: Open SUNY Textbooks.

Schulz S.A. and Flanigan R.L., 2016. Developing competitive advantage using the triple bottom line: a conceptual framework. Journal of Business & Industrial Marketing, 31 (4), 449–458.

Truong H., Sampaio P., Carvalho M., Fernandes A., Binh D. and Vilhenac E., 2016. An extensive structural model of supply chain quality management and firm performance. International Journal of Quality & Reliability Management, 33 (4), 444–464.

Wilson J.P., 2015. The triple bottom line: Undertaking an economic, social, and environmental retail sustainability strategy. International Journal of Retail & Distribution Management, 43 (4/5), 432–447.

AUTHORS BIOGRAPHY

Juan M. Cogollo Flórez is a Professor at the Department of Quality and Production, Instituto Tecnológico Metropolitano-ITM, Medellín, Colombia. He received a MSc in Management Engineering in 2011, from the Universidad Nacional de Colombia. He is currently a doctoral student in Engineering - Industry and Organizations at Universidad Nacional de Colombia. His current research interests are Performance Measurement, Supply Chain Quality Management and Advanced Statistical Quality Control.

Alexander A. Correa Espinal is a Full Professor at the Department of Organizations Engineering, Universidad Nacional de Colombia, Medellín, Colombia. He received the PhD in Statistics and Operational Research in 2007 from the Universitat Politècnica de Catalunya (Spain) and the MSc in Industrial Engineering in 1999 from the Universidad de los Andes (Colombia). His current research interests are Advanced Design of Experiments, Advanced Statistical Quality Control and Total Quality Management.

SIMULATION OF VIRTUAL HUMAN HAND EVOLUTION AFTER STROKE

Esteban Peña-Pitarch[a], NeusTicó-Falguera[b]

[a] Escola Poltècnica Superior d'Enginyeria de Manresa, Department of Mechanical Engineering, Manresa, SPAIN
[b]XarxaAssistencialAlthaia, Manresa, Spain

[a]esteban.pena@upc.edu, [b]29346ntf@gmail.com

ABSTRACT
Different types of neurological deficits and sequels in the upper extremities that affect the activities of daily living (ADL) in patients who have undergone stroke have been analyzed from a subjective clinical point of view.
The aim of this work is to show a novel environment to simulate the initial improvement of the upper limb functions a few days after stroke and simulate the functional recovery of patients under a rehabilitation program.
Eighteen patients in the first seven days of stroke were selected. Five men and thirteen women participated in the experiments. After six months, six of them recovered all the functionality of the hand tested with the ARAT test and wearing a Ciberglobe system, five recovered part of the hand functionality, and seven did not show apparent recovery of the functionality of the hand.

Keywords: stroke, hand simulation, virtual environment

1. INTRODUCTION

Stroke is one of the leading causes of death in industrialized countries as well as of disability and economic cost in adults. Stroke is associated with a connotation of poor prognosis and difficult recovery. It has an impact at personal, family, social and work level, in addition to producing a high expenditure for all health and social services.
Therapeutic advances of the last years, both in prevention and in diagnostic and therapeutic complexity, have determined a change in stroke management towards the multidisciplinary approach and the creation of specific units that has led to a significant decrease in mortality and its sequels.
Most studies on the evolution and functional prognosis of patients with acute stroke are focused on the assessment of gait recovery and the performance of daily life activities. Predictors of survival, hospital discharge, hospital stay, and overall motor recovery have also been described. Studies on the prediction of recovery of specific neurological deficits, such as upper extremity (UE) function, have been increasing in recent years. In this sense, in the last years there are more studies on the specific evaluation of the functional recovery of the UE after having suffered a stroke. This increase in the prevalence of this type of clinical research could also be due to the recent development of validated predictive measures of motor function of UE useful also to establish appropriate therapeutic programs (Chen and Winstein 2009).
Approximately 70-80% of patients with stroke have deficits in UE in the acute phase and 40% in the chronic phase (Nakayama et al. 1994, Broeks et al. 1999). These deficits limit voluntary movement, coordination, sensitivity, level of physical activity, as well as the realization of activities of daily living (Feys et al. 1998). This aspect implies a limitation and difficulty in their reintegration in their socio-labor environment (Nakayama et al. 1994) and it affects their quality of life (Nichols-Larsen et al. 2005).
Carrying out studies on prognostic factors of the functionality of UE paresis in people who have suffered a stroke is important because of its incidence, its prevalence, its sequels and disability, and its difficulty to predict recovery and functional prognosis of UE.
Man has evolved throughout history thanks to his brain and his hands. The main function of the hand is the grip. This has allowed humans -due to the opposing thumb and because it is longer than in the rest of animals- to make a grip with more precision and fineness. In addition, the hand is an indispensable sensory organ for the recognition of forms, volumes and distances, it sends to the cerebral cortex the evaluation and interpretation of the information that receives. It also constitutes the basis of a very peculiar sense that is the stereognosis (knowledge of the relief, thickness and space). The hand has the ability to recognize objects without having the view to participate.
Therefore, the involvement of UE and specifically the hand in diseases such as stroke implies in these patients a significant alteration in the performance of many activities of daily living, as well as motor, sensory and body expression limitations that can seriously affect the relationship of these people with their environment.
The aim of the present study is the simulation in patients at the beginning and at the end of six months of suffering from a stroke. The intention of this work is to show a novel virtual environment to simulate the improvement of the functions of the upper extremity a few days after having suffered a stroke and to simulate its recovery under a rehabilitation program.
This document continues with a section of materials and methods. Section 3 is dedicated to the results, the next section is the discussion and it ends with the conclusions.

2. MATERIALS AND METHODS

Patients admitted due to acute stroke and motor involvement of UE in the neurology department of Sant Joan de Deu Hospital who met the inclusion criteria and none of the exclusion criteria were selected.

The inclusion criteria were:

- Patients older than 18 years.
- Patients who have suffered a stroke for the first time with motor deficit in the UE, admitted to the service of Sant Joan de Deu hospital, with confirmation of neuroimaging brain injury during the first 48 hours.
- Patients without cognitive impairment that makes it difficult for them to understand and follow up the assessments.
- Patients who before the stroke were independent in their activities of daily living (ADL).
- To accept to participate in the study and signing the informed consent.

The exclusion criteria were:

- Patients with deficits and sequelae in their UE of any pre-stroke etiology.
- Patients with subsequent follow-up and control difficulties.
- Patients with terminal illness with a life expectancy of less than six months.

Withdrawal criteria were also taken into account, such as:

- Appearance of a new stroke during follow-up.
- Onset of concomitant pathology affecting the patient's vital prognosis and/or subsequent follow-up.
- That the patient decides voluntarily.
- Lack of compliance and collaboration on the part of the patient.
- Death of the patient.

In total, 18 patients who met the inclusion criteria were selected. The first follow-up visit was performed 3-4 days after the stroke and data were collected during hospital admission. The next visit and data collection was performed at 7 days, at 3 months and the last one was at 6 months' post-stroke. In this last visit, one of the patients was able to perform the whole validated Action Research Arm Test (ARAT) table. It is in this last visit that we will focus the simulation and in which we will create the virtual simulation environment.

2.1. Assessment of joint balance with Cyber-Globe II® glove

In order to obtain information on the angles of the joints of the hand, the 18-sensor Cyber-Globe II® glove was used, which also has a resolution of one degree, with a ratio of 90 measurements/s. An interface for capturing data with the glove was made, as shown in Figure 1.

After obtaining all the data with the glove, the data were converted to the hand model proposed by the authors [6]. The hand model is 25 degrees of freedom (DOF), not counting flexion/extension (F/E) and adduction/abduction (Ad/Ab) of the wrist, as shown in figure 2.

Figure 1: Screenshot of the data collection and numbering of the sensors.
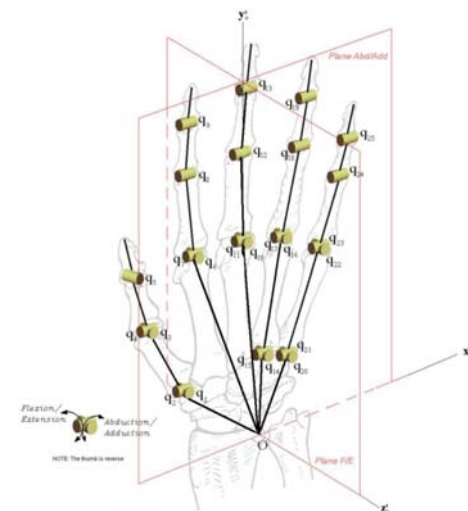
Figure 2: Hand with 25 DOF. Rear view right hand.

The implementation of a virtual model of 25DOF allows a very realistic simulation of the hand. The procedure for passing the readings from the 18 glove sensors to the 25 DOF model (hand only) will be the same for each joint. This procedure is based on linear interpolation, since there are some DOFs that depend on the readings of one or more sensors, and takes into account the maximum and minimum reading of the movement of each sensor and the maximum and minimum of the range of motion of each joint of the hand (Peña-Pitarch et al. 2014).

Once the assessment of the joint balance of each union and its subsequent conversion to the hand model with the 25 DOF was made, its implementation in the virtual environment of Blender -free code program- was carried out.

2.2. Simulation

Our approach for each of the patients has been using the model shown in figure 3. In this model in addition to the 25 DOF of the hand were added 4 more DOF, two for the wrist, in order to simulate the F/E of the same and the other for Ab/Ad. About the other two DOFs remaining, one is for the F/E of the elbow and the other is for supination/ pronation of the arm. In this work we have only focused on the study of the arm, not counting the shoulder, since we have considered more DOF focused on object grabbing. The future simulation of the trunk up to the hand is not ruled out.
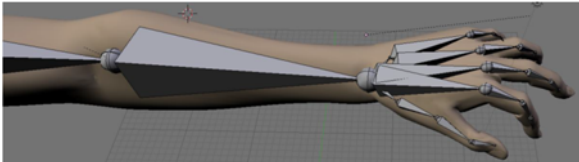
Figure 3: Left arm implemented in Blender with 29 DOF

The mathematical formulation applied to the model was previously published in (Peña-Pitarch et al. 2014) and it is as follows. The position vector is defined by $\mathbf{p}(\mathbf{q}^i)$ which with respect to the local coordinate system will be:

$$\begin{bmatrix} \mathbf{p}(\mathbf{q}^i) \\ 1 \end{bmatrix} = {}^0 A_1 A_2 \cdots {}^{n-1} A_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

Where $\mathbf{q}^i = [q_1 \ldots q_n]^T$, $i = $ I, II, III, IV, V, VI, VII, where I is the thumb, II is the index, III is the middle, IV the annular and V the pinkie, VI is for the wrist and VII for the arm. Where n is the total number of DOF, in this case it is 29.

To move it to a global coordinate system located on the shoulder, the position vector is now:

$$\begin{bmatrix} \mathbf{w}(\mathbf{q}^i) \\ 1 \end{bmatrix} = [{}^0\mathbf{H}_i] \begin{bmatrix} \mathbf{p}(\mathbf{q}^i) \\ 1 \end{bmatrix} \quad (2)$$

Figure 4 shows a screenshot of Blender, where on the left side we can see the code implemented with Python and allows to simulate each patient once the data obtained by the glove has been converted to the described model of 29 DOF, where the two of the elbows have been approximated by means of a goniometer.

Figure 4: Screenshot of a moment of the simulation of a patient.

3. RESULTS

The prognostic capacity of each of the joints in the functional recovery at 6 months according to the ARAT was evaluated both for the motion without gravity and for the movement against gravity when performing F/E of the fingers as well as the gain of movement experienced by each of these joints at 6 months' post-stroke. The joints corresponding to the metacarpo-phalangic and interphalangeal joints of all the fingers were evaluated.

3.1. Joints q_7, q_8, q_9. Index finger (F/E).

The mean q_7 joint and during the six-month follow-up the ARAT group≥10 experienced a gain of 65 degrees in the F/E movement range without severity. At three months the gain of movement was 26 degrees higher in the ARAT group≥10 than in the ARAT group<10. However, at six months the ARAT≥10 group experienced a gain of 41 degrees in the same movement but against gravity. As for the joint q_8 the gain was 21 degrees without gravity and between 19 to 39 degrees in the movement against gravity, for the group ARAT≥10. The joint q_9, the same group obtained a gain of 14 degrees and against gravity a gain of between 13 and 26 degrees.

3.2. Joints q_{11}, q_{12}, q_{13}. Middle Finger (F/E).

Always referring to the same group (ARAT≥10), in the joint q_{11} a gain of 22 degrees was obtained in the motion without gravity and between 14 and 17. The joint q_{12} without gravity experienced 21 degrees of gain and between 19 and 30 degrees in motion against gravity. Similarly, q_{13} gained 14 degrees and between 13 and 20 degrees with the same movement.

3.3. Joints q_{17}, q_{18}, q_{19}. Finger ring (F/E).

Similarly, the q_{17} joint experienced a gain of 18 degrees without gravity and between 15 and 24 degrees against gravity. The joints q_{18} and q_{19}, had gains of 22 and 15 degrees respectively and between 15 and 31 degrees.

3.4. Joints q_{23}, q_{24}, q_{25}. Pinky finger (F/E).

Likewise, the joint q_{23} experienced a gain of 19 degrees without gravity and between 12 and 24 degrees against gravity. The joints q_{24} and q_{25}, had gains of 28 and 18 degrees respectively and between 13 and 35 degrees.

3.5. ARAT bivariate analysis table 6 months. Clinical and functional characteristics at 3-4 days associated with ARAT.

Assessment 3-4 days	ARAT <10 N=6	ARAT ≥10 N=12	p-value
Deep sensitivity			
No altered	2 (14,3%)	12 (85,7%)	0,005[b]
Altered	4 (100,0%)	0 (0,0%)	
Urinary incontinence			
No	3 (20,0%)	12 (80,0%)	0,025[b]
Yes	3 (100%)	0 (0,0%)	
Fecal incontinence			
No	4 (26,7%)	11 (73,3%)	0,245[b]
Yes	2 (66,7%)	1 (33.3%)	
Hemineglect			
No	2 (14,3%)	12 (85,7%)	0,005[b]
Yes	4 (100%)	0 (0,0%)	
Muscular tone			
Normal	0 (0,0%)	7 (100%)	0,038[b]
Flaccidity	6 (54,5%)	5 (45,5%)	
NIHSS			
NIHSS	14,8 (DE=4,8)	6,4 (DE=3,3)	0,005[a]

Assessment 3-4 days	ARAT <10 N=6	ARAT ≥10 N=12	p-value
NIHSS sensitivity			
Normal	1 (20,0%)	4 (80,0%)	0,615[b]
Hypoaesthesia	5 (38,5%)	6 (61,5%)	
NIHSS motor UE			
Normal	0 (0,0%)	3 (100,0%)	0,013
Claudicate	0 (0,0%)	4 (100%)	
Not overcome gravity	1 (20,0%)	4 (80,0%)	
Motionless	5 (83,3%)	1 (66,7%)	
NIHSS orientation			
Responds two orders	3 (23,1%)	10 (76,9%)	0,123
Responds one order	1 (33,3%)	2 (66,7%)	
Not performancy	2 (100%)	0 (0,0%)	
NIHSS orders			
Performs two tasks	3 (20,0%)	12 (80%)	0,025[b]
Performs one task	3 (100%)	0 (0,0%)	
NIHSS conjugated gaze			
Normal movement	3 (20,0%)	12 (80,0%)	0,025[b]
Parcial paralysis	3 (100%)	0 (0,0%)	
NIHSS Visual fields			
No defects	2 (16,7%)	10 (83,3%)	0,082
Parcial hemianopsia	1 (50,0%)	1 (50,0%)	
Full hemianopsia	3 (75,0%)	1 (25,0)	
NIHSS Extension-inatencion			0,004
Without modifications	2 (15,4%)	11 (84,6%)	
Alteration a modality	4 (100%)	0 (0,0%)	
Severe hematution	0 (0,0%)	1 (100%)	
No mechanical pain	6 (33,3%)	12 (66,7%)	-
No neuropathic pain	6 (33,3%)	12 (66,7%)	-
Finger extensors			
0-2	6 (42,9%)	8 (57,1%)	0,245[b]
3-5	0 (0,0%)	4 (100%)	
Flexors fingers			
0-2	6 (50,0%)	6 (50,0%)	0,054[b]
3-5	0 (0,0%)	6 (100%)	
Wrist extensions			
0-2	6 (42,9%)	8 (57,1%)	0,245[b]
3-5	0 (0,0%)	4 (100%)	
Flexors wrist			
0-2	6 (50,0%)	6 (50,0%)	0,054[b]
3-5	0 (0,0%)	6 (100%)	
Elbow extensors			
0-2	6 (50,0%)	6(50,0%)	0,054[b]
3-5	0 (0,0%)	6 (100%)	
Flexorselbow			
0-2	6 (54,5%)	5 (45,5%)	0,038[b]
3-5	0 (0,0%)	7 (100%)	
Abductionshoulder			
0-2	6 (54,5%)	5 (45,5%)	0,038[b]
3-5	0 (0,0%)	7 (100%)	
Flexors shoulder			
0-2	6 (50,0%)	6 (50,0%)	0,054[b]
3-5	0 (0,0%)	6 (100%)	
Fugl Meyer UE	4,8 (DE=2,0)	31,8 (DE=22,3)	0,003[a]
Barthel index	7,5(DE=3,5)	27,5(DE=10,6)	0,001[b]
Ranking scale			
0-2	0 (0,0%)	2 (100%)	0,529[b]
3-5	6 (37,5%)	10 (62,5%)	

[a]U Mann-Whitney; [b]Fisher exact prove; [c]p-value bilateral Monte Carlo.

4. DISCUSSION

The measurement of the flexion and extension movement of the finger joints is a part of the basic exploration of many clinicians in their daily explorations. This measurement can be done quickly, simply and next to the patient's bed. The instrumentation of the measurement of the amplitude of the active range of motion has allowed us to obtain a more accurate data of the active movement of each joint and to implement it in the virtual environment as well as to simulate this movement for each patient.

When analyzing the range of motion (difference between minimum and maximum F/E) of each joint, significant differences were observed at 3 months between the two ARAT groups in the interphalangeal joints of the four fingers, and in the metacarpal joints -phalangeal of the ring finger and pinky in both positions, without and against gravity. But these data do not provide much information because at 3 months it is considered that many patients have recovered most of the functionality, and it is not an early assessment that allows us to define individualized programs of rehabilitation treatment. Only in the evaluation of the initial week there are significant changes in the proximal and distal interphalangeal joints of the index and the annular of the hand in the movement against gravity. Of these two fingers, the interphalangeal joints with more joint gain are those of the ring finger. The gain was 22 degrees with respect to 19 in the proximal interphalangeal joint, and 15 with respect to 13 degrees in the distal interphalangeal joint of the ring finger with respect to the index finger.

Thus, in the analysis of the range of motion between the two ARAT groups, the joints with the greatest capacity predicted at the week are the proximal and distal interphalangeal of the index finger and the annular one in the position against gravity, and of these two the ring finger.

The active range of flexion of each finger of the hand, flexion of the back and elbow, pronation and supination of the elbow, and flexion and extension of the wrist were examined in (Beebe et al. 2009). These authors observed that the active flexion of the middle finger and the presence of abduction in the lower back evaluated in the first 3 weeks post stroke had better predictive capacity at three months of the stroke than the rest of the fingers and joints of UE. In contrast, in another study, it was the active extension of the index and middle fingers against severity at 3 weeks of stroke which were strongly predictive of recovery at 13 weeks post stroke (Lang et al. 2006). Mirbagheri et al. (2008) identified the active range of motion and maximal voluntary contraction of elbow flexion and extension movements at 4 weeks of stroke as predictors of UE motor recovery.

The table in section 3.5 shows the evolution of the clinical and functional characteristics of the patients according to the ARAT and for each one of the evaluations that were performed.

The group of patients with ARAT ≥ 10 had a mean NIHSS score of 6.4 (SD = 3.3), whereas the group of patients with an ARAT <10 had a mean of 14.8 (SD = 4.8) in the 3-4 days' assessment. At 3-4 days, in the ARAT group <10 the patients had a mean FM-UE of 4.8 (SD = 2.2), and in the ARAT group ≥10, a mean of 31.8 (SD = 22.3). In the Barthel index, the mean was 7.5 (SD = 3.5) in the ARAT group <10 and 27.5 (SD = 10) in the ARAT group> 10.

Each patient in the ARAT group ≥10 presented alteration of the deep sensitivity, but had no urinary incontinence or hemineglect. All patients in the ARAT group <10 had an ERM ≥ 3 in all follow-ups that were made (See table section 3.5).

The movement of the fingers of the hand is important to acquire the skill and for the manipulation of objects that determine the proper functionality of the ES. The hand has multiple functions; the most important are touch, which is a sensitive function and grip, which is a motor function. To carry out these functions, the hand adopts different positions according to the type of grip it has to

do. In all these positions involve more or less fingers, but in general the little finger intervenes when it is necessary to grasp objects of greater weight and volume. All fingers are important for manipulation and grasping, but it is possible that the one that does not so much determine the functionality of UE is the little finger because it is only used when objects are heavy and bulky.

The results of this study highlight the predictive capacity for the recovery of UE function at 6 months after stroke.

5. CONCLUSIONS

The biomechanical evaluation of the fingers and their simulation in a virtual environment may facilitate the stratification of the patients in groups at risk according to the prognosis of the recovery of the paresis UE. This fact would help health professionals to make a more individual planning of neuro-rehabilitating treatment of patients who have suffered a stroke. Adequate patient selection would increase the efficiency of rehabilitation services.

The data obtained with the Cyber-Globe II® instrumentation glove are useful for designing technical aids or orthoses that help to promote independence in the activities of the daily life of the patients who need it. These data have been adapted to the virtual model with 29 DOF, which has allowed the simulation of each patient, giving an approximation of the abilities and limitations in their ADL.

The virtual simulation of the arm and hand in patients with stroke gives a new objective tool to physicians that allows simulating the evolution of deficits in some patients. The relevance of this work for patients affected by these deficits is that in the first visit to the doctor after the stroke it is possible to know through the simulation the evolution of their functional recovery. Another relevance is that the arm and hand model has been implemented with parametric lengths and can be extrapolated to other affected patients with the same deficits. Finally, ergonomists can be given data on hand movements and help design new products for people with reduced mobility due to a partial recovery after having followed the rehabilitation program.

REFERENCES

Chen S, Winstein C., 2009. A systematic review of voluntary arm recovery in hemiparetic stroke: critical predictors for meaningful outcomes using the international classification of functioning, disability, and health. J NeurolPhysTher. 33(1), 2–13

Nakayama H.,Jorgensen, H., Raaschou, H., Olsen, T., 1994. Recovery of upper extremity function in stroke patients: the Copenhagen Stroke Study. Arch Phys Med Rehabil. 75(4),394–398.

Broeks J., Lankhorst, G., Rumping, K., Prevo, A., 1999. The long-term outcome of arm function after stroke: results of a follow-up study. DisabilRehabil. 21(8), 357–364.

Feys, H., De Weerdt, W., Selz, B., Cox Steck, G., Spichiger, R., Vereeck, L., 1998. Effect of a therapeutic intervention for the hemiplegic upper limb in the acute phase after stroke: a singleblind, randomized, controlled multicenter trial. Stroke. 29(4), 785–792.

Nichols-Larsen, D., Clark, P., Zeringue, A., Greenspan, A., Blanton, S., 2005. Factors influencing stroke survivors" quality of life during subacute recovery. Stroke. 36(7),1480–1484.

Peña-Pitarch, E., Tico-Falguera, N., Yang, J., 2014. Virtual human hand: model and kinematics. Computer Methods BiomechBiomedEngin. 17(5), 568–579.

Beebe, J., Lang, C., 2009. Active Range of Motion Predicts Upper Extremity Function 3 Months After Stroke. Stroke [Internet]. 40(5), 1772–1779.

Lang, C., Wagner, J., Edwards, D., Sahrmann, S., Dromerick, A., 2006. Recovery of grasp versus reach in people with hemiparesis poststroke. Neurorehabil Neural Repair.20(4), 444–454.

Mirbagheri, M., Rymer, W., 2008. Time-course of changes in arm impairment after stroke: Variables predicting motor recovery over 12 months. ArchPhysMedRehabil. 89(8), 1507–1513.

MULTI-POLE MODELLING AND INTELLIGENT SIMULATION OF A FLUID POWER FEEDING SYSTEM WITH A PNEUMO-HYDRAULIC ACCUMULATOR

M. Harf[(a)], G. Grossschmidt 2[(b)]

Tallinn University of Technology, Estonia
[(a)] Institute of Software Science
[(b)] Institute of Mechanics and Industrial Engineering

[(a)] mait@cs.ioc.ee, [(b)] gunnar.grossschmidt@ttu.ee

ABSTRACT

An approach based on multi-pole modelling and intelligent simulation is proposed for design of a feeding subsystem with a pneumo-hydraulic accumulator for a fluid power system. Multi-pole mathematical models of feeding system components are presented. Modelling and simulation is explained on the hydraulic feeding system including electric motor, variable displacement axial piston pump, three-directional flow regulating valve and hydraulic accumulator together with hydraulic resistors and check valve. An intelligent visual simulation environment CoCoViLa supporting declarative programming in a high-level language and automatic program synthesis is used as a tool. Simulation examples of dynamics illustrating the behaviour of the accumulator in charging and discharging processes are presented and discussed. Using the proposed models and methods fluid power systems can be designed that are less sensitive to shock effects and high amplitudes of oscillations in the system.

Keywords: fluid power feeding system with a pneumo-hydraulic accumulator, multi-pole model, intelligent programming environment, simulation.

1. INTRODUCTION

To make fluid power feeding systems more flexible it is reasonable to use a hydraulic pump together with hydraulic accumulator. An accumulator enables a hydraulic system to cope with extremes of demand using a less powerful pump, to respond more quickly to a temporary demand, and to reduce shock effects and amplitudes of oscillations in a system.

A hydraulic accumulator is a pressure storage reservoir in which a hydraulic fluid is held under pressure that is applied by an external source. The external source can be a spring, a raised weight, or a compressed gas.

A compressed gas accumulator consists of a cylinder with two chambers that are separated by an elastic diaphragm, a totally enclosed bladder, or a floating piston. One chamber contains hydraulic fluid and is connected to the hydraulic line. The other chamber contains an inert gas under pressure that provides the compressive force on the hydraulic fluid.

In the paper diaphragm and bladder accumulators are considered, the dependences and models concern floating piston accumulators as well.

The stages of working of a bladder hydraulic accumulator are shown in Figure 1.
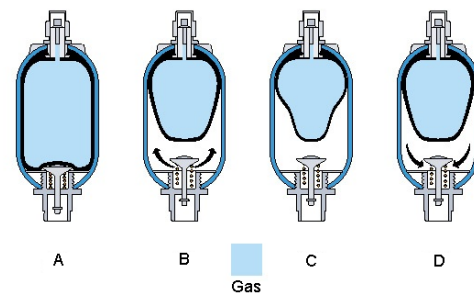
Figure 1: Stages of working of a bladder hydraulic accumulator

Stage A: The accumulator is pre-charged.
Stage B: The hydraulic system is pressurized. As system pressure exceeds gas pre-charge hydraulic pressure fluid flows into the accumulator.
Stage C: System pressure peaks. The accumulator is filled with fluid to its design capacity.
Stage D: System pressure falls. Pre-charge pressure forces fluid from the accumulator into the system.

In (Mamčic and Bogdevičius 2010) review and analysis of hydraulic accumulators and a number of links to scientific works are presented. The paper focuses on pressure pulsations in hydraulic systems, the means reducing them and examines the structure of hydraulic accumulators, including their features and differences.

The analysis of pneumo - hydraulic accumulator efficiency, applied as element of hybrid driving system is presented in (Chrostowski and Kedzia 2004).

Dynamics of accumulators together with hydraulic tubes is analyzed and natural frequencies are calculated in (Murrenhoff 2005).

In (Barnwal, Kumar, Kumar, and Das 2014) effect of hydraulic accumulator on the parameters of a transmission system is considered. The study deals with the surge absorbing characteristics of a hydraulic accumulator and is focused to finding out the suitable size of accumulator which will give less pulsation. The hydraulic system is modelled using MATLAB-SimHydraulics software.

In the current paper an approach is proposed, which is based on using multi-pole models with different oriented causalities (Grossschmidt and Harf 2009, 2014) for describing components of fluid power

systems. In the paper multi-pole modelling of feeding system with a pneumo-hydraulic accumulator is proposed.

2. MULTI-POLE MODELS

In general a multi-pole model represents mathematical relations between several input and output variables (poles). The nearest to physical nature of various technical systems is using multi-pole mathematical models of their components and subsystems.

In hydraulic and mechanical systems variables are usually considered in pairs (effort and flow variable). Multi-pole models enable to express both direct actions and feedbacks.

Each component of the system is represented as a multi-pole model having its own structure including inner variables, outer variables (poles) and relations between variables.

Using multi-pole models allows describe models of required complexity for each component. For example, a component model can enclose nonlinear dependences, inner iterations, logic functions and own integration procedures. Multi-pole models of system components can be connected together using only poles. It is possible directly simulate statics or steady state conditions without using differential equation systems.

The multi-pole model concept enables us to describe mathematical models visually which facilitates the model developing.

3. SIMULATION ENVIRONMENT

CoCoViLa is a flexible Java-based simulation environment that includes both continuous-time and discrete event simulation engines and is intended for applications in a variety of domains (Kotkas, Ojamaa, Grigorenko, Maigre, Harf, and Tyugu 2011). The environment supports visual and model-based software development and uses structural synthesis of programs (Matskin and Tyugu 2001) for translating declarative specifications of simulation problems into executable code.

Designer do not need to deal with programming, he can use the models with prepared calculating codes. It is convenient to describe simulation tasks visually, using prepared images of multi-pole models with their input and output poles.

4. SIMULATION PROCESS ORGANIZATION

Using visual specifications of described multi-pole models of fluid power system components one can graphically compose models of various fluid power systems for simulating statics or steady state conditions and dynamic responses.

When simulating statics or steady state conditions fluid power system behaviour is simulated depending on different values of input variables. Number of calculation points must be specified.

When simulating dynamic behaviour, transient responses in certain points of the fluid power system caused by applied disturbances are calculated.

Disturbances are considered as changes of input variables of the fluid power system (pressures, volumetric flows, load forces or moments, control signals, etc.). Time step length and number of steps are to be specified. For integrations in dynamic calculations the fourth-order classical Runge-Kutta method is used in component models.

Static, steady state and dynamic computing processes are organized by corresponding process classes (static Process, dynamic Process). To follow the system behaviour, the concept of state is invoked. State variables are introduced for each component to characterize its behaviour at the current simulation step.

A simulation task requires sequential computing states until some satisfying final state is reached. A final state can be computed from a given initial state if a function exists that calculates the next state from known previous states. This function is to be synthesized automatically by CoCoViLa planner.

A special technique is used for calculating variables in loop dependences that may appear when multi-pole models of components are connected together. One variable in each loop is split and iteratively recomputed to find its value satisfying the loop dependence.

State variables and split variables must be described in component models. When building a particular simulation task model and performing simulations state variables and split variables are used automatically.

5. MULTI-POLE MATHEMATICAL MODELS OF A PNEUMO-HYDRAULIC ACCUMULATOR

5.1. Multi-pole models of pneumo-hydraulic accumulators

A multi-pole model of a pneumo-hydraulic accumulator has input pressure p and output volumetric flow Q (Figure 2a) or input volumetric flow Q and output pressure p (Figure 2b). Additionally the models have the outputs of gas volume V and of temperature T of gas.
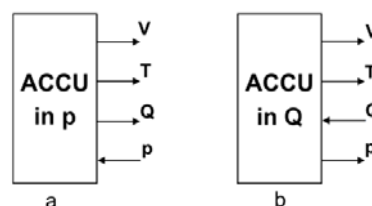
Figure 2: Multi-pole models of a pneumo-hydraulic accumulator

The model (a) is used for calculating static characteristics. The model (b) is used for calculating dynamic transient responses.

The model (a) is more natural also for dynamic. But if the model is directly connected to the resistor with check valve, the iterative calculation process solving the loop dependence during the simulation is not stable. If using model (b) calculations turn out to be stable.

5.2 Mathematical models and characteristics for statics

Here the following notations taken from (Murrenhoff 2005) are used:

$p0$ - gas pre-charge pressure to accumulator,
$p1$ - minimum pressure from accumulator,
$p2$ - maximum pressure from accumulator,
$p3$ - safety valve pressure,
$V0$ - maximum gas volume at pressure $p0$,
$V1$ - gas volume at minimum pressure $p1$,
$V2$ - gas volume at maximum pressure $p2$,
$V3$ - gas volume when safety valve turns on at pressure $p3$.

Gas pre-charge pressure to accumulator is usually taken

$$p0 = 0.9*p1.$$

Gas volume is calculated using formula

$$V = V0 * (p0/p)^{1/k},$$

where
k - polytrope exponent,
$k = 1$ in case of isothermal process,
$k = 1 \ldots 1.4$ in case of polytropic process,
$k = 1.4$ in case of adiabatic process.

Fluid volume in the accumulator is expressed as

$$Vf = V0 – V.$$

Maximum available fluid volume from the accumulator is expressed as

$$Vfmax = V1 – V2.$$

Static characteristic of an accumulator representing dependence of the gas volume on the pressure is shown in Figure 3.
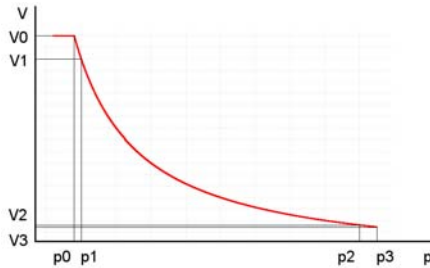
Figure 3: Static dependence of the gas volume against the pressure of in accumulator

This characteristic is calculated as a result of simulation accumulator statics using model from Figure 2a. For pressures $p0 = 0.9e6$ Pa, $p1 = 1e6$ Pa, $p2 = 5$ Pa and $p3 = 5.5$ Pa the volumes are equal $V0 = 1e-3$ m³, $V1 = 0.9e-3$ m³, $V2 = 0.18e-3$ m³, $V3 = 0.164e-3$ m³, $Vfmax = 0.72e-3$ m³ at isothermal process.

5.3 Mathematical models for dynamics

For dynamics (accumulator model from Figure 2b) we have an adiabatic process ($k = 1.4$).
Volume elasticity of the gas

$$CA = dV / dp = – V0 * p0^{1/k} / (k * pold^{1/k + 1}),$$

where

$pold$ – pressure at previous simulation step,
and volume elasticity of the fluid

$$CF = – Vf * βm,$$

where

$$Vf = V0 * (1 – (p0/pold)^{1/k}),$$

$βm$ – compressibility factor of fluid consisting air.

Sum of gas and fluid elasticities

$$C = CA + CF.$$

In the formulas pressure at previous simulation step $pold$ is used, as the pressure p at current simulation step will be calculated later.
The output pressure is calculated as

$$p = pold + Q * dt / C,$$

where
dt – simulation time step length.
The output gas volume

$$V = V0 * (p0/p)^{1/k}.$$

The output gas temperature (in ºC) is calculated as

$$T = (Told +273.15) * (p/pold)^{(k-1)/k} – 273.15,$$

where

$Told$ – gas temperature at previous simulation step.

6. MULTI-POLE MATHEMATICAL MODELS OF A SIMPLE FLUID POWER FEEDING SYSTEM COMPONENTS

Functional scheme of a simple fluid power feeding system with an accumulator is shown in Figure 4.
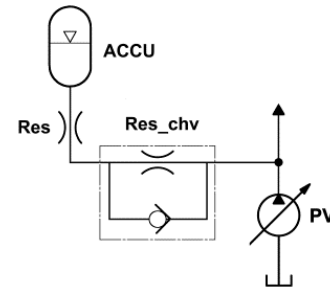
Figure 4: Functional scheme of a simple fluid power feeding system with an accumulator

The simple feeding system with an accumulator includes hydraulic accumulator **ACCU** together with hydraulic resistor **Res** at accumulator inlet, hydraulic resistor with check valve **Res_chv** and a variable displacement axial piston pump **PV**.
The feeding system of proposed configuration is usable for such fluid power systems where pressure at pump depends on load (load sensing systems, etc.).
Oriented graph of the hydraulic feeding system with an accumulator for dynamics is shown in Figure 5.

The oriented graph contains all the hydraulic system components and a hydraulic interface element IEH4. The graphs show all the oriented relations between variables and all the loop dependencies.
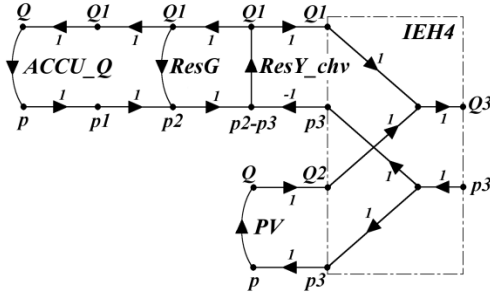
Figure 5: Oriented graph of simple hydraulic feeding system with an accumulator for dynamics

Mathematical model of a hydraulic resistor ResG

$$p2 = p1 - (RL + RT*abs(Q1)) * Q1,$$

where
RL, RT – hydraulic linear and square flow resistances (Grossschmidt and Harf 2010).

Mathematical model of a hydraulic resistor with check valve ResY_chv:

hydraulic resistance of check valve if (p2 <= p3, accumulator charging)

$$R = \Delta p_n / Q_n,$$

where

Δp_n – nominal pressure drop,

Q_n – nominal volumetric flow,

volumetric flow through check valve

$$Q1 = - (p3 - p2) / R,$$

where
p2, p3 – pressure at left and right port,

volumetric flow through hydraulic resistor if (p2 > p3, accumulator discharging)

$$Q1 = \mu * \pi * d^2 / 4 * ((2*(p2 - p3) / \rho)^{\frac{1}{2}},$$

where
μ – flow coefficient,
d – inner diameter of resistor,
ρ – fluid density.

In interface element IEH4

$$Q3 = Q1 + Q2.$$

Output volumetric flow of variable displacement axial piston pump PV

$$Q = \omega * V * \eta_{vol} \ m^3/s,$$

where
ω – angle velocity rad/s,
working volume of the pump

$$V = V_{max} * \tan (al) / \tan (al_{max}) \ m^3/rad,$$

where

V_{max} – maximum working volume m^3/rad,

al – position angle of the swash plate rad,

al_{max} – maximum position angle of the swash plate rad,

volumetric efficiency coefficient

$$\eta_{vol} = 1 - k_{vol}* p,$$

where

k_{vol} – coefficient characterizing the dependence from p,

p – inlet pressure.

7. SIMULATION OF DYNAMICS OF FLUID POWER FEEDING SYSTEM WITH AN ACCUMULATOR

7.1. Simulation of simple feeding system
Simulation task of a simple feeding system with a pneumo-hydraulic accumulator for dynamic is shown in Figure 6.
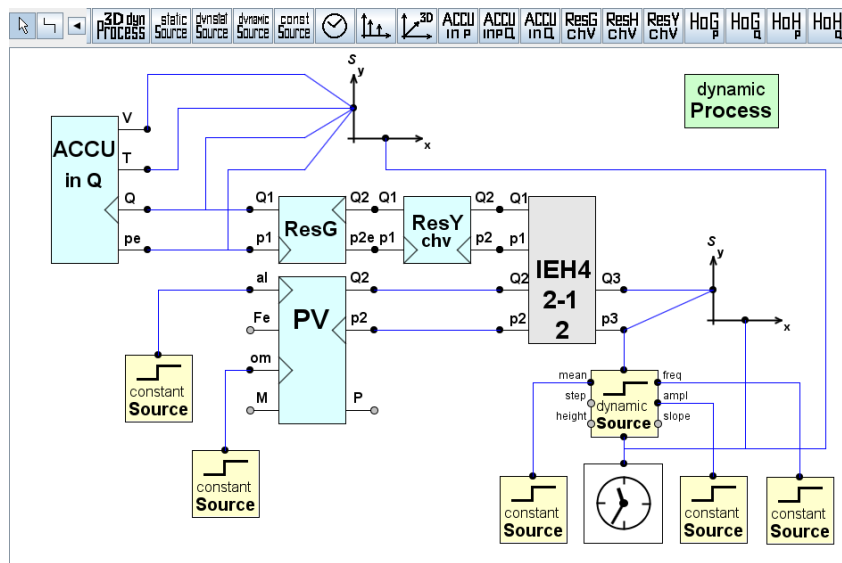
Figure 6: Simulation task of a simple feeding system with a pneumo-hydraulic accumulator for dynamic

In Figure 6 multi-pole models are as follows: accumulator ACCU_inQ, resistor ResG, resistor with check valve ResY_chv, hydraulic interface element IEH4_2-1_2 and variable displacement pump PV (reaction force Fe of the pump swash plate, reaction moment M of the pump and output power P of the pump are not used in the present task).

Dynamic input is denoted as "dynamic Source", constant inputs are denoted as "constant Source". Time is given by "Clock". Simulation process is managed by "dynamic Process". Parameters denoted by suffix "e" (pe in ACCU_inQ and p2e in ResG) are to be computed by using splitting and iterative calculation (see Chapter 4).

In all the examples below the following feeding system parameters are used.

For accumulator ACCU_inQ: p1 = 1e6 Pa, p2 = 5e6 Pa, V0 = 1e-3.

For resistor ResG: d = 0.01 m; for resistor ResY_chv: d = 0.003 m, μ = 0.7, Δp_n = 1e5 Pa, Q_n =2e-4 m³/s.

For hydraulic pump PV: V_{max} =10.027e-6 m³/rad, al_{max} = 0.3264 rad, al = 0.23 rad, ω = 154 rad/s, k_{vol} = 2e-9 1/Pa.

The following initial values are used.

For accumulator ACCU_inQ: init pe = 3e6, initV = 2.55e-4 m³, initT = 40 ºC.

For resistor ResG: initp2e = 3e6 Pa.

For resistor ResY_chv: initQ1 = 0.

For hydraulic pump PV: initQ =10.55e-4 m³/s.

Physical properties of working fluid (density ρ, kinematic viscosity v and coefficients of fluid compressibility) are calculated at each simulation step depending on average of input and output pressure in the component. In all the simulations hydraulic fluid HLP46 at temperature 40 ºC is used. Cinematic viscosity at temperature 40 ºC v = 46E–6 m²/s, density at temperature 15 ºC ρ_{15} = 875 kg/m³, volume of air, relative to the entire volume at p = 0, vol_0 = 0.08.

Simulation parameters: time step Δt = 1e-6 s, calculation step 4e5.

Simulated output volumetric flow of the feeding system in case of sinusoidal input pressure is shown in Figure 7.

Figure 7: Simulated output volumetric flow of the feeding system and sinusoidal input pressure

The sinusoidal input pressure (graph 1) parameters are: medial value 3e6 Pa, amplitude 5e5 Pa and frequency 10 Hz. The output volumetric flow (graph 2) is as a sum of pump flow and accumulator flow. The accumulator flow accounts the majority of the flow. So the change of output volumetric flow follows mainly the change of accumulator volumetric flow. In case of increasing input pressure the output volumetric flow drops. Charging of the accumulator occurs. Dropping the volumetric flow causes the fluid power system outlet velocity to decrease. In case of decreasing of input pressure the output volumetric flow increases. Discharging of the accumulator occurs with lower amplitude of output volumetric flow. As a result the accumulator works as an absorber of oscillations.

The simulated graphs showing the behaviour of accumulator variables are presented in Figure 8.

Accumulator gas volume (graph 1) is oscillating with phase shift to input pressure. The oscillations are asymmetric, they stabilize during 0.4 s.
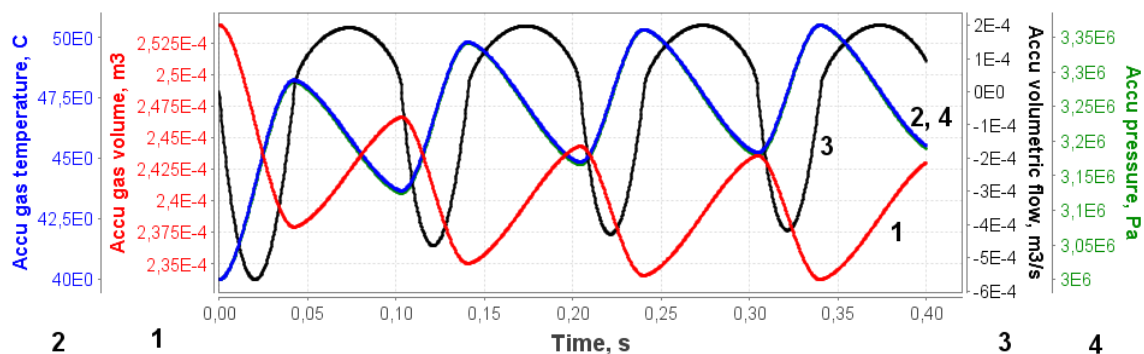
Figure 8: Simulated graphs of accumulator variables in case of sinusoidal input pressure

Graphs of accumulator gas temperature (graph 2) and accumulator pressure (graph 4) overlap. These oscillations are in opposite phase to gas volume oscillations at graph 1.

Oscillations of the accumulator volumetric flow (graph 3) have unsymmetrical amplitudes. At the negative volumetric flow charging and at the positive volumetric flow discharging of the accumulator occurs.

7.2. Simulation of feeding system with three-directional flow control valve

A simulation experiment was performed to demonstrate using an accumulator in a more complex hydraulic system. A hydraulic drive with three-directional flow control valve considered in (Harf and Grossschmidt 2015) was used as basis. A fragment of the drive equipped with an accumulator is shown in Figure 9. The pump **PV** is driven by electric motor **ME** through clutch **CJh**. The outlet of the pump is provided with three-directional flow regulating valve **FRV** and safety valve **SV**. The feeding system is supplemented with hydraulic accumulator **ACCU** together with hydraulic resistor **Res** and hydraulic resistor with check valve **Res_chv**.
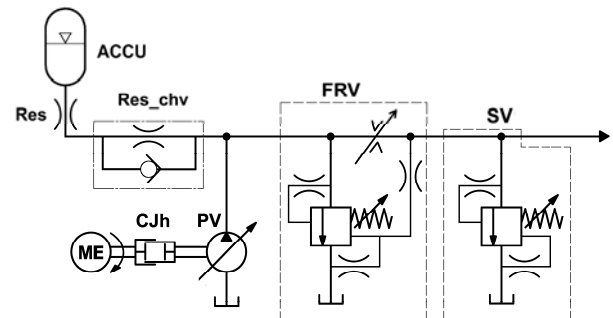
Figure 9: Functional scheme of a fluid power feeding system with an accumulator, three-directional flow control valve and safety valve

In Figure 10 multi-pole models are as follows: electric motor ME, clutch CJh, variable displacement pump PV, accumulator ACCU_inQ with resistor ResG and resistor with check valve ResY_chv, three-directional flow control valve FRV (pressure compensator spool VQAS22, pressure compensator slot RQHC, regulating throttle orifice ResYOrA, resistors ResG_Ch and ResH), safety valve SV (safety valve spool VS and throttle edge of safety valve spool RV) and interface elements IEH.
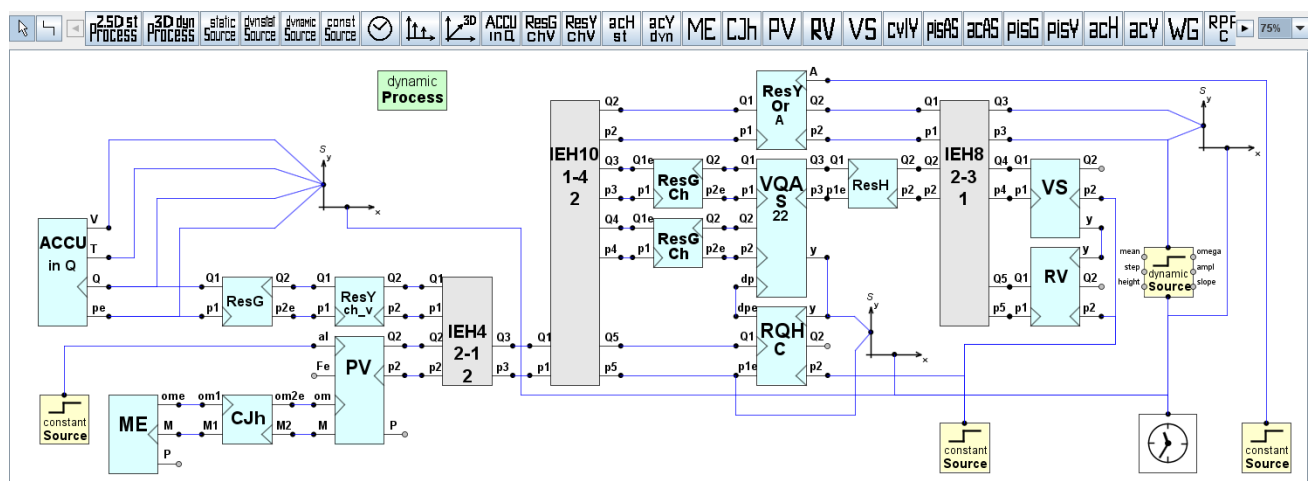
Figure 10: Simulation task of a feeding system with an accumulator

Simulation results are shown in Figures 11 ... 13.
The simulated graphs showing the behaviour of the accumulator are presented in Figure 11.
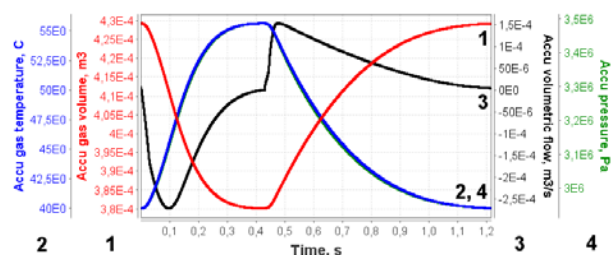
Figure 11: Simulated graphs of accumulator variables in case of impulse input pressure

Accumulator gas volume (graph 1) decreases when the input pressure impulse (graph 1 in Figure 13) stands at

maximum. Gas volume starts to increase when the impulse falls and achieves the initial value at 1.2 s. Graphs of accumulator gas temperature (graph 2) and accumulator pressure (graph 4) overlap. When the impulse height is achieved temperature and pressure increase. Temperature and pressure start to decrease when the impulse falls.

Accumulator volumetric flow (graph 3) decreases during the impulse rise. When the impulse height is achieved the output volumetric flow is going to restore the initial level. When the impulse falls, the output volumetric flow increases. After the impulse pressure reaches back to the baseline the output volumetric flow slowly decreases to the initial value at time 1.2 s.

Simulated graphs of three-directional flow control valve FRV are shown in Figure 12.
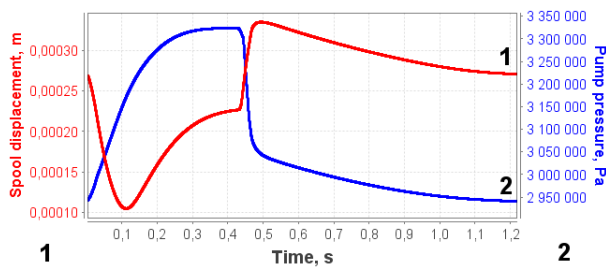
Figure 12: Simulated graphs of three-directional flow control valve

Spool displacement (graph 1) determines pump pressure (graph 2). Pump pressure is applied to accumulator.

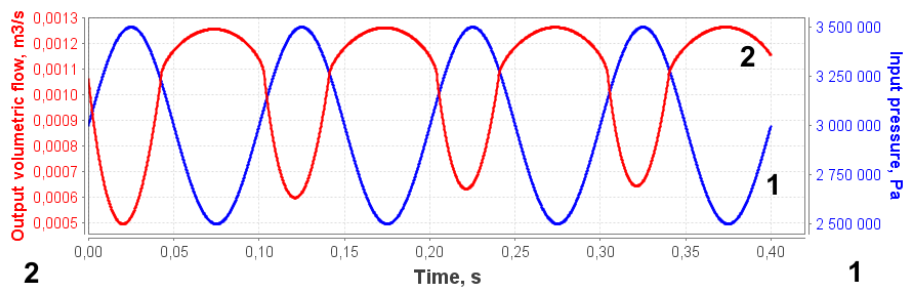Simulated output volumetric flow of the feeding system in case of impulse input pressure is shown in Figure 13.
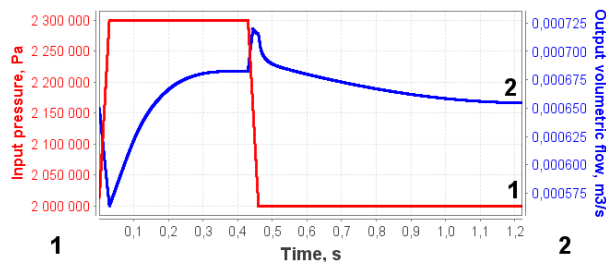
Figure 13: Simulated output volumetric flow of the feeding system in case of impulse input pressure

Impulse input pressure (graph 1) parameters are: impulse rising and falling time 0.03 s, impulse duration 0.4 s, baseline pressure 2e6 Pa and impulse height 3e5 Pa. Feeding system output volumetric flow (graph 2) decreases during the impulse rise. Charging of the accumulator occurs. When the impulse height is achieved the output volumetric flow is going to restore the initial level. When the impulse falls, the output volumetric flow jumps up. After the impulse pressure drops back to the baseline the output volumetric flow slowly decreases to the initial value. Charging of the accumulator occurs rapidly because the check valve of ResY_chv is opened. Discharging of the accumulator occurs slowly because the check valve of ResY_chv is closed.

CONCLUSION

In the paper multi-pole modelling and intelligent simulation of a fluid power feeding system with a three-directional flow control valve and pneumo-hydraulic accumulator has been considered. Mathematical multi-pole models of the system components (accumulator, inlet resistor, resistor with check valve, hydraulic pump) are described.

An intelligent simulation environment CoCoViLa supporting declarative programming in a high-level language and automatic program synthesis is used as a tool for modelling and simulation.

Visual simulation task of a feeding system with a pneumo-hydraulic accumulator for dynamic is presented.

Simulations have been performed and resulting graphs for cases of sinusoidal and impulse disturbances of the input pressure are presented. The graphs demonstrate

that using the observable feeding system smoothes volumetric flows to fluid power systems.

Using methodology and simulation system described in the paper enables to perform simulations with accumulators of different parameters. As a result accumulator with optimal parameters can be found for each particular fluid power feeding system.

The feeding subsystem of proposed configuration is usable for fluid power systems where pressure at pump depends on load (systems with three-directional flow regulating valves, load sensing systems, etc.).

ACKNOWLEDGEMENTS

REFERENCES

Barnwal, M. K., Kumar, N., Kumar, A. and Das, J. 2014. Effect of Hydraulic Accumulator on the System Parameters of an Open Loop Transmission System. 5th International & 26th All India Manufacturing Technology, Design and Research Conference (AIMTDR 2014), December 12–14, 2014, IIT Guwahati, Assam, India, 304-1 – 304-5.

Chrostowski, H. and Kedzia, K. 2004. The analysis of pneumo - hydraulic accumulator efficiency, applied as element of hybrid driving system. Scientific papers of the University of Pardubice, Series B, The Jan Perner Transport faculty 10.

Grossschmidt, G. and Harf, M. 2009. COCO-SIM - Object-oriented Multi-pole Modeling and Simulation Environment for Fluid Power Systems, Part 1: Fundamentals. International Journal of Fluid Power, 10(2), 2009, 91 - 100.

Grossschmidt, G. and Harf, M. 2010. Simulation of hydraulic circuits in an intelligent programming environment (Part 1, Part 2). 7th International DAAAM Baltic Conference "Industrial engineering", 22-24 April 2010, Tallinn, Estonia, 148 -161.

Grossschmidt, G. and Harf, M. 2014. Effective Modeling and Simulation of Complicated Fluid Power Systems. The 9th International Fluid Power Conference, 9. Ifk, March 24-26, 2014, Aachen, Germany, Proceedings Vol 2, 374-385.

Harf, M. and Grossschmidt, G. 2015. Multi-pole Modeling and Intelligent Simulation of a Hydraulic Drive with Three-directional Flow Regulating Valve. 10th International DAAAM Baltic Conference "Industrial engineering", 12-13 May 2015, Tallinn, Estonia, 27 -32.

Kotkas, V., Ojamaa A., Grigorenko P., Maigre R., Harf M. and Tyugu E. 2011. "CoCoViLa as a multifunctional simulation platvorm", In: SIMUTOOLS 2011 - 4th International ICST Conference on Simulation Tools and Techniques,

March 21-25, Barcelona, Spain: Brussels, ICST, 2011, 1-8.

Mamčic, S. and Bogdevičius, M. 2010. Simulation of dynamic processes in hydraulic accumulators. Transport, 2010, Vilnius, Lithuania, 25(2): 215–221.

Matskin, M. and Tyugu, E. 2001. Strategies of structural synthesis of programs and its extensions, Computing and Informatics, Vol. 20, 2001, 1-25.

Murrenhoff, H. 2005. Grundlagen der Fluidtechnik, Teil 1: Hydraulik, 4. neu überarbeitete Auflage. Institut für fluidtechnische Antriebe und Steuerungen, Aachen, 2005.

AUTHORS BIOGRAPHY

Mait Harf

Dipl. Eng. degree obtained from Tallinn University of Technology in 1974. Candidate of Technical Science degree (PhD) received from the Institute of Cybernetics, Tallinn in 1984. His research interests are concentrated around intelligent software design. He worked on methods for automatic (structural) synthesis of programs and their applications to knowledge based programming systems such as PRIZ, C-Priz, ExpertPriz, NUT and CoCoViLa.

Gunnar Grossschmidt

Dipl. Eng. degree obtained from Tallinn University of Technology in 1953. Candidate of Technical Science degree (PhD) received from the Kiev Polytechnic Institute in 1959. His research interests are concentrated around modelling and simulation of fluid power systems. His list of scientific publications contains 94 items. He has been lecturing at the Tallinn University of Technology 55 years, as Assistant, Lecturer, Associate Professor, Head of the chair of Machine Design and Senior Researcher.

A STOCHASTIC RISK ANALYSIS THROUGH MONTE CARLO SIMULATION APPLIED TO THE CONSTRUCTION PHASE OF A 600 MW GAS TURBINE PLANT

Fahimeh Allahi[(a)], Lucia Cassettari[(b)], Marco Mosca[(c)]

[(a),(b),(c)] DIME, Scuola Politecnica, University of Genoa
Genoa, ITALY

[(a)] allahi@ dime.unige.it, [(b)] cassettari@dime.unige.it, [(c)] marcotulliomosca@gmail.com

ABSTRACT
Construction projects are characterized by great uncertainty. Appropriate risk analysis techniques are required to estimate the adequate coverage level against the occurrence of extra costs to increase the progress of the project in the tenders. The project margin increases when an excessive provision leads to a more comprehensive coverage of the risks.

The purpose of this research is to apply an innovative analysis method based on Monte Carlo Simulation (MCS) to a real project to demonstrate the advantages of a study in a stochastic regime. The amount of contingency determined by the proposed approach is more accurate compared with the previous method used by the company. In the illustrated application, MCS has been applied even to the study of the work progress status.

Keywords: Contingency, Stochastic Risk Analysis, Monte Carlo Simulation, Construction Project.

1. INTRODUCTION
A risk is an uncertain event or condition that, if it occurs, has an effect on at least one of the project objectives. Contingency cost is the estimated amount of budget or time setting aside to cover the total risk of projects (Project Management Institute 2013; Association for Project Management 2008; Eldosouky, Ibrahim and Mohammed 2014). Fundamentally, contingency cost is an essential reservation for uncertainties in the projects (Thompson and Perry 1992) and it is demonstrated the total financial obligation for the project manager (Baccarini and Love 2014). The procedure of project risk management consists of identifying, quantitative and qualitative risk analysis, response planning and mitigating risks which are caused the successful of the project (Maytorena, Winch, Freeman and Kiely 2007). By applying a risk analysis method such as the Monte Carlo Simulation (MCS), PERT, Failure Models and Effect Analysis (FMEA), decision trees and sensitivity analysis (PMBOK® Guide 2013; Muriana and Vizzini 2017; Baccarini 2005; Bakhshi and Touran 2014), the estimated contingency amount can be obtained. As an example, while a project contains tasks with risk factors, PERT method can determine the risk of overcoming the

estimated time. Furthermore, the MCS has been applied in the study of the stochastic system behavior by means of its reproduction in a controllable environment. It is defined as a mathematical model that consists of the equations that describe the relationships between the components of the system parameters and their bond. The purpose is to verify the performance of the system in the face of certain inputs, to gather information on its output and make possible predictions. The MCS is used when it is not possible to analyze the object system analytically to make an estimate of the entire output probability distribution. The basic steps for the application of the Monte Carlo method can be described as follows:

• Identification of external factors
• Model definition
• Allocation of probability distributions
• Settings of the simulations and performance of the experiments
• Verification of results

This proposed methodology for risk management and project control allows working in a stochastic regime that increases the progress of the project.

The research is structured in the following steps: the next part illustrates a description of the company and the different phases of the project. In addition the MCS analysis is applied for the contingency provision. In the last part, the analysis and the obtained results from the application of the methodology to an installation of four gas turbines 600MW are described.

2. RISK ANALYSIS METHOD FOR THE CASE STUDY
The proposed method is applied in the EPC project (turnkey system) of an international Construction Company with over 30,000 employees in seven plants. The organizational structure of the Company is quite complex, considering the high number of employees, 2,937 units and the wide range of functions which require an ordering and management on many different correlated levels. In the recent years, the Company has taken steps to strengthen the operational methodologies and tools to support the management. The EPC contract

had involved the design and construction of a power plant in open cycle "turnkey" 600 MW in Egypt. The control unit is composed of four equipped units with four gas turbines that are totally designed and constructed by the Company.

The Company usually carried out a Risk Analysis so structured:

1. Identification of the activities needed to complete the job order, thus creating a list of tasks and their dependencies (prior and subsequent activities) and the necessary resources (in the construction phase are the work hours to complete that particular task);

2. Analysis of environmental conditions that may affect the site's activities (socio-political situation, type of customer, logistical constraints, local workforce specialization);

3. Mitigation of the risk of delayed timing and consequently the risk of overcoming the cost of the site budget by increasing the percentage of time and therefore the resources planned in the initial ideal program. The ideal program is modified by finding suitable multipliers "K" that vary according to the environmental analysis result. This step allowed the Company to have a more realistic forecast (Figure 1).
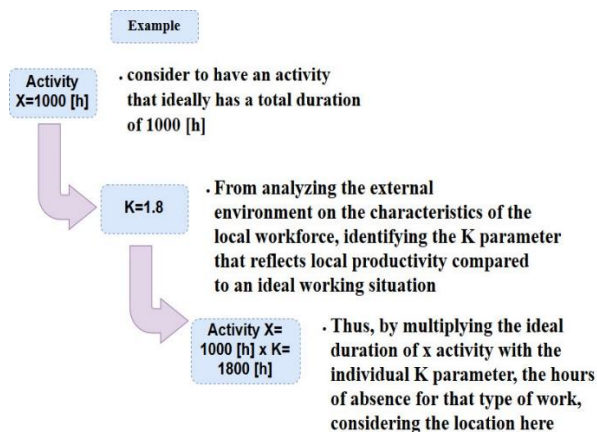
Figure 1: Example of the Company's Risk Management approach

The Authors applied MCS to manage the phase of development of the project, which is stochastic, with the aim of reducing the risks of delay of the contractual delivery date. In particular, risk analysis has been applied to two different phases of the project:

1. The phase of bid
2. The work progress control

2.1. The Phase of Bid

This phase is a fundamental activity to estimate the price of the order to success in the project. It is important to note that the methodology will be applied only to the construction phase, which it is allocated about 30% of the total project budget.

In the following, the main steps of the methodology for the bid phase, entitle:

• Identification of necessary data to the model;
• Applying Monte Carlo analysis for the contingency provision.

2.1.1. Identification of Necessary Data to the Model

When the total required budget for the construction phase is determined, it should be split into the different program activities. Allocation budget to the individual activities is done by taking into account the specific characteristics of each activity such as the duration, the type of processing, the fixed costs and the variable costs. Table 1 presents the identified weight percentage of the total required budget for each activity.

Table 1: Weight Percentage of the Total Required Budget for Some Tasks

Task	Weight	Task	Weight
Excavation for turbine Hall unit 1-4	0.86%	Fuel oil plumps shelter install.	0.81%
Foundation Hall including half of unit 1-4	1.95%	Compressor install	1.00%
Stack unit 1-4 foundations	0.59%	Fogging system install	1.38%
G.T. 1-4 foundations	3.11 %	Underground instrument	7.36%
Steel struct. turbine hall 1-4	2.05%	Cable ways and install	2.93%
Local control system (half) elevation 1-4	1.70%	Yard area and finishing HVAC	8.05%
Main transformer G.T. 1-4 foundations	3.42%	GIS subs. and cable connection	2.31%

In order to obtain the necessary input data for the simulator and acquire an accurate estimation, it has been switched from a deterministic analysis to a stochastic one.

Each activity was then associated with a probability distribution taking into account both the optimistic case, in which the allocated budget is not fully spent, and the bad case, where the execution of the activity requires the allocation of an extra budget .

The most suitable probability density function starting from a 3-point-estimate containing minimum value, maximum and most likely value, is the Triangular Distribution. Therefore, each activity is then assigned with a triangular distribution and the Table 2 is obtained.

Table 2: Applying the Triangular Distribution to Each Task

Weight of single task	Triangular Distribution		
	Min	Real	Max
0.86%	0.77%	0.86%	1.16%
1.95%	1.76%	1.95%	2.63%
0.59%	0.53%	0.59%	0.80%
3.11 %	2.80%	3.11%	4.20%
2.05%	1.85%	2.05%	2.77%
1.70%	1.53%	1.70%	2.30%
3.42%	3.08%	3.42%	4.62%
1.38%	1.24%	1.38%	1.86%
6.04%	5.44%	6.04%	8.23%
2.39%	2.15%	2.39%	3.23%
2.95%	2.66%	2.95%	3.98%
2.21%	1.99%	2.21%	2.98%
1.28%	1.15%	1.28%	1.73%
1.97%	1.67%	1.97%	2.71%

2.1.2. Applying Monte Carlo Analysis for the Provision Contingency

By using the described input data, it is possible to apply the Monte Carlo method. In addition, to apply a number of experimental runs on the model to obtain the valid results, the method of the Mean Square Pure Error (MSPE) in the repeated run (Mosca, Bruzzone and Cassettari 2009; Cassettari, Giribone and Mosca 2010; Cassettari, Mosca and Revetria 2012; Mosca, Giribone, Revetria, Cassettari and Cipollina 2008) should be done. Furthermore, Figure 2 presents the MSPE curves necessary to identify the sample size in order to obtain the statistical stabilization both of Mean Square Pure Error of the Mean (MSPEMED) and Mean Square Pure Error of the Standard Deviation (MSPESTDEV). It occurs at around 1000 runs. Therefore, the MCS results obtained using @RISK with the features of 5 reps and 10,000 runs are shown in Figure 3.
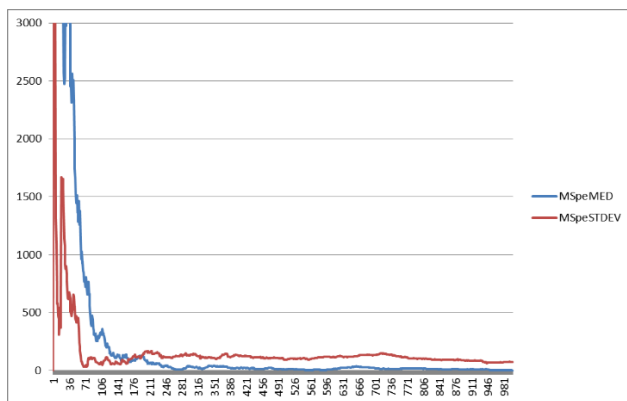
Figure 2: MSPE Curve for the Input Data
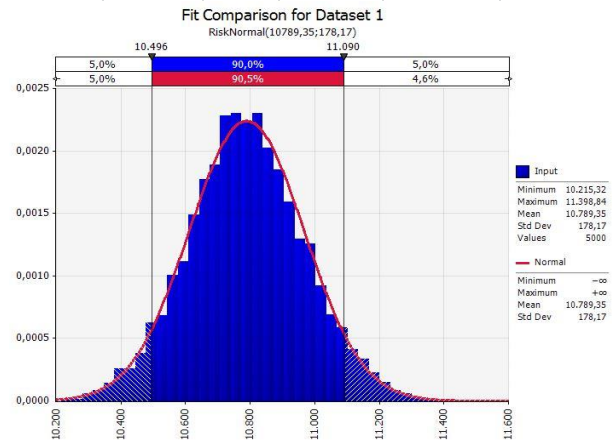
Figure 3: Monte Carlo Simulation by @ Risk Sofware

The obtained probability distribution curve covers a range of between € 10.2mn and € 11.6mn. In order to have an 80% coverage probability, the Company should therefore allocate an amount of not less than 11m euros, representing about 8% of the total value of the order.

2.2. The Phase of the Work in Progress Control

The main risk of this phase are the delays that may be the result of various reasons. Following are cited some of the main causes involving delays in the pipeline:

- Equipment failures;
- Errors on the part of employees;
- Weather conditions;
- Delays in the procurement of materials.

The objective of the proposed risk analysis to the phase of the progress management is to identify the likelihood of unexpected upstream in order to have the possibility to make changes to the program and the construction budget to complete the project within the deadline. This is crucial as the excess of the end date of the project involves huge penalties from the customer.

The main steps of the applied methodology to the phase of the advancements management are:

- Study of the Construction program and data identification
- Identification of the critical path

2.2.1. Study of Construction Program and Data Identification

To define the necessary input date for the simulator, each activity should be associated with a real deterministic time. As mentioned above the Company was used to augment the actual time with an incremental time by means of a standard percentage of increase K. Therefore, the first step was to eliminate the effect of the coefficient K and consequently to identify the most likely duration "TM" (an average duration that does not take account of external factors (Figure 4)).

Once the average time is obtained it is possible to transform the duration of each activity from deterministic to stochastic. The next step is to decide what type of probability distribution to use. It has opted for a non-symmetrical triangular distribution for all activities.

However, the variability of the duration of the activity was differentiated according to the characteristics of the project task.

In particular, for all civil works to take into account the impacting weather variability on outdoor works, it is considered intervals as follows:

Max duration = TM * 1.4
Min duration = TM * 0.8

As for the electro-mechanical assemblies, the following extreme values are considered:

Max duration = TM * 1.2
Min duration = TM * 0.9

6	G.T.1-2-3-4 Foundations	93
7	Foundation Hall including mesa unit 1-2-3-4	119
8	Steel Struct. Turbine hall 1-2-3-4	90
9	Local control system (mesa) elevation 1-2-3-4	120
10	Stack Unit 1-2-3-4 foundations	96
11	Main Traformers foundation G.T. 1-2-3-4	138
12	Commons Civil Works	341g
13	Common Electrical control building	221
14	1° Access for Cable Ways install	0
15	Electrical Substation GIS	187
16	GIS Control Bldg	105

Figure 4: Example of the Most Likely Duration Estimation

2.3. Applying Monte Carlo Analysis to Predict the Final Date of the Project

In this phase, the described input data and MCS should apply in the @RISK software with features of 5 reps and 1,000 runs corresponding to the MSPE curves in Figure 5. The MCS has been applied to each of the four critical paths (one for each gas turbine) in order to evaluate the duration of the four units construction.

The MCS risk analysis has been repeated at 5 different instants of time in order to take into account during the progress of the project of the activities which already completed. In addition it recalculates with an increasing level of reliability of the expected date of delivery of the four gas turbines.

- T = 0: from June 1, 2016
- T = 1: from August 31, 2016
- T = 2: from September 30, 2016
- T = 3: from October 31, 2016
- T = 4: from November 30, 2016

Some of the MSPE Curves and MCS results on the different critical path associated with these instants of time are reported in the following.
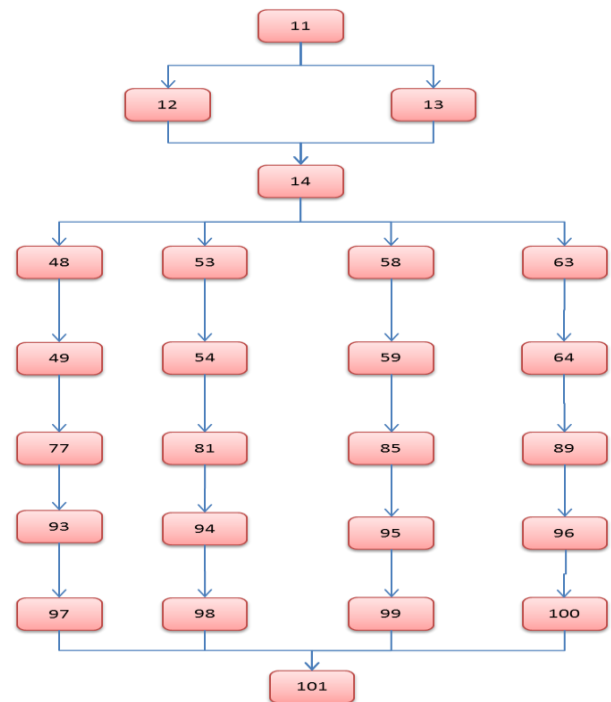
Figure 5: Critical Paths for Each Gas Turbine

Critical path 1 (T=0)
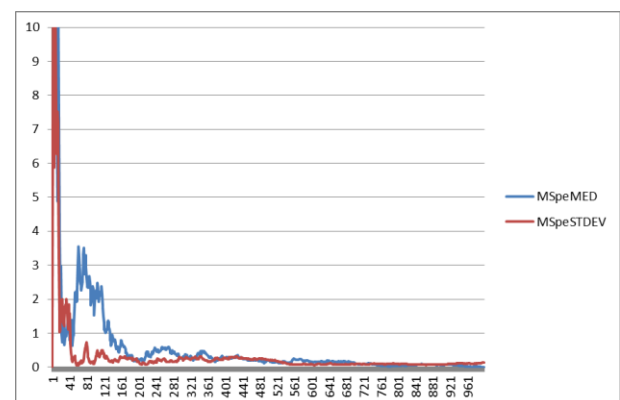
Figure 6: MSPE Curve for Critical Path 1 (T=0)

From T = 0 to T = 1, which is from June to August 2016, the first civil works were completed. There were no major problems or delays and some of activities had been concluded in advance with a positive impact on the overall project duration. In Figure 6 and Figure 7 the results of MSPE and Monte Carlo analysis for Critical path 1 for T = 0 and Figure 8 and Figure 9 for critical path=1 for T=1 are illustrated.

Figure 7: Monte Carlo Simulation for Critical path 1 (T=0)

Figure 9: Monte Carlo Simulation for Critical Path 1 (T=1)

Figure 8: MSPE Curve for Critical Path 1 (T=1)
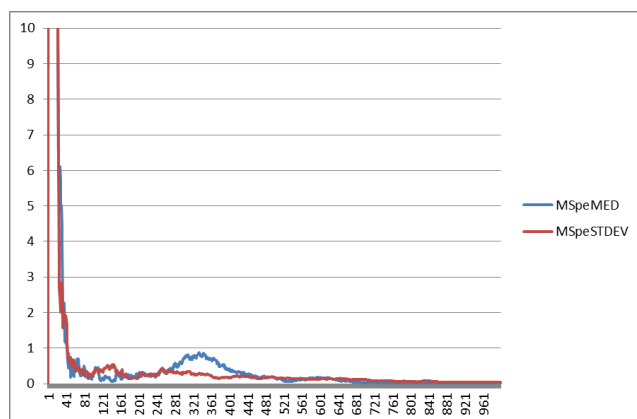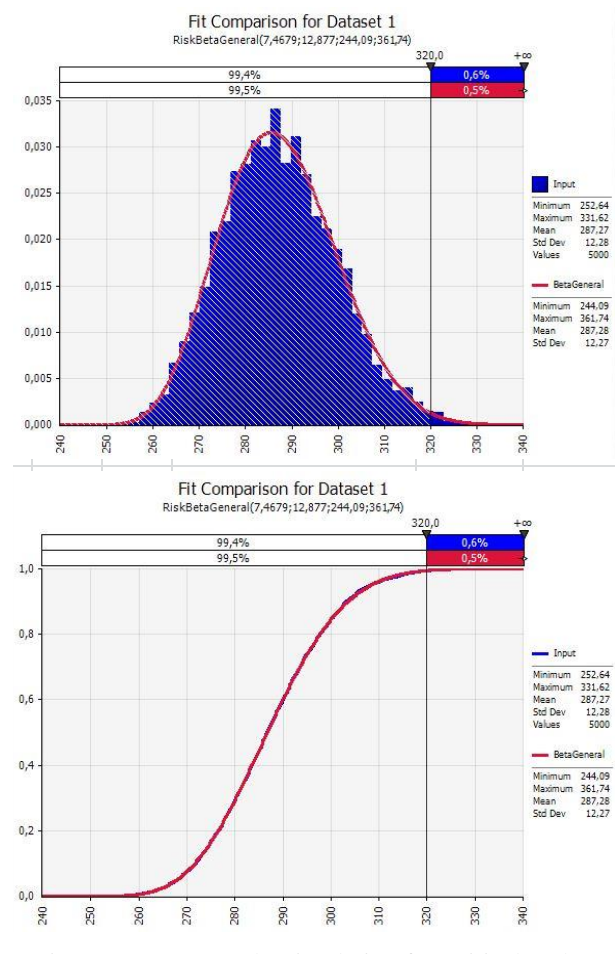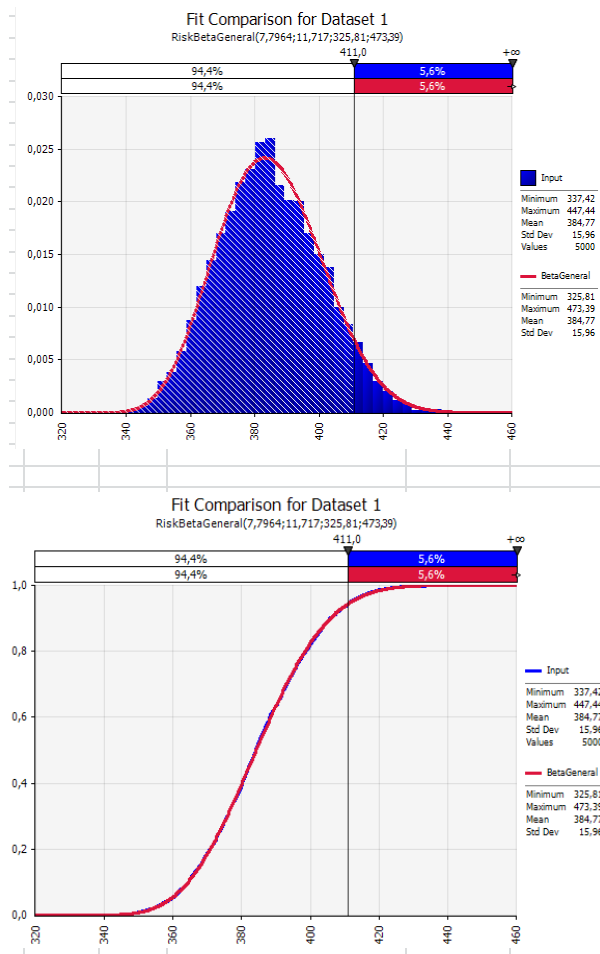
From the instant T = 1 to the instant T = 2, i.e. from September to October 2016, the realization of civil works continued. There were no particular problems or delays during the implementation of the program activities.

From the instant T = 2 to T = 3, i.e. from October to November 2016, almost all civil works were completed without any unexpected details, but the obtained time advantage in the first phase of the order had a slight decrease, as shown in the curve of project total cost (Figure 10).

Figure 10: Project Total Cost
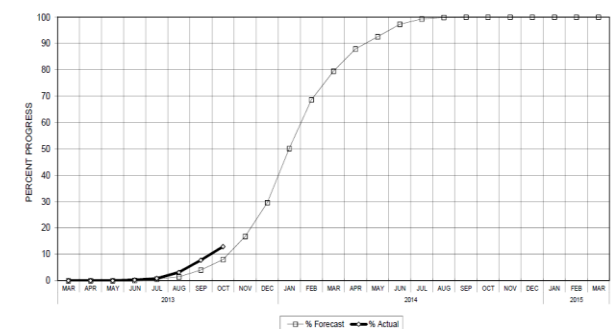
From time T = 3 to T = 4 i.e. from November to December 2016 continued the civil works and construction of steel structures of the four gas turbines has started. Although the overall situation of the construction phase has to be in advance of the program yet, the previously accumulated advantage has been greatly reduced.

The progress in the pipeline and the total order curves progress updated at T=4 are illustrated in Figure 11.
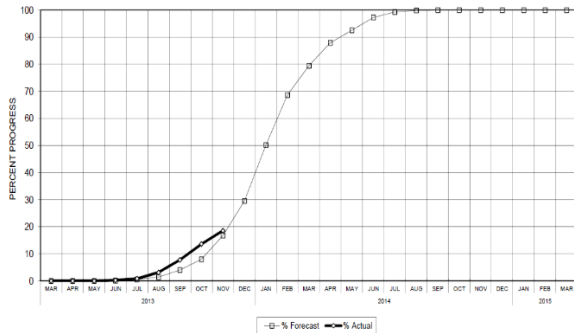
Figure 11: Total Order Curve of Progress

The objective of the proposed risk analysis was to identify for each critical path and each instant of time, the probable date of delivery of the plant and also the respective percentages of risk that the effective date would exceed.

In Particular, the obtained data from the carried out analysis are:

- ΔT: this data indicates the duration in days that divides the current date and the delivery date of the plant (for example, if you consider the time instant T = 2 of the first critical path the ΔT will equal 290 because the days are passing from late September 2016 to August 2017).
- Variability Range

This data indicates the variability of project final date determined by the analysis:

• Optimistic, realistic and pessimistic time
• P(x) delivery on time
• P(x) delayed delivery

Table 3 summarizes all the results just described.

At this point to understand when it would be appropriate to make changes to the program it is necessary to identify an additional element named P(x) threshold.

The P(x) introduces the probability of delay threshold that shows whether or not to make changes to the program of activities. For instance when P(x) threshold > P(x) delayed delivery, the advancements in the pipeline are under control then it will not allowed to make changes to the program. In addition, when P(x) threshold < P(x) delayed delivery, the construction site of the advancements are having significant delays so it is appropriate to begin to change the program by allocating additional resources to make up for lost time.

Table 3: Summary of Progress Simulation Results

Summary of Progress Simulation Results									
ID	Instant time T	N critical path	ΔT in days: Data TOAC - Instantaneous advancement	Variable Range	Optimistic T	Realistic T	Pessimistic T	P(x) delivery on time	P(x) delayed delivery
01	0	1	411	48	337	385	433	94.40%	5.60%
02	0	2	421	49	353	402	451	87.60%	12.40%
03	0	3	432	50	362	412	462	87.40%	12.60%
04	0	4	442	50	376	426	476	81.60%	18.40%
11	1	1	320	47	240	287	334	99.40%	0.60%
12	1	2	330	48	256	304	352	96.90%	3.10%
13	1	3	341	48	266	314	362	97.70%	2.30%
14	1	4	351	50	277	327	377	95.70%	4.30%
21	2	1	290	33	231	264	297	98.60%	1.40%
22	2	2	300	36	245	281	317	93.90%	6.10%
23	2	3	311	36	255	291	327	93.80%	6.20%
24	2	4	321	36	268	304	340	91.40%	8.60%
31	3	1	259	30	204	234	264	99.30%	0.70%
32	3	2	269	33	218	251	284	94.00%	6.00%
33	3	3	280	33	228	261	294	94.10%	5.90%
34	3	4	290	34	251	285	319	92.10%	7.90%
41	4	1	229	26	182	208	234	98.50%	1.50%
42	4	2	239	29	196	225	254	90.90%	9.10%
43	4	3	250	29	206	235	264	89.50%	10.50%
44	4	4	260	30	218	248	278	87.10%	12.90%

The Figure 12 illustrates an example comparing for each instant of time and any critical path with the threshold P(x) delayed delivery:
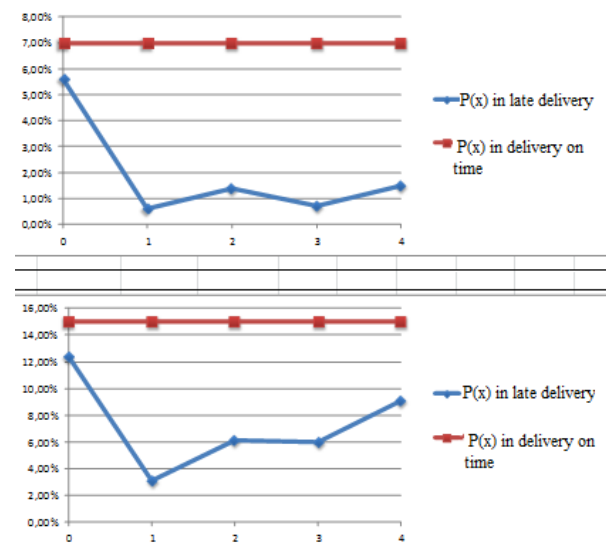
Figure 12: Comparison with the Threshold P(x) for Critical Path 1 (top chart) and Critical Path 2 (bottom chart)

As can be seen from the charts, for the analyzed time period it was not necessary to make changes to the program for any critical path at each instant of time. So it has been concluded that the advances in the pipeline in the period from June 2016 to December 2016 mirrored the predictions made at the

Finally, by the Figure 13, it is understood that how the variability of results decreases with the approaching of the delivery dates. This is very important because it allows the passing of time to identify more accurately the project end date.
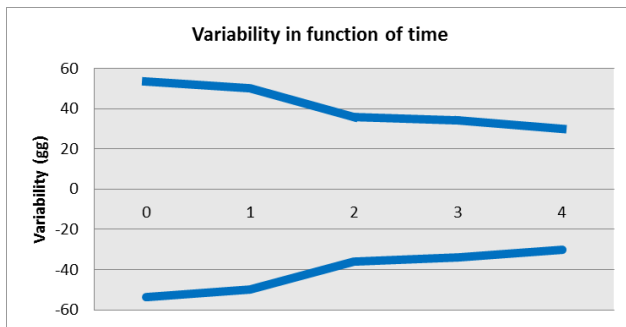
Figure 13: Range of Variability

3. CONCLUSION

The proposed study illustrate the application of a stochastic risk analysis based on Monte Carlo method to a real case study. The study aims to highlight the benefits and results obtained through a stochastic analysis compared to traditional deterministic analysis.

In Particular, two project phases were considered in this research: the process of bid and the phase of work in progress control; Risk Analysis with Monte Carlo method has been applied to both.

The analysis on the first phase has led to the allocation of a contingency equal to 8% of the costs of construction of the plant; this percentage represents a quantitative estimate of all the possible risks that may occur in the pipeline.

The MCS analyzes in the progress phase identified the project final dates for each critical path calculated at five different time instants, from June 2016 to December 2016; This has allowed to verify the evolution of the program and modify the program to avoid penalties. Moreover, it can be understand from the months analyzed showed that was not necessary to change the program of activities.

Finally, it has been possible to note how the variability of the results decreases with the approaching of the implant delivery dates; this aspect is fundamental in order to identify more precisely the final date of the project over time.

REFERENCES

Allahi F., Cassettari L., Mosca M., 2017. Stochastic Risk Analysis and Cost Contingency Allocation Approach for Construction Projects Applying Monte Carlo Simulation. Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2017, 5-7 July, 2017, London, U.K., pp 385-391.

Association for Project Management, 2008. Interfacing Risk and Earned Value Management. Association for Project Management, UK.

Baccarini D., 2005. Estimating Project Cost Contingency – Beyond the 10% syndrome. Proceedings of Australian Institute of Project Management National Conference. Nov 9th, Victoria (Australian Institute of Project Management, Australia).

Baccarini D., Love P.E.D., 2014. Statistical Characteristics of Cost Contingency in Water Infrastructure Projects. Journal of Construction Engineering and Management, 140(3), 1-9.

Bakhshi P., Touran A., 2014. An overview of budget contingency calculation methods in construction industry. Procedia Engineering 85, 52-60.

Cassettari L., Giribone P.G., Mosca M., Mosca R., 2010. The stochastic analysis of investments in industrial plants by simulation models with control of experimental error: Theory and application to a real business case. Applied Mathematical Sciences, 4(76), 3823–3840.

Cassettari L., Mosca R., Revetria R., 2012. Monte Carlo Simulation Models Evolving in Replicated Runs: A Methodology to Choose the Optimal Experimental Sample Size. Mathematical Problems in Engineering, 2012, 0-17.

Eldosouky I.A., Ibrahim A.H., Mohammed H.E., 2014. Management of construction cost contingency covering upside and downside risks. Alexandria Engineering, 53(4), 863–881.

Maytorena E., Winch G.M., Freeman J., Kiely T., 2007. The influence of experience and information search styles on project risk identification performance. IEEE Trans. Engineering Management, 54 (2), 315–326.

Mosca R., Giribone P., Revetria R., Cassettari L., Cipollina S., 2008. Theoretical Development and Applications of the MSPE Methodology in Discrete and Stochastic Simulation Models Evolving in Replicated Runs. Journal of Engineering, Computing and Architecture, 2(1), 1934-7197.

Mosca R., Bruzzone G.A., Cassettari L., Mosca M., 2009. Risk analysis for industrial plants projects: An innovative approach based on simulation techniques with experimental error control. Proceedings of 21st European Modelling and Simulation Symposium, EMSS 2009.

Muriana C., Vizzini G., 2017. Project risk management: A deterministic quantitative technique for assessment and mitigation. International Journal of Project Management, .doi: 10.1016/j.ijproman.2017.01.010, (In Press, Corrected Proof).

Project Management Institute (PMI), 2013. A guide to the Project Management Body of Knowledge: (PMBOK® Guide). Fifth ed. PMI, Project Management Institute, Newtown Square, PA.

Thompson P.A., Perry J.G., 1992. Engineering construction risks. London, Thomas Telford.Mercer P.A. and Smith G., 1993. Private view data in the UK. 2nd ed. London: Longman.

AUTHORS BIOGRAPHY

Fahimeh Allahi is a Ph.D. student in Industrial Engineering at the University of Genoa- Department of Mechanical Engineering, Production Engineering, and Mathematical Modelling (DIME). Her research field is focused on risk management and Derivate based on Monte Carlo Simulation.

Lucia Cassettari earned her degree in Management Engineering in 2004 at the University of Genoa. She obtained

the title of Ph.D. in Mathematical Engineering and Simulation in 2009. She is actually a researcher in University of Genoa-Department of Mechanical Engineering, Production Engineering, and Mathematical Modelling (DIME). Her research field is focused on simulator-based applications for industrial plants with particular attention on the application of DOE and Optimization techniques to industrial plant problems using Simulation. Her main experience is in research projects involving BPR (Business Process Reengineering), Activity Based Management, Cost Control, Logistics, and Decision Support Systems. Her current teaching activity is in the Industrial Plant Management and Planning & Control degree courses.

Marco Mosca earned his degree in Management Engineering in 1999 at the University of Genoa. He obtained the title of Ph.D. in Innovation Management in 2011. He has more than 10 years of working experience: consultant in Accenture, manager in Bombardier Group working in Italy, Switzerland, Austria and Germany and Managing Director of an Italian Manufacturing company. He is now Visiting Professor at Polytechnic School of University of Genoa. His current teaching activity is in the Operations Management degree course.

SIMULATION-BASED OPTIMIZATION OF A FOUR STAGE HYBRID FLOW SHOP WITH SEQUENCE-DEPENDENT SETUP TIMES AND AVAILABILITY CONSTRAINTS

Paul Aurich[(a)], Abdulrahman Nahhas[(b)], Tobias Reggelin[(c)], Marco Krist[(d)]

[(a),(c)] Department of Logistics and Material Handling Systems, Otto-von-Guericke University Magdeburg, Germany
[(b)] Magdeburg Research and Competence Cluster for Very Large Business Applications, Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany
[(d)] Tectron GmbH, Germany

[(a)]paul.aurich@mail.de, [(b)]abdulrahman.nahhas@ovgu.de, [(c)]tobias.reggelin@ovgu.de
[(d)]marco.krist@thermofisher.com

ABSTRACT
In this paper, a deterministic four stage hybrid-flow-shop scheduling problem with sequence-dependent setup times of a printed circuit board assembly is discussed. Since the problem came from an industrial company, availability constraints have been taken into consideration. There are many reasons which can cause an unavailability of machines. In this article stochastic breakdowns and preventive maintenance were considered. Furthermore, deterministic breakdowns will be investigated, in order to compare deterministic and stochastic breakdowns in terms of the robustness and the stability of the solution and the required computational effort. The objective of the problem is to minimize the makespan, the total tardiness and the total setup time of the first stage. To generate an optimized production schedule, the metaheuristics simulated annealing, tabu search and differential evolution as well as a sequencing algorithm were combined with a discrete event simulation model. This paper is a continuation and extension of our previous work (Aurich et al. 2016; Nahhas et al. 2016).

Keywords: metaheuristic, simulation-based optimization, hybrid flow shop, breakdown, maintenance

1. INTRODUCTION & LITERATURE REVIEW
The development of digitization over the last decades and their connection with technical systems leads to the possibility to record the process data of a production system at any moment. This makes it possible to react faster to unforeseen events. In contrast to this, in most of the publications which were discussed in the field of scheduling, availability constraints were not taken into consideration. But real production systems are always subject to interruptions. Based on the data, which now can be detected, it is possible to define preventive maintenance times more precisely and notice breakdowns quickly. Therefore availability constraints should be taken into account.
Two types of processing cases are differed in the literature, when a machine becomes available again after an interruption. The first type according to (Lee 1996) is called *resumable*, in which the processing of a job can be continued without any loss in time. In opposite to that it is called *non-resumable*, if a restart of the processing from the beginning is necessary, which propagate a loss of time. (Saidy et al. 2008) surveyed a wide range of typical scheduling problems with the addition of resumable and non-resumable availability constraints.
Considering stochastic breakdowns, makes it difficult to create a mathematical model, because of the structural and functional complexity. However, simulation techniques are often used as an alternative solution. In addition, using simulation techniques the modeler is capable of mimicking the exact behavior of a considered system. According to (März et al. 2011), four possibilities of combining simulation with optimization techniques can be differentiated:

- Optimization is integrated into the simulation
- Simulation as evaluation function of optimization
- Simulation results as initial value of the optimization
- Optimization results for configuring the simulation

Since this article focusses on a hybrid flow shop (HFS) scheduling problem the literature review is concentrated on parallel machine (P), flow shop (F) and hybrid flow shop scheduling problems. A hybrid flow shop production environment consists of k stages in series. Each production stage comprises j parallel machines. Each job i should be processed in all stages and each job can be processed by any machine of a stage (Ruiz and Vázquez-Rodríguez 2010; Pinedo 2012).
The problem $F2|.|C_{max}$ is the only flow shop problem which can be solved in polynomial time. (Johnson 1954) developed the so called Johnson Rule to solve this problem. However, already the problem $F3|.|C_{max}$ is NP-hard in a strong sense (Garey and Johnson 1979). Also the problem $HFS2(P)|.|C_{max}$ is NP-hard in a strong sense, even if there is only one machine in the first stage und two parallel machines on the second stage, i.e. problem $HFS2(1, P2)|.|C_{max}$ studied by (Hoogeveen et al. 1996). Several other cases of the two stage HFS were

studied in the literature (Gupta 1988; Li 1997; Allaoui and Artiba 2006).

(Allaoui and Artiba 2004) considered different HFS with four stages, a maximum number of five machines at each stage and 50 jobs. Furthermore resumable and non-resumable availability constraints were taken into consideration. They investigated the impact of the initial schedule on simulated annealing with different objectives, such as minimization of the makespan, total completion time, mean flow time, mean waiting time, mean tardiness and maximum tardiness. To generate the initial solutions the SPT, LPT and EDD dispatching rules were used. (Gholami et al. 2009) developed a simulation-based optimization approach to solve several HFS problems with sequence-depend setup times and stochastic breakdowns. Resumable processing was considered. He adapted the random key genetic algorithm to build a schedule for the first production stage. To assign the jobs for the following stages the SPT cycling heuristic and a Johnson-Rule-based heuristic were implemented. The simulation is used as an evaluation function of the optimization. (Gholami et al. 2009) noticed that the first available machine rule would not be efficient, if sequence-dependent setup times have to be considered at all stages.

2. PROBLEM DESCRIPTION

The considered deterministic scheduling problem can be described as a four stage hybrid flow shop scheduling problem. In the classical hybrid flow shop all jobs have to be processed on one machine at each stage. In contrast to this the considered HFS is as special form, where all jobs have to be processed on the first and second stage, but only specific jobs have to be processed on the third or fourth stage or on each of them.

The machine environment consists of four stages (see Figure 1). The first stage contains four identical parallel surface mount device placement machines (SMD). These are the critical resources in the considered production system. The second stage accommodates five identical parallel automated optical inspection machines (AOI). The further stages contain only one machine each, a selective soldering machine (SS) in the third stage and a coating machine (CM) in the fourth stage. In the first and fourth stage jobs are scheduled with sequence-dependent major and minor setup times.

(Tang 1990) and (Wittrock 1990) introduced the concept of major and minor setup times. They both investigated parallel machines scheduling problems, where several jobs could be grouped into different families depending on their part-types. In the first stage the setup time s_f depends on family-type f of a job. The jobs are cluster into families based on their raw materials. In the fourth stage the setup time depends on the coating-type c of job. The company uses two different coating-types. In the second and third stage jobs are scheduled with sequence-independent setup times.

The job nature and the considered restrictions can be described with the following assumptions:

- The number of jobs in a certain scheduling period is fixed.
- The processing time $p_{i,j,k}$ of each job i on machine j of stage k is known and fixed.
- The priority of a job represents it´s desired delivery date d_i.
- It is not allowed to process jobs from the same family-type on different machines of the first stage at the same time.
- The buffer size between two stages is unlimited.
- Preemption and splitting of jobs are not allowed.

Unlike most of the papers which deal with scheduling problems, this paper takes availability constraints into consideration. More specific preventive maintenance and stochastic breakdowns can affect a machine during or not during the processing of a job. When the machine becomes available after an interruption, the processing of a job continuous without any loss in time, consequently it is resumable rs.

The objective functions of the analyzed HFS are to minimize the makespan C_{max} (1), to minimize the total tardiness T (2) and to minimize the total setup time of the first stage $\sum s_f$.

$$C_{max} = \max C_i \ \forall \ i = 1, \ ..., \ n \tag{1}$$

$$T = \sum_{i=1}^{n} T_i \ , \ T_i = \max(C_i - d_i, \ 0) \tag{2}$$

According to the classification of (Graham et al. 1979) the problem can be described with:

HSF4 (IP4, IP5, 1, 1)$|s_f$, s_c , $rs|C_{max}$, $T, \sum s_f$.
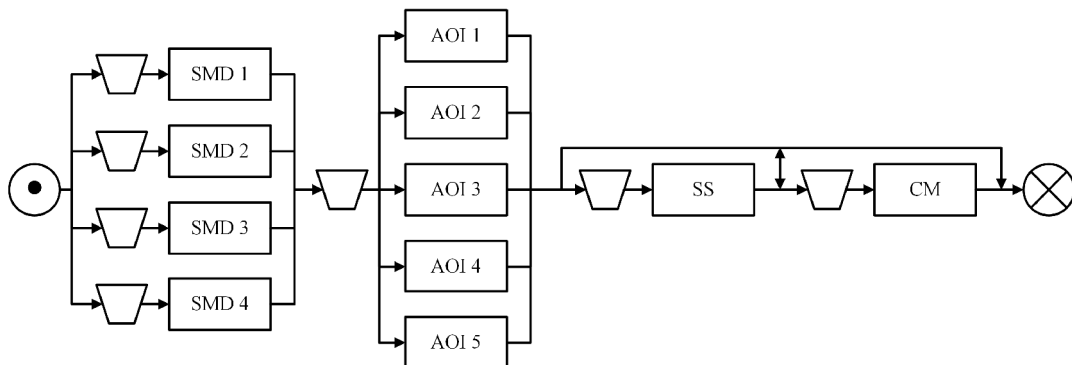
Figure 1: Production System

3. SOLUTION APPROACHES

Since the considered problem is NP-hard, it is not possible to develop a polynomial algorithm, which can provide an optimal solution in a reasonable time. In order to reduce the complexity of the problem the allocation and sequencing decisions of the first stage are separated from each other. On the following stages the jobs are assigned to the machines according to the first available machine (FAM) rule.

To solve the allocation problem of the first stage the metaheuristics simulated annealing, tabu search and differential evolution were implemented. To deal with the four independent single machine problems with sequence-dependent setup times in the first stage a sequencing algorithm was developed. The job sequence for the stages two to four is generated based on the earliest due date rule. The implementation of all metaheuristics, dispatching rules and the sequencing algorithm is done inside the simulation model in order to avoid an increasing computational time because of the data exchange between optimization and simulation tool. The simulation model was implemented in ExtendSim 9.1. The discrete event simulation approach is adapted to build the simulation model, in which each job is aggregated into a single object The combination of optimization and simulation model cannot be classified based on the classification of (März et al. 2011). This is because two methods of combination were used. On one side the optimization is integrated inside the simulation. On the other side the simulation is an evaluation function for the metaheuristics.

3.1. Initialization

Before the optimization takes place the user must insert a dataset into an excel document. Moreover, the user must decide which metaheuristic should be used and accordingly the control parameter must be setup. Then an initial allocation of families is executed in excel. This can be done randomly or with some easy sorting rules, for instance sorting the families based on the number of jobs with the same family-type or the total processing time of a family on the first stage. Hereafter all informations were send to the simulation model and a single simulation run is executed to measure the objective values of the initial allocation.

3.2. Functionality of the metaheuristics

Depending on the control parameters of the used metaheuristic a multi run simulation is setup. The execution of the metaheuristic takes place at the end of each simulation run. Based on the decision strategy of the metaheuristic it is decide if the current solution is used for the next iteration.

3.2.1. Simulated Annealing

Simulated Annealing (SA) is a nature inspired optimization technique, mimicking a thermodynamically cooling process. It was first introduced by (Kirkpatrick et al. 1983) and (Černý 1985) to solve the traveling salesman problem. The neighborhood search (NHS) of the adapted SA is a random single point operator NHS. This means that a randomly chosen family is randomly allocated to a new machine at the first stage. The decision strategy of the SA is divided into the following cases:

1. The new schedule dominates the old one in all objective values. The new solution is accepted and used as the next start solution.
2. The old schedule dominates the new one.
3. Neither the old schedule nor the new one dominates the other.

For cases two and three, the Boltzmann distribution is used to decide whether to accept a new solution or not. A weighted sum of the observed objective values was used since the Boltzmann distribution contains only one value.
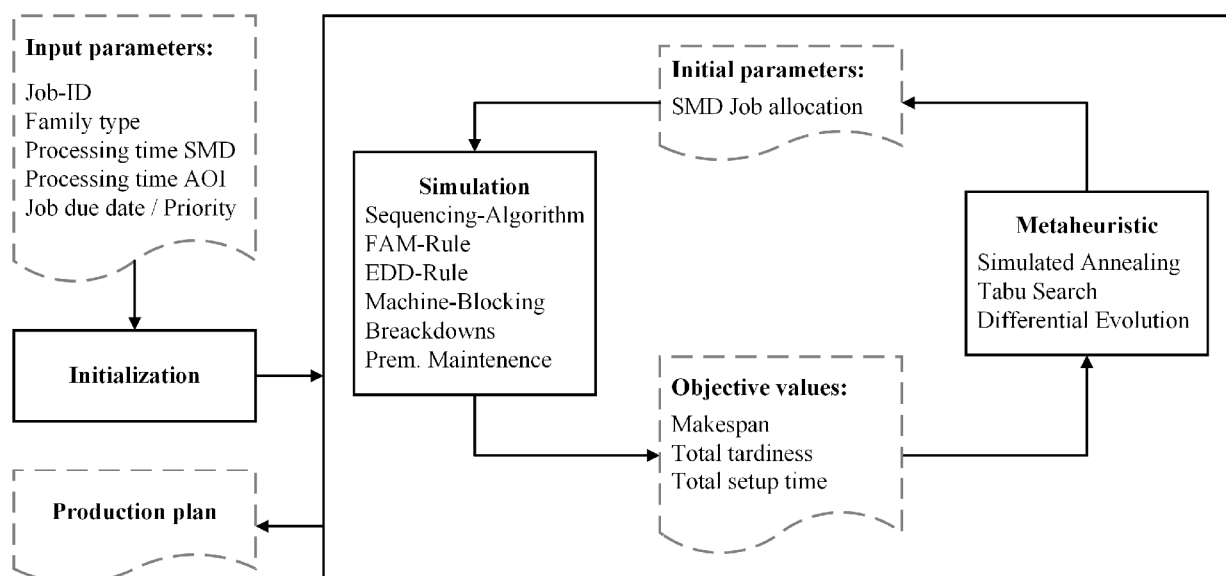
Figure 2: Metaheuristic Simulation-Based Optimization

3.2.2. Tabu Search

Tabu Search (TS) is a deterministic local search technique guided by a fixed or adaptive memory structure. It was developed by (Glover 1977), in order to solve combinatorial optimization problems. As decision strategy for the adapted TS the best neighbor strategy is used. The NHS is again a single point operator neighborhood search. This means that in each iteration each family is once allocated to each machine on the first stage. When the neighborhood search is finished the neighbors are compared to each other in order to find the best neighbor. If no neighbor exists that dominates all other neighbors in all objective values, the weighted sum is used to identify the best solution.

3.2.3. Differential Evolution

Differential Evolution is a stochastic population based optimization technique, introduced by (Storn and Price 1997). In contrast to other metaheuristics it is comparatively new.

For the NHS the user can decide between DE/best/1 and DE/rand/1. DE/best/1 means that the new individual is generated based on the best individual from the last generation and two randomly chosen ones. DE/rand/1 means that the individual is generated depending on three randomly chosen individuals from the previous generation. The selection strategy for the DE is a greedy selection between the current individual and its predecessor. If the current solution dominates the previous one or if the weighted sum outperforms the other, the new individual is chosen, else the predecessor is selected for the next generation.
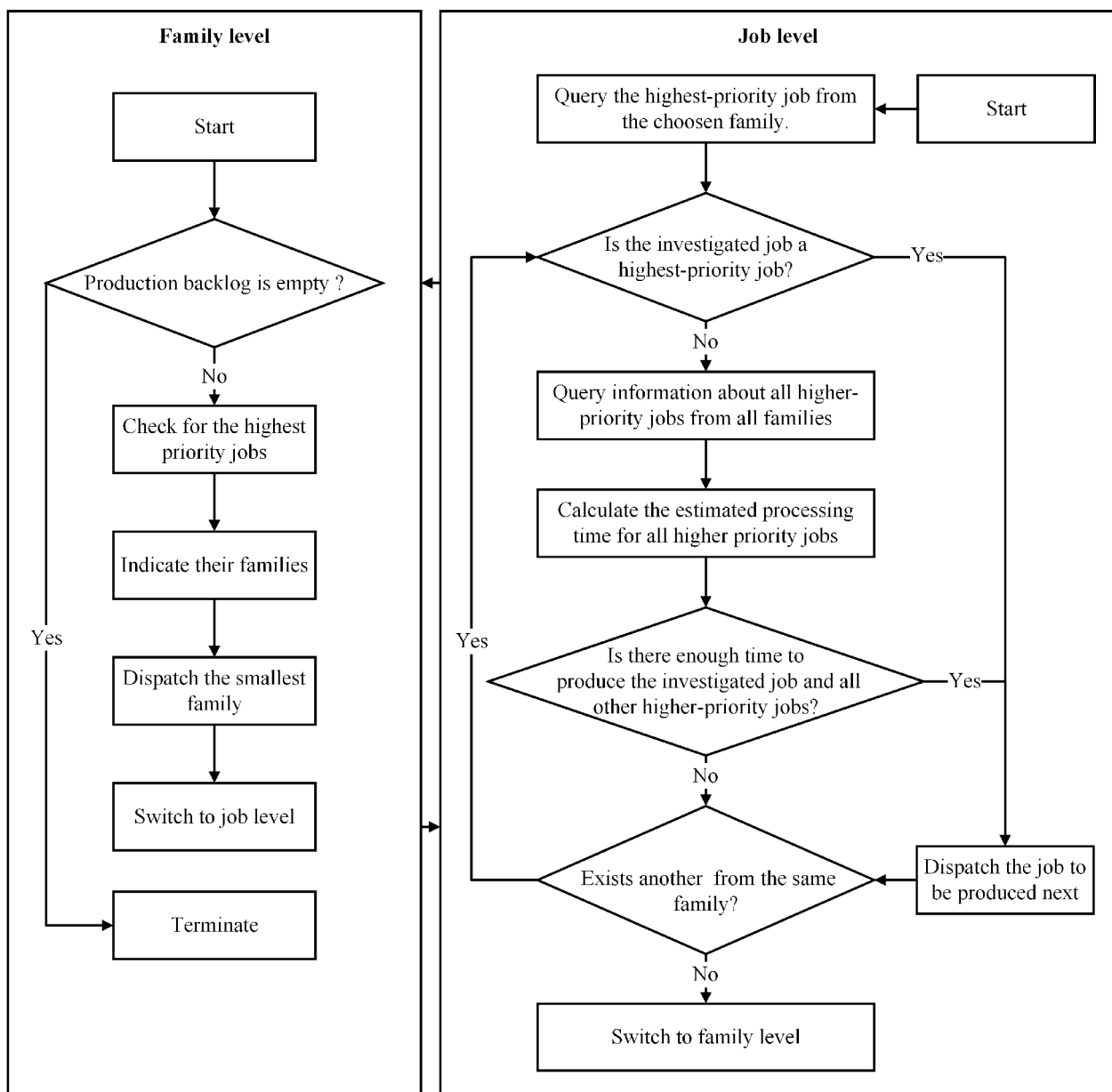
Figure 3: Sequencing Algorithm

3.3. Sequencing algorithm

The sequencing algorithm was developed in order to minimize the objective functions and furthermore to improve the machine utilization of the first stage and was derived from the one formally presented in (Aurich et al. 2016; Nahhas et al. 2016). To meet these conditions all jobs of a family should be processed successively to avoid major setups. But a strict successive processing of jobs from the same family would lead to delivery time violations of many jobs from other families. The algorithm has been developed to resolve this tension.

The behavior of the sequencing algorithm can be divided in two logical levels; a family level and a job level (see Figure 3). When the sequencing algorithm is initiated, it first executes the family level. Here, the smallest family which contains at least one of the highest priority jobs is chosen. The smallest family is the one with the least total processing time of jobs. The reason to choose the smallest family is a chance to completely produce all jobs of it before reaching a critical point. A critical point is met, when it is no longer possible to produce a job from the same family without violating the delivery date of jobs from other families. When the family level of the sequencing algorithm has chosen a family the job level is executed. On the job level the algorithm chooses jobs from the same family according to their priority, using the EDD rule. The sequencing algorithm keeps operating in the job level until all jobs of the family are produced or a critical point is met.

3.4. Machine Blocking

The machine blocking is an extension of the methodologies which were discussed in sections before. The goal of this extension is to avoid a delay of jobs. It is implemented on the stages two, three and four. When a job leaves stage *k*, it is checked if a job is processed in stage *k-1*, which will be delayed, when it has to wait in stage *k*. If this case occurs the free machine of stage *k* is blocked for the job. The blocking only takes place when all machines of a stage are full except one machine and all jobs in the queue of stage *k* are not critical. Critical means that a job must be processed as fast as possible to meet its due date. More precisely the maximum

blocking time is up to 120 minutes and the job a machine is blocked for, has to be critical.

4. DESIGN OF EXPERIMENTS

The datasets which were used for the experiments are real records from the company's production pool. Each dataset represents a production schedule of three weeks. The characteristics of those datasets are shown in Table 1. In general the dataset contain a relatively large amount of jobs, with heterogeneous processing times.

In order to compare the impact of the deterministic and the stochastic breakdowns, several runs with all datasets and solution methodologies were executed. A deterministic breakdown happens at the end of a day for 135 minutes. This represents a 45 minute breakdown in each of the three shifts of a day. The behavior of the stochastic breakdowns can be described with the time between failure (TBF) and time to repair (TTR) scheme. The TBF is based on a normal distribution with a mean of 480 minutes and a standard deviation of 120 minutes. For the TTR a mean of 45 minutes and a standard deviation of 15 minutes were used. That is about 135 minutes per day, which is similar to the deterministic behavior. Thus, the two breakdown variants are approximately comparable. The use of the normal distribution follows from an analysis of the machine breakdowns. In the stochastic case the objectives of each schedule are the mean values of teen simulation-runs. To check the quality of an optimized schedule, 200 stochastic simulation runs were executed.

For the SA the following ranges of control parameter values were used: temperature $T \in \{15,30\}$, linear cooling rate $\alpha \in \{0.1\}$ and step size $n \in \{10,25\}$. The low α and the relative high n lead to a slow cooling and thus prevent a large amount of entropy.

The control parameters of the TS were set with the following ranges: tabu list length $TL \in \{5,20\}$ and number of iterations $i \in \{30,75\}$.

The experiments with the DE were executed with the following ranges of parameter values: number of generations $G \in \{25,100\}$, population size $NP \in \{20,50\}$, crossover rate $CR \in \{0.05\}$, mutation factor $F \in \{0.3\}$.

Table 1: Input Datasets

	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**
Number of jobs	181	179	194	170
Number of families	21	23	24	29
SMD: process time interval | mean	7 - 2912 | 306	6 - 1376 | 366	3 - 1151 | 228	3 - 1957 | 326
AOI: process time interval | mean	9 - 3747 | 381	8 - 1605 | 456	4 - 1495 | 286	4 - 2173 | 403
SS: process time interval | mean	50 - 3250 | 845	133 - 2808 | 1083	11 - 3641 | 648	79 - 2804 | 1133
SS: proportion of jobs	15%	11%	18%	12%
C: process time interval | mean	137 - 8973 | 1322	67 - 2870 | 1044	70 -2302 | 786	70 -2302 | 1076
C: proportion of jobs	9%	9%	13%	11%

5. COMPUTATIONAL RESULTS

Table 1 and Figure 4 show the computational results of the applied approaches. One of the research questions was, whether metaheuristics are able to optimize the job allocation, while the sequencing is done by a special algorithm and which metaheuristic performs best.

Overall the computational results of the applied metaheuristics are very similar. The formulation of a statement, which metaheuristic outperforms another in terms of the objective functions, is very difficult for this scheduling problem and these datasets.

Table 2: Computational Results of the different solution approaches

		Makespan (minutes)	Total Tardiness (minutes)	Penalties (number)	Major-Setups (number)	Computational Time (hours)
Dataset 1						
deterministic Breakdowns	SA	29618	352	0,432	18,395	0,73
	TS	29561	250	0,267	19,86	0,45
	DE	29946	291	0,384	18	0,28
stochastic Breakdowns	SA	29921	323	0,458	19,075	9,03
	TS	29703	341	0,465	20	8,53
	DE	29744	404	0,53	23,21	9,03
Dataset 2						
deterministic Breakdowns	SA	27851	0	0,005	24,635	0,54
	TS	28174	2	0,05	22,56	0,65
	DE	27894	68	0,63	24,475	0,73
stochastic Breakdowns	SA	27913	3	0,025	24,015	5,42
	TS	28235	0	0,005	22,94	9,75
	DE	27949	1	0,01	24,1	9,03
Dataset 3						
deterministic Breakdowns	SA	27983	3	0,055	27,755	0,68
	TS	28462	64	0,665	25	0,98
	DE	27886	6	0,01	25	0,73
stochastic Breakdowns	SA	28148	1	0,06	25,14	13,55
	TS	27483	27	0,22	25,25	9,75
	DE	28385	3	0,065	25	9,03
Dataset 4						
deterministic Breakdowns	SA	28543	10	0,325	31,93	0,73
	TS	28464	4	0,135	32	1,18
	DE	28952	7	0,255	30,285	0,68
stochastic Breakdowns	SA	28775	1	0,045	31	6,78
	TS	28821	2	0,08	31	11,78
	DE	28534	1	0,085	34	7,22

The TS slightly outperforms the other approaches in the deterministic case in terms of total tardiness and makespan. When stochastic breakdowns took place, no solution approach dominates over all datasets. In summary, all methods are suitable for solving the considered problem. But, for example, an increasing number families would raise the complexity of the problem, especially the computational time of the TS would increase rapidly, because of the neighborhood search structure, cf. (Nahhas et al. 2016). A short tabu list leads to better results for the TS, this also caused by the small number of families. Furthermore another strategy for the generation of the start population of the DE should be tested. The current random generation seems to be improvable.

The results for the makespan are very similar over all approaches, regardless of whether deterministic or stochastic breakdowns are considered. The evaluation of the provided Gantt chart shows that third and fourth production stage are highly loaded. Therefore they must be regarded as a kind of bottleneck, consequently the stages three and four have a main impact on the makespan. The sooner a job first arrives at one of these stages, rather at stage four, the smaller is the makespan.

The first and second stages are often in an idle state, while the last jobs are produced at stage three or stage four. For this reason the company should maybe rethink the selection of objective functions. For example the total completion time could be a good substitute. On the other hand is the considered environment a dynamic system and a period of three weeks will be optimized. Since an optimization run will take place at the end of each week, more jobs could be added, which doesn´t need to be processed on stage three or four.

Another research questions was if it is necessary to consider stochastic breakdowns or whether deterministic breakdowns also lead to stable production schedules. The experiments of both procedures showed very similar results in terms of makespan. In regard to total tardiness and total major setup times it seems that the deterministic procedure tends to better results. The computational time increases tenfold at stochastic procedure. It could be shown that the considering of deterministic breakdowns led to stable results, while the computational effort rapidly decreases. This procedure suits very well for the consideration of machine breakdowns.
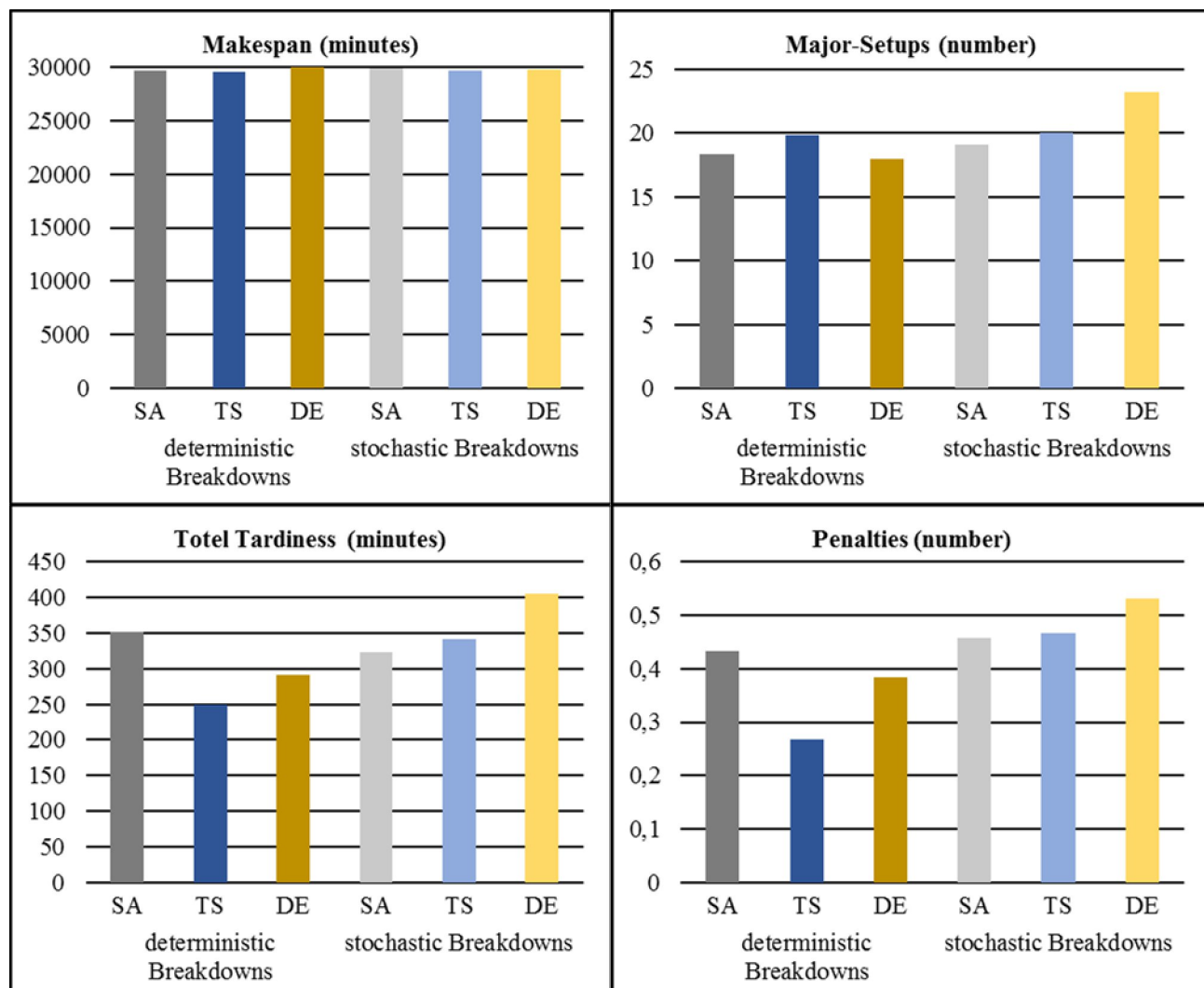

Figure 4: Computational Results of the different solution approaches for the first dataset

6. CONCLUSION

In this paper, a four stage hybrid flow shop scheduling problem was solved with a simulation-based optimization approach. Sequence-dependent setup times, stochastic and deterministic machine breakdowns and preemptive maintenance were considered. The allocation problem of the first production stage was optimized with the metaheuristics simulated annealing, tabu search and differential evolution. The sequencing algorithm proposed by (Nahhas et al. 2016) was implemented to solve the sequencing problem of the machines on the first stage. For the following stages the first available machine rule was used for the allocation and the earliest due date rule was applied for the sequencing. The results of several experiments showed, that a deterministic representation of breakdowns suffices to reach stable production schedules.

REFERENCES

Allaoui H., Artiba A., 2004. Integrating simulation and optimization to schedule a hybrid flow shop with maintenance constraints. Computers & Industrial Engineering, 47 (4), 431–450.

Allaoui H., Artiba A., 2006. Scheduling two-stage hybrid flow shop with availability constraints. Computers & Operations Research, 33 (5), 1399–1419.

Aurich P., Nahhas A., Reggelin T.; Tolujew, Juri, 2016. Simulation-based optimization for solving a hybrid flow shop scheduling problem: Proceedings of the 2016 Winter Simulation Conference. Winter Simulation Conference 2016, 2809–2819, 11.-14. December. Arlington, VA.

Černý V., 1985. Thermodynamical approach to the traveling salesman problem. Journal of Optimization Theory and Applications, 45 (1), 41–51.

Garey M. R. and Johnson D. S., 1979. Computers and intractability. A guide to the theory of NP-completeness. New York: Freeman.

Gholami M., Zandieh M., Alem-Tabriz A., 2009. Scheduling hybrid flow shop with sequence-dependent setup times and machines with random breakdowns. The International Journal of Advanced Manufacturing Technology, 42 (1-2), 189–201.

Glover F., 1977. Heuristics for integer programming using surrogate constraints. Decision sciences, 8 (1), 156–166.

Graham, R. L.; Lawler, E. L.; Lenstra, J. K.; Kan, A.H.G.Rinnooy, 1979. Optimization and Approximation in Deterministic Sequencing and Scheduling: a Survey. In Hammer P. L., Johnson E. L., Korte B. H. (Eds.): Discrete Optimization II, Proceedings of the Advanced Research Institute on Discrete Optimization and Systems Applications of the Systems Science Panel of NATO and of the Discrete Optimization Symposium co-sponsored by IBM Canada and SIAM Banff, Aha. and Vancouver (Annals of Discrete Mathematics, 5), vol. 5. North Holland: Elsevier, 287–326.

Gupta J. N. D., 1988. Two-Stage, Hybrid Flowshop Scheduling Problem. Journal of the Operational Research Society, 39 (4), 359–364.

Hoogeveen J. A., Lenstra J. K., Veltman B., 1996. Preemptive scheduling in a two-stage multiprocessor flow shop is NP-hard. European Journal of Operational Research, 89 (1), 172–175.

Johnson S. M., 1954. Optimal two- and three-stage production schedules with setup times included. Naval Research Logistics, 1 (1), 61–68.

Kirkpatrick S., Gelatt C. D., JR, Vecchi M. P., 1983. Optimization by simulated annealing. Science (New York, N.Y.), 220 (4598), 671–680.

Lee C.-Y., 1996. Machine scheduling with an availability constraint. Journal of Global Optimization, 9 (3-4), 395–416.

Li S., 1997. A hybrid two-stage flowshop with part family, batch production, major and minor set-ups. European Journal of Operational Research, 102 (1), 142–156.

März L., Krug W., Rose O., Weigert G., 2011. Simulation und Optimierung in Produktion und Logistik. Praxisorientierter Leitfaden mit Fallbeispielen. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Nahhas A., Aurich P., Reggelin T.; Tolujew, Juri, 2016. Heuristic and metaheuristic simulation-based optimization for solving a hybrid flow shop scheduling problem: Proceedings of the International Conference on Modeling and Applied Simulation 2016. The 15th International Conference on Modeling and applied Simulation, 95–103, 26.-28. September. Cyprus.

Pinedo M. L., 2012. Scheduling. Theory, Algorithms, and Systems. Boston, MA: Springer US.

Ruiz R., Vázquez-Rodríguez J. A., 2010. The hybrid flow shop scheduling problem. European Journal of Operational Research, 205 (1), 1–18.

Saidy D., Reza H., Taghi Taghavi-Fard M., 2008. Study of scheduling problems with machine availability constraint. Journal of Industrial and Systems Engineering, 1 (4), 360–383.

Storn R., Price K., 1997. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. Journal of Global Optimization, 11 (4), 341–359.

Tang C. S., 1990. Scheduling batches on parallel machines with major and minor set-ups. European Journal of Operational Research, 46 (1), 28–37.

Wittrock R., 1990. Scheduling parallel machines with major and minor setup times. International Journal of Flexible Manufacturing Systems, 2 (4).

AUTHORS BIOGRAPHY

PAUL AURICH is a master student of Industrial Engineering with specialization in production systems at the Otto von Guericke University Magdeburg. His main work interests are scheduling problems and simulation based optimization solution strategies as well as automated modelling. Paul Aurich holds a bachelor degree in Industrial Engineering with specializing in logistics from Otto von Guericke University Magdeburg. His email address is paul.aurich@mail.de.

ABDULRAHMAN NAHHAS is a PhD-student of Business Informatics at the Otto von Guericke University Magdeburg. His main work interests are scheduling problems and simulation based optimization solution strategies. Abdulrahman Nahhas holds a master degree in Business Informatics from Otto von Guericke University Magdeburg. His email address is abdulrahman.nahhas@hotmail.com.

TOBIAS REGGELIN is a project manager at the Fraunhofer Institute for Factory Operation and Automation IFF and Otto von Guericke University Magdeburg. His main research and work interests are logistics system modeling and simulation and the development and conduction of logistics management games. Tobias Reggelin received a doctoral degree in engineering from Otto von University Magdeburg. Furthermore, he holds a diploma degree in industrial engineering in logistics from Otto von Guericke University Magdeburg and a master's degree in Engineering Management from Rose-Hulman Institute of Technology in Terre Haute, IN. His email address is tobias.reggelin@ovgu.de.

MARCO KRIST is a logistic manager of the company Thermo Fisher Scientific. His main work interests are optimization of production systems and supply chain planning. Marco Krist holds a diploma degree in industrial engineering in logistics from Otto von Guericke University Magdeburg. His email address is marco.krist@thermofisher.com.

A REINFORCEMENT LEARNING APPROACH TO SCHEDULING DUAL-ARMED CLUSTER TOOLS WITH TIME VARIATIONS

Ji-Eun Roh[a], Tae-Eog Lee[b]

[a],[b]Department of Industrial and Systems Engineering,
Korea Advanced Institute of Science and Technology (KAIST), 305-701 Daejeon,
Republic of Korea

[a]rje0531@kaist.ac.kr, [b]telee@kaist.ac.kr

ABSTRACT

The cluster tool, which consists of multiple process chambers are widely used in the semiconductor industry. As the process of wafers becomes more sophisticated, the operation of cluster tools is also being improved. To effectively operate cluster tools, several rule-based schedules, such as the swap sequence have been developed. However, scheduling in time variance environment is not fully considered. In this paper, we propose a cluster tool modeling method, which can handle time variance in dual-armed cluster tool. Then, we present a reinforcement learning process based on the proposed cluster tool model to find new operational schedules in specific configurations. To measure the performance of the newly obtained schedule, makespan is compared under the new policy and the swap policy. The makespan reduced compared to the conventional swap policy, which implies that the reinforcement learning well learned the operation schedule in the time variance environment.

Keywords: cluster tool, Markov decision process, scheduling, reinforcement learning

1. INTRODUCTION

Along with the innovative development in the semiconductor manufacturing industry, quality issues in the wafer manufacturing process have been discussed. The technologies of each process have rapidly developed and improved the production quality of the wafers. To avoid quality issues due to batch production, the cluster tool that process one wafer at a time are now widely used in the semiconductor industry (Lee 2008). Chambers in cluster tools do not process wafers in units of batches but process them individually, so they can meet the quality standard. The cluster tool consists of usually four to six processing chambers and one transport robot. In each chamber, only one wafer is processed at a time, and one of the process steps specified before the start of the process is processed. In addition, the transport robot moves the wafers inside the cluster tool. To start the processes, a wafer enters the chamber that is responsible for the first process step. After the process is completed in the chamber, the wafer is transported to the next chamber by a transport robot. The transport robot repeats the process of unloading the processed wafer from the chamber and loading the wafer in the next appropriate chamber which is in charge of the next process step. Once all the required processes are completed with the proper chamber sequence, the wafer process in the cluster tool finally is completed.

Such a configuration leads cluster tools to have several issues in operation. Since cluster tools consist of only processing chambers and a single transport robot, only one robot operation is possible at a time. Hence, the wafers cannot be moved at any time, and can only be moved when the transport robot moves them. Even if the multiple chambers are ready for the process, multiple wafers cannot be delivered at the same time. In addition, the cluster tools do not contain any buffer space in their interior space, so the only way to store the wafers outside the chamber within the cluster tool is to hold it with the transport robot. When the robot loads a wafer into the chamber, the chamber starts the process, and the process ends naturally after the process time. The chamber processes for a period of time without any decision. Therefore, we can say that the overall cluster tool schedule depends on the order decisions of the transport robot. As the transport robot unloads or loads the chambers, the status of the cluster tool system has changed. This can be seen as a discrete event system because the state of the system changes as the robot takes action. Petri nets, finite state machines (FSMs), Markov decision processes (MDPs), etc. are used to model discrete event systems (Murata 1989; Choi and Kang 2013; Puterman 2014).

Several studies about cluster tools scheduling have used a timed Petri net (TPN) model, in which the places or transitions have a time delay. A TPN is classified into deterministic and stochastic TPN whether the time delay is deterministic or stochastic (Murata 1989). Each of them is a modeling technique that deals with different kinds of time delays. Many studies have been conducted on finding policies for the deterministic environment. For example, (Lee, Lee and Shin 2004, Zuberek 2004) have modeled analyzed cluster tools by the deterministic TPNs, and Jung and Lee (2012)

proposed a mixed integer programming (MIP) model to find the optimal policies in deterministic TPNs. In stochastic environment, there also has been several studies. To deal with time variation in cluster tool, (Kim and Lee 2008) suggested an extended Petri net and developed a graph-based procedure to verify the schedulability condition. Qiao, Wu, and Zhou (2012) have introduced two-level architecture to deal with time variation, and proposed some heuristic algorithms to find a schedule for one of the architectures. Molly (1982) suggested that the stochastic TPNs are isomorphic to finite Markov processes under the certain conditions.

However, new methods for finding good policies in specific systems are studying nowadays. Reinforcement learning (RL) is widely used in finding policies in many fields such as manufacturing systems, autonomous vehicle control, finance, and games. This reinforcement learning is one of the solutions to find optimal policies in MDP models (Sutton and Barto 1998). Hence, modeling the system behavior by MDPs, then find the policies for the system by RL is widely used (Moody et al. 1998; Mnih et al. 2015).

In this paper, we propose a model using MDPs, which can handle time variations in cluster tools. Then, we report how we learn the new robot sequence using the MDP model to minimize the average makespan in time variation environment, and measure the performance of the obtained sequence. This study is an attempt to schedule the cluster tool using reinforcement learning. Since the model is designed to represent the stochastic process time of the cluster tool, the schedules obtained from reinforcement learning can be expected to be well applied in a stochastic environment.

2. MODELING DUAL-ARMED CLUSTER TOOLS WITH TIME VARIATIONS

To model dual-armed cluster tools with time variations, we use a MDP model. After introducing the problem of configuring the MDP model, the model we proposed is reported.

2.1. Markov Decision Process

A Markov decision process (MDP) is a tuple $< S, A, P, r >$, where S denotes a set of states, A a set of actions, P a state transition probability distribution, and r a reward function, respectively. The process follows the Markov property, which means the transition probability and reward functions depend only on the current state and the action, not the past history. In this paper, we address only the stationary environments, which means the system properties does not change as time goes by. According to Puterman (2014), at every decision epoch, a state $s \in S$, which is a representation of a system, is observed and an action $a \in A_s$ has to be chosen by a decision maker from the set of allowable actions in the state s. As a result of the action, the system state at the next decision epoch changes by some transition probability $P : S \times A \times S \to [0, 1]$. $P_{s,s'}^a$ is determined by the current state s and the chosen action

a. The decision maker gets a reward $r : S \times A \times S \to \mathbb{R}$, and $r_{s,s'}^a$ is determined by the current state s, action a, and the next state s'. Here, we have a concept of a decision rule, which describes a procedure for action selection in each state. This rule is called a stationary policy, $d : S \to A_s$. The decision maker passes through a sequence of states s_t, which are determined by transition probabilities $P_{s,s'}^a$ and the actions $a_t = d(s_t)$ the decision maker chooses, then the sequence of reward $r_{s,s'}^a$ is obtained. The popular performance metric is discounted reward, which is the sum of the discounted reward gained over the entire time horizon when we allow the specific policy. The discounted reward of a policy d starting at state i is defined as $J_d(i) \equiv \lim_{k \to \infty} \mathrm{E} \left[\sum_{j=1}^{k} \gamma^{(j-1)} r_{s_j,s_{j+1}}^{d(s_j)} \mid s_1 = i \right]$, where $\gamma \in [0,1]$ is a discount factor. By Bertsekas (1995), it is proved that $J_d(i) = \left[\sum_{s' \in S} P_{s,s'}^{d(s)} \left(r_{s_j,s_{j+1}}^{d(s_j)} + \gamma J^d(s') \right) \right]$. The function J_d is called the value function for policy d. There are many variants of MDP model, and one of the them is a semi-Markov decision process (semi-MDP). In semi-MDP, temporal factors are included in the modeling. The original MDP assumes that each transition takes the same time through all the states; however, semi-MDP considers the transition time to follow arbitrary probability distributions.

2.2. Cluster tool modeling with MDP

If we simply insert the state of the chambers and actions the robot takes into MDPs as previous studies have done with deterministic TPN models, the MDPs does not properly represent the behavior of the cluster tools. This is due to characteristics of cluster tools. The cluster tools are not simply changing the systems by a discrete event; cluster tools are changing the systems by the time element. The general MDP assumes that every transition takes the same time, however, the transitions in cluster tools cannot be assumed to be the same. Considering the different transition times between states, we may think of simple ways to model the cluster tool system.

1. Model with MDPs which have a constant transition time, one second. Adjust transition probabilities to represent the time element.
2. Model with semi-MDPs so we can insert time information to the transition distribution.

To verify above two methods, we first consider the example case as shown in Figure 1.

The example represents a status change in a single chamber, which has two states A and B, and two actions 1 and 2. Consider the case where the environment state changes from A to B after 50 seconds. By using the first modeling method listed above, we can set transition probability to 1/50; thus, the state changes from A to B occurs after 50 seconds in average. However, the average cannot express the actual individual transition time explicitly.

$a^i \in \{0,1\}$: action
$p_a^i \in [0,1]$: transition probability
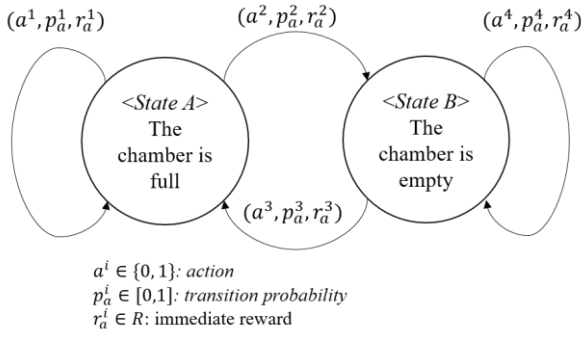$r_a^i \in R$: immediate reward

Figure 1: Simple Example for Cluster Tool System

By using the second modeling method, we can set transition time distribution as 50 seconds. When using the second modeling method, we can express that the transition occurs after 50 seconds in Figure 1 if we set the transition time distribution to 50 seconds (as a constant). However, the cluster tool is a tool with multiple chambers, so the model should reflect changes in state in other chambers. If 20 seconds have elapsed while the status changes in another chamber, and so the chamber of the example now needs to change its state after 30 seconds, the transition time distribution should be changed to 30 seconds. Every time the transition occurs in the other chamber, we have to calculate and change the distribution; the environment of the system changes, which does not satisfy the markov property.

The two MDP modeling methods above are not appropriate for modeling the cluster tool operations in time variations, since they do not represent the transition time explicitly or do not satisfy the Markov property. To satisfy the Markov property, a state should contain sufficient information, so that the state does not require past history to obtain the next state or reward. Therefore, state is designed to contain information about chamber and robot usage, and the remaining process time.

In this study, We propose to represent the state as $(W, C_1, C_2, \ldots, C_n, R, S_1, S_2, Z_1, Z_2, \ldots, Z_n)$, where n is the number of processing chambers in cluster tool. W is the number of wafers remaining in the loadlock. C_i is the state of the chamber, which represents whether the ith chamber is empty ($C_i = 0$), full ($C_i = 1$), or completed processing ($C_i = 2$) for $i \in \{1, 2, \ldots, n\}$. R is the number of wafers held by the transport robot. Since we are dealing with a dual-armed cluster tool, R can have a value of 0, 1, or 2. S_1 and S_2 are the next process steps of the wafers held by the robot. These represent the process steps that each wafer should visit in the next step. Z_i is the expected remaining process time of the ith chamber for $i \in \{1, 2, \ldots, n\}$. The value is calculated by subtracting the elapsed time from the initially set process time, because it is not known what the actual process time will be in a time variation environment. In addition, define S to be the state space of the proposed state.

Hence, the proposed state structure contains the information about the chamber and the wafer inside it,

and includes information on which chambers the wafers on the robot should be sent. Finally, it roughly contains information when the chambers are going to be finished. Then, the model satisfies the Markov property.

To better understand the structure of the states, we show state representation of Figure 2 by using the proposed states structure. Suppose that the black and white wafers indicate that they are in the first and second process steps, respectively, and the hatched wafer indicates that the process is in progress. According to the proposed structure, the state is ($W = 3, C_1 = 2, C_2 = 1, C_3 = 1, C_4 = 2, R = 1, S_1 = 2, S_2 = 0, Z_1 = 0, Z_2 = 4, Z_3 = 4, Z_4 = 0$). This means that the three wafers are remained in the loadlock, the process is going on in the second and the third chambers, whereas the first and the forth chamber have completed its process. The robot holds a single wafer waiting to enter the second process step, and the remaining process times of the chambers are four for the second and the third chambers.
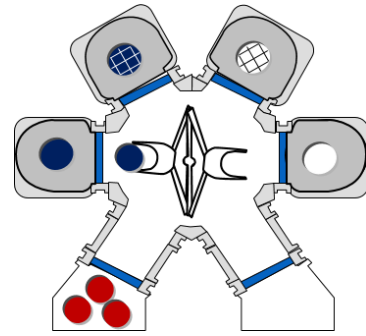
Figure 2. A Dual-armed Cluster Tool with Four Chambers

A is a set of actions that the transport robot can perform, which is $\{Wait, U_j, L_j, SW_i\}$, where $j \in \{0, 1, 2, \ldots, n\}$ and $i \in \{1, 2, \ldots, n\}$. They all represent the robot tasks: U_j, L_j, and SW_i indicate unload, load, and swap operations on the jth or ith chamber. The unload, load, and swap operations are the common robot tasks in other modeling methods. However, in this study, $Wait$ action is added. $Wait$ represents the robot waiting until the state changes, such as finishing the process in a chamber. To make the model transition independent on remaining process time, we suggest $Wait$ to be an action of the robot. If without a $Wait$ action, the state of the cluster tool changes when the process completes, even if no action is selected. Furthermore, for all states $\{Wait, U_j, L_j, SW_i\}$ can always be selected as an action; therefore, $3n + 3$ actions are available to be chosen in all states.

Transition probability is denoted by $P_{s,s'}^a$, where s is the current state, a is the current action and s' is the next state. It stands for probability for transition to state s' when we choose action a in state s. In this study, we do not consider unexpected event occurrence such as robot failure, so the system always transposes exactly to the specified state. Hence, when the action is one of U_i, L_i, and SW_i, $P_{s,s'}^a$ has a value of 1 for the appropriate s'. Here, the appropriate state s' is when s'

is likely to be the next state in state s. Furthermore, when the s' is not likely to be the next state in s, $P^a_{s,s'} = 0$. However, when the action is *Wait*, transition probability $P^a_{s,s'}$ does not only have a value of 0 or 1. Assume that the action is *Wait*. If at least one chamber is 'full', which stands for $C_i = 1$, the chambers should be selected which change their state to 'completed processing'. In other words, it chooses which chamber the process ends with. Choosing the chamber with the shortest remaining process time is the simplest and most intuitive. However, since we are dealing with a time variation environment, we cannot say that the process is finished after Z_i, exactly the expected remaining process time. Therefore, the chamber selection should reflect the fact that unexpected chamber processes may end sooner. In addition to selecting the chamber, the transition time needed to be stochastically generated to reflect the time variation environment. In general, the process of the chamber is completed within a range of times, including the average time. Therefore, we choose to generate the transition time using a beta distribution that can generate the average value within the bounded range.

Reward function is set to get 1 if a wafer completes all processes, otherwise, the function is set to get -0.1 in all transitions except for the *Wait* actions. When the *Wait* action is selected, the reward is set to be $-0.1 \times$ (transition time).

Detailed transition rules and reward definitions are introduced in Sections 3.2 and 3.3.

3. LEARNING POLICIES FOR ROBOT MOVES IN CLUSTER TOOLS BY REINFORCEMENT LEARNING

There are various ways to solve MDP. First, when we know the model, we can define a total return. Then by dynamic programming, we can find the optimal value function of Bellman optimality equation (Bellman 1954). However, the proposed model has a time element in the state, it is not possible to solve the MDP through dynamic programming due to the curse of dimensionality. Therefore, we applied reinforcement learning (RL) to find the policy for the proposed MDP model. RL requires an environment to interact with. So we built an interactive environment based on the proposed MDP model. We then set up tasks to perform the learning, and learned the robot policy using Q-learning. After learning, we performed another experiment to measure the performance of the policy.

3.1. Reinforcement learning
Reinforcement learning (RL) is useful when the classical dynamic programming is not enough to solve the MDP problems. Dynamic programming is not suitable for the problem, when the environment is difficult to build a perfect model due to its unknown transition probability (curse of modeling), or when the environment has too many states to solve (curse of dimensionality). RL adds the concepts of stochastic approximation, temporal differences, and function approximation to classical dynamic programming, so that it can solve the MDP problem even if the transition probability is not explicitly represented, and the state dimension is too large (Gosavi 2014).

3.2. The environment
Since we find the robot policy by RL, we need an environment that returns the next state and reward for action according to state for learning progress. The environment required for RL is either an environment consisting of real equipment or an environment built by computer simulation. The best way to interact with the environment is to build the real world environment, then conduct the experiments with it. However, using a real world cluster tool is nearly impossible due to its value. A single tool is so expensive and needs large area to install; therefore, even the company is not able to arrange enough cluster tools for their manufacturing facilities. To replace the real world environment, we built a virtual environment, which responds to some stimulation.

The simulation environment mimics the real world; however, not fully, only partially. To make the partially reflected simulation environment to involve the core elements we want, we needed to be careful making the simulation environment. We made some assumptions in building an environment. First, we do not consider machine failure. Process time can vary depending on a machine and time; however, we do not handle machine failure in this study. Second, the environment is stationary, which means system dynamics does not change as time goes by. No matter how much time has passed, they follow the same distribution. Third, the time dynamics of the systems are governed by the rules we set; hence, it may not fully reflect the real world cluster tool behaviors.

The simulation environment shows the change of the environment depending on the agent's action. It shows the appropriate reward and the next state according to the selected action. Some actions are valid only in particular states. For example, robots can take a *Wait* action only when more than one chamber are filled with the wafer, and are not completed processing, which means that chamber i with a C_i value of 1 exists. Detailed rules for robot actions are specified in Table 1.

Table 1: Possible Situations for the Robot Actions

Robot Actions	Possible Situations
Wait	When at least one chamber is filled with a wafer has not been completed the process
Unload	When the chamber has a processed wafer, but still have the wafer inside in it
Load	When the robot holds some wafers, and one of the chambers corresponding to the steps that the wafers are going to enter is empty
Swap	When the chamber has a processed wafer, but still has the wafer inside in it. At the same time, the robot holds only one wafer to enter the chamber

Table 1 shows the possible situations for the action. The action is valid only when the right box condition is satisfied. If we select an action that is not in those possible situations, the state will not change. We set the environment to prevent state transitions when inappropriate actions are taken. We call an action taken in an impossible situation an *invalid action*, and an action taken in a possible situation a *valid action*. For example, if a chamber has processed wafers but the wafers are still in it and at the same time the robot has only one wafer in its arm, swap action on the chamber is valid action, while the rest actions are not.

All transitions are treated to be deterministic move except for $Wait$ action, as mentioned in Section 2.2. Other actions cause the environment to have the same response each time a state is entered. However, when a $Wait$ action is chosen, the state transition is performed stochastically and the transition probability is defined differently depending on the state. When only one chamber is in the 'full' state, the chamber is selected to complete the process. However, when more than one chambers are in 'full' state, changes in environment is determined by the algorithm below.

- With a probability 0.9, the chamber with the minimum remaining process time completes the process
- With a probability of 0.1, the selected chamber completes the process with a probability of inversely proportional to the remaining process time

Every time the $Wait$ action is chosen, the transition time is set to be stochastically selected along the $beta(50,50)$ distribution.

3.3. Task description
In general, a set of 25 wafers becomes a 'wafer cassette', which is the unit of production for cluster tools. Therefore, the goal is to quickly finish a single cassette process rather than a single wafer process. In fact, in real companies usually produces 100 to 200 wafers once, which means produces 4~8 wafer cassette without stopping the operation of the cluster tool. The problem of minimizing the total makespan can be thought of as a reduction in the total makespan itself and a reduction in the makespan of each subtask. When the whole makespan is continuously measured and returned at the end, the reward structure is heavily sparse to learn the policy. Therefore, in this study, we follow the method of minimizing the makespan of each subtask, which is completing a single wafer. In addition, we follow the structure of receiving a reward at the end of each wafer process. In order for all subtasks to ultimately reduce the total makespan, the total return is defined as the discounted reward. So learning is done in the direction of receiving the reward as soon as possible.

Furthermore, after 100 ~ 200 wafers are produced in real world, the equipment configuration may be changed. In the case of an episodic task that ends after the production of a certain number of wafers, the whole process can be divided into start-up, steady, and close-down cycles (Kim, Lee and Kim 2016). However, in this study, we examine whether there is a policy change when the environment changed from deterministic to stochastic in the steady cycle. Therefore, we consider the learning task as a continuous task to learn a steady cycle.

The most important goal of this problem is to learn to choose an action that maximizes the total return. But we should learn to prevent the system from selecting a behavior that makes it deadlock and learn to prevent invalid actions.

3.4. Learning procedure
The cluster tool behavior was learned by computer simulation-based reinforcement learning. Learning is achieved through the interaction of agents and environments. The agent repeats the process of selecting the action, receiving the corresponding response, and updating the Q value for the state and the action. The action selection and Q value update follow the traditional Q-learning method (Sutton and Barto 1998). The algorithm for Q-learning is illustrated in Figure 3. We used an ε-greedy policy to select an action from the current state, where the ε decreased over the time steps.

Even if using the same learning method, depending on how the environment is set, the selectable action set changes and the response to the action changes. In this study, we set the environment to have two chambers, and a wafer flow pattern (1,1) with all process times to be 4 seconds. Here, the wafer flow pattern refers to the number of parallel chambers in each process. The initial state is set to be full loaded, so all the chambers are filled with wafers that have not yet begun to process. Since we only deal with the steady state, we only used $(C_1, C_2, R, S_1, S_2, Z_1, Z_2)$ as the state. With this environment settings, the agent continued learning using Q-learning in the direction of maximizing the total return.

Initialize $Q(s,a), \forall s \in S, a \in A(s)$, arbitrarily, and $Q(terminal-state, \cdot) = 0$
Repeat (for each episode):
    Initialize S
    Repeat (for each step of episode):
        Choose A from S using policy derived from Q (e.g., ε-greedy)
        Take action A, observe R, S'
        $Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma \max_a Q(S',a) - Q(S,A)]$
        $S \leftarrow S'$
    until S is terminal

Figure 3: Q-Learning Algorithm

However, there are other issues that need to be taken care of by agents: deadlock actions and invalid actions.

We want to ensure that when an agent chooses an action in the current state, it does not choose an invalid action and a deadlock action that causes the next state to become deadlock. The agent goes through the process of defining possible action sets for the current state by a two-step look ahead method. To remove the invalid actions for the current state, the agent requests the environment for the results for all the actions, then saves only which actions are valid (one-step look ahead). For an action that is valid for current state, a second look ahead is needed to check if the deadlock occurs. It is time-consuming to always have a two-step look ahead in every learning step. To further speed up the procedures, we followed the deadlock prevention rules to the second look ahead procedure. As we apply the rules, deadlock action checking is needed only when all of the loading actions for the current state are invalid and the number of the holding wafer (R) is 1. In addition, only unloading actions among valid actions need to be checked. If any of the loading actions are valid, the agent does not need to do a deadlock check. A way to check if a given unloading action is a deadlock is to request a response for an appropriate loading action. If the unloaded wafer has a process step k to be visited, check that it is possible to load to the chambers that are in charge of the kth process step. If any of the loading actions are valid, the given unloading action is not a deadlock action, otherwise, the given unloading action is a deadlock action.

For example, consider the case illustrated in Figure 2, where chamber size (n) is 4, the current state is $(3, 2, 1, 1, 2, 1, 2, 0, 0, 4, 4, 0)$, and the wafer flow pattern is (2,2). Since the value of n is 4, the cardinality of the available action sets for all states is 15, which is the value of $3n + 3$. Among these 15 actions, the agent has to avoid invalid actions and deadlock actions. The agent first removes the invalid actions from the action set, and then removes the deadlock actions from the remaining action set. Through interaction with the environment, the agent gets the invalid actions and stores only the valid actions, which is $\{Wait, U_0, U_1, U_4, SW_4\}$. Since there is no loading action among the valid actions, actions $\{U_0, U_1, U_4\}$ become the candidates for the deadlock actions, by the deadlock checking rule that we mentioned above.

Therefore, the agent needs to check the validity of the appropriate loading actions for states (b), (c), and (d) in Figure 4. In state (b), unloaded wafer has to visit the 1st process step; hence, request responses for actions

$\{L_0, L_1\}$, which are in charge of the 1st process step. In this way, responses for actions $\{L_2, L_3\}$ and $\{L_4\}$ are requested to check the deadness of the state (c) and (d), respectively. Then, the agent gets that any loading actions are not valid for state (b) and (c), and $\{L_4\}$ is valid for state (d). Therefore, the unloading actions corresponding to the state (b) and (c) are the deadlock actions, so the actions $\{U_0, U_1\}$ should be removed from the remaining valid action set, $\{Wait, U_0, U_1, U_4, SW_4\}$. Finally, actions $\{Wait, U_4, SW_4\}$ can be chosen from the current state S_0.
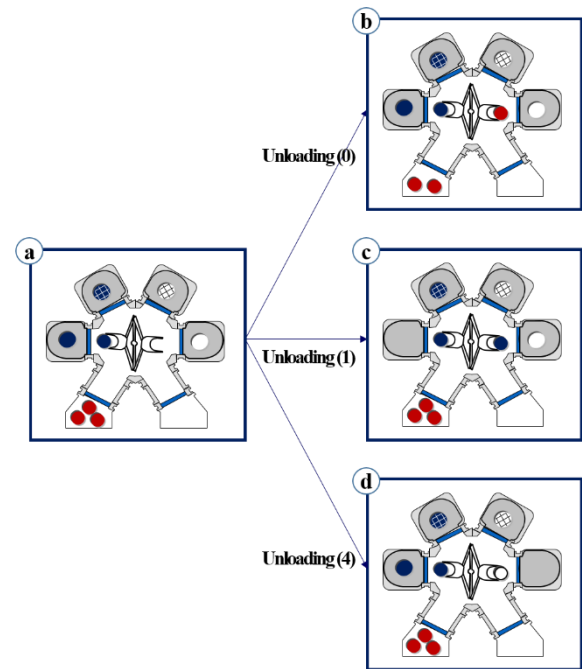
Figure 4: Checking Deadlock Action Candidates

We conducted Q-learning based on the above action selection rule. The agent always chooses a non-invalid and non-deadlock action, regardless of the exploration rate ε. The reason for constructing this additional action selection part before the action selection part in Q-learning is the learning time. In the cluster tool, there are too many invalid or deadlock actions in the whole action. If we do not forbid those inappropriate actions, a computation expense to explore for unnecessary action occurs and unnecessary learning time occurs. Therefore, by adding hand-crafted adjustments specific to this system, learning time has been shortened.

Table 2: Q-values for Some States

	State S $(C_1, C_2, R, S_1, S_2, Z_1, Z_2)$	Q-value $[Q(s,a_1), Q(s,a_1), \dots, Q(s,a_8)]$	Action $argmax_a Q(s,a)$
1	(0, 0, 0, 0, 0, 0, 0)	[-1., 4.43624915, -1., -1., -1., -1., -1, -1, -1]	1
2	(0, 0, 1, 0, 1, 0, 0)	[-1., 4.62024937, -1., -1., 4.59535455, -1., -1., -1., -1.]	1
3	(0, 0, 1, 0, 2, 0, 0)	[-1., 5.17879609, -1., -1., -1., 5.16569827, -1., -1., -1.]	1
4	(0, 0, 1, 1, 0, 0, 0)	[-1., 4.62429555, -1., -1., 4.5976508, -1., -1., -1., -1.]	1
5	(1, 0, 2, 3, 2, 3, 0)	[6.16012302, -1., -1., -1., -1., 6.64482254, 6.63412537, -1., -1.]	5
6	(1, 0, 2, 3, 2, 4, 0)	[6.05787908, -1., -1., -1., -1., 6.42254697, 6.42956666, -1., -1.]	6
7	(1, 2, 1, 1, 0, 4, 0)	[5.84405062, -1., -1., 5.71535707, -1., -1., -1., -1., -1.]	0
8	(1, 2, 1, 2, 0, 0, 0)	[6.39532166, -1., -1., 6.0765975, -1., -1., -1., -1., 6.40542711]	8
9	(1, 2, 1, 2, 0, 1, 0)	[6.31353027, -1., -1., 6.31195477, -1., -1., -1., -1., 6.48067444]	8
10	(2, 1, 0, 0, 0, 0, 4)	[5.67589726, 5.68395603, 5.53326838, -1., -1., -1., -1., -1., -1.]	1
11	(2, 1, 1, 0, 1, 0, 1)	[6.15144211, -1., 6.06476367, -1., -1., -1., -1., 6.23425531, -1.]	7
12	(2, 2, 1, 2, 0, 0, 0)	[-1., -1., -1., 6.30849807, -1., -1., -1., -1., 6.46170915]	8
13	(2, 2, 1, 3, 0, 0, 0)	[-1., 7.02139286, 6.88018823, 6.88533456, -1., -1., 7.03105732, -1., -1.]	6

3.5. Performance measurement

After the learning is performed based on the proposed MDP model, $Q(s,a)$ values for every state $s \in S$ and action $a \in A$, are obtained. The policy we obtained is a greedy policy that selects the action with the highest Q-value in each state, i.e., $max_a Q(s,a)$. To report the policy, we have to list the selected actions for states as shown in Table 2. However, there are too many states to list all of them as a table, and it is difficult to intuitively understand what actions to take in some cases. Therefore, after we obtained the policy, we used the measures that represent the performance of the policies, then compared the performance of the acquired policy against the existing swap sequence.

Through learning, $Q(s,a)$ for $\forall s \in S, \forall a \in A$ is continuously updated; define $Q(s,a)$ for each thousand wafers produced as $q_0, q_1, \dots, q_t \in \mathbb{R}^{|S| \times |A|}$. However, we do not use all $q_0, q_1, \dots, q_t$ to get the policy; we only use the $Q(s,a)$ value after the convergence. To verify the convergence of a sequence $q_0, q_1, \dots, q_t$, $|q_k - q_{k-1}|$ for $k \in \{1, \dots, t\}$ was checked. If all elements in the matrix $|q_k - q_{k-1}|$ is relatively small, the q_k can be considered to be converged. Figure 5 shows the maximum element in matrix $|q_k - q_{k-1}|$ for $k \in \{1, \dots, 543\}$. The value converges to 0, which means the sequence $q_0, q_1, \dots, q_t$ converged. We used the converged $Q(s,a)$ value after the 543th iteration, i.e., used q_{543} to obtain the policy.

To indicate the performance of the obtained policy, the time it takes to process 50 wafers, makespan, is used. Makespan is measured in the same environment as the learning period. We used two processing chambers, wafer flow pattern (1,1), average process time of four,

and the chamber process time is set to follow the $beta(50,50)$ distribution. Because the process time is randomly generated, makespan can be measured differently, even if the measurement is implemented under the same policy. Therefore, we measured makespan a thousand times under the designated policies and compared the averages.
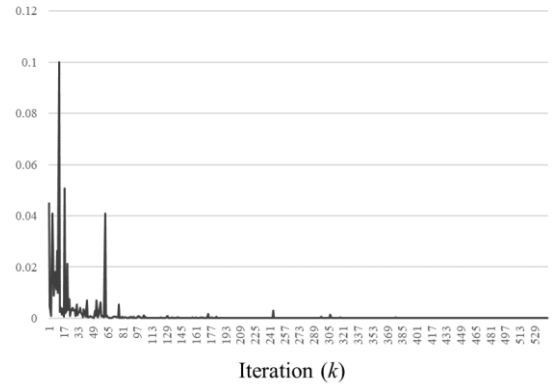
Figure 5: Q-value difference maximum value

4. ANALYSIS OF ROBOT POLICY OBTAINED BY LEARNING

Figure 6 represents the average q-values of the obtained policy and the swap policy. Figure 7 and Table 3 represent the makespan differences of the two policies as a histogram and a table, respectively. M_p and M_s represent makespan under the obtained policy and the swap policy.

We confirmed that the makespan of the policy obtained through Q-learning based on the proposed MDP model

is on average shorter than the swap policy as shown in Figure 6.
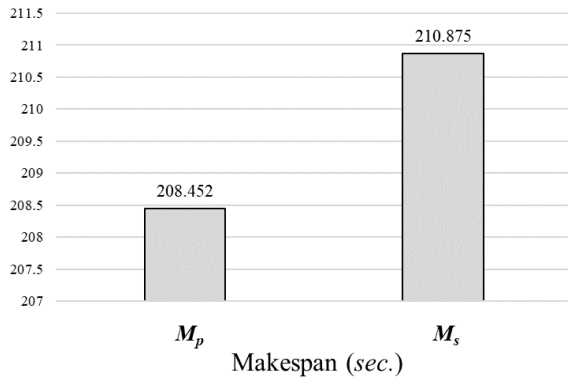
Figure 6: Average Q-values of Two Policies

Except for 2.3% cases, the others acquired the same or shorter makespan with the obtained policy. This means that the above reinforcement learning has found a good robot sequence and obtained a slightly better policy than the widely used swap policy. The average of makespan is not significantly different as illustrated in Figure 6. It seems that the small difference is due to the settings, the average process time and the number of the chambers in our learning environment. Both of the settings may have reduced the time variation effect to the environment. If we were to increase the average process time and number of chambers as much as the actual cluster tool, then the time variance would have increased, and the difference in makespan may have increased.

However, the fact that makespan is reduced in most cases means that as a policy run in a time variance environment, the policy obtained through learning in a time variance environment is better than the policy obtained in a deterministic environment.
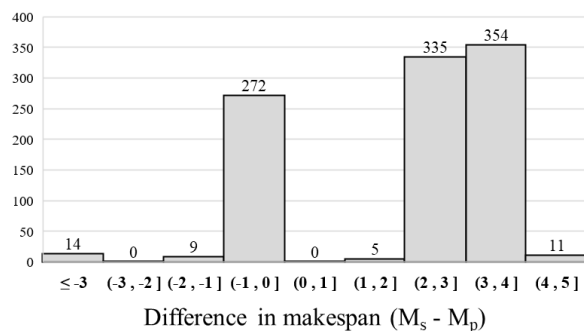
Figure 7: Histogram Graph Showing the Q-value Differences

Table 3: Frequency Distribution of the Q-value Difference

Frequency Distribution of $(M_s - M_p)$			
$M_s < M_p$ Case		$M_s \geq M_p$ Case	
Bin	Count	Bin	Count
[-5, -4)	1	[0, 1)	272
[-4, -3)	10	[1, 2)	0
[-3, -2)	3	[2, 3)	5
[-2, -1)	0	[3, 4)	335
[-1, 0)	9	[4, 5)	354
		[5, 6)	11
23 (2.3%)		977 (97.7%)	

5. CONCLUSIONS

Cluster tool scheduling has been usually studied under the deterministic environment. In this paper, we modeled dual-armed cluster tool behaviors as a Markov decision process in the time variance environment, then applied reinforcement learning to obtain robot policies. The newly obtained robot policies reduces the makespan of 50 wafer processes compare to the conventional swap sequence in most cases. From this study, it seems that the proposed MDP models can adequately express the cluster tool behaviors in time variance environment. This kind of view makes it possible to look at the cluster tool from a slightly different perspective; hence we can continue to try scheduling cluster tools through reinforcement learning.

However, since the above model includes the remaining time in the state in order to consider the time variance, the state space can be very large. Furthermore, the environment settings we used in learning couldn't fully express the real world cluster tool behaviors. Therefore, the learning with more general environment settings is needed. To conduct such learning in further research, functional approximation seems to be applied due to the time elements in the states.

REFERENCES
Bellman, R., 1954. The theory of dynamic programming (No. RAND-P-550). RAND CORP SANTA MONICA CA.

Bertsekas, D. P., 1995. Dynamic programming and optimal control (Vol. 1, No. 2). Belmont, MA: Athena scientific.

Choi, B. K., & Kang, D., 2013. Modeling and simulation of discrete event systems. John Wiley & Sons.

Gosavi, A., 2003. Simulation-based optimization. parametric optimization techniques and reinforcement learning.

Jung, C., & Lee, T. E., 2012. An efficient mixed integer programming model based on timed Petri nets for diverse complex cluster tool scheduling problems. IEEE Transactions on Semiconductor Manufacturing, 25(2), 186-199.

Kim, D. K., Lee, T. E., & Kim, H. J., 2016. Optimal scheduling of transient cycles for single-armed cluster tools with parallel chambers. IEEE Transactions on Automation Science and Engineering, 13(2), 1165-1175.

Kim, J. H., & Lee, T. E., 2008. Schedulability analysis of time-constrained cluster tools with bounded time variation by an extended Petri net. IEEE Transactions on Automation Science and Engineering, 5(3), 490-503.

Lee, T. E., Lee, H. Y., & Shin, Y. H., 2004, December. Workload balancing and scheduling of a single-armed cluster tool. In Proceedings of the 5th APIEMS Conference (pp. 1-15). Gold Coast, Australia.

Lee, T. E., 2008, December. A review of scheduling theory and methods for semiconductor manufacturing cluster tools. In Proceedings of the 40th conference on winter simulation (pp. 2127-2135). Winter Simulation Conference.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S., 2015. Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.

Molloy, M. K., 1982. Performance analysis using stochastic Petri nets. IEEE Transactions on computers, 31(9), 913-917.

Moody, J., Wu, L., Liao, Y., & Saffell, M., 1998. Performance functions and reinforcement learning for trading systems and portfolios. Journal of Forecasting, 17(56), 441-470.

Murata, T.,1989. Petri nets: Properties, analysis and applications. Proceedings of the IEEE, 77(4), 541-580.

Puterman, M. L., 2014. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons.

Qiao, Y., Wu, N., & Zhou, M., 2012. Real-time scheduling of single-arm cluster tools subject to residency time constraints and bounded activity time variation. IEEE Transactions on Automation Science and Engineering, 9(3), 564-577.

Sutton, R. S., & Barto, A. G., 1998. Reinforcement learning: An introduction (Vol. 1, No. 1). Cambridge: MIT press.

Zuberek, W. M., 2004. Cluster tools with chamber revisiting-modeling and analysis using timed Petri nets. IEEE Transactions on semiconductor manufacturing, 17(3), 333-344.

HOLISTIC PLANNING OF PRODUCTION AND INTRALOGISTICS SYSTEMS THROUGH AUTOMATED MODELING WITHIN AND AMONG THE TOOLS OF THE DIGITAL FACTORY

David Weigert[a], Paul Aurich[b], Tobias Reggelin[c]

[a],[b],[c] Otto von Guericke University Magdeburg, Magdeburg (Germany)

[a]david.weigert@ovgu.de, [b]paul.aurich@ovgu.de, [c]tobias.reggelin@ovgu.de

ABSTRACT
The automated and semi-automated model generation has been discussed and developed for decades. With AutomationML (AML), an open, object-oriented, XML-based storage and exchange format is provided, which should also allow an exchange among visualization, construction and simulation tools. Separate cross-section functions and proprietary software solutions of the individual tools make it difficult to define a common transfer point. The presented concept and tool describes the development of an application-oriented and source-open middleware. The focus of the ongoing implementation phase is the development of these uniform, digital planning methods and tools by AML. The first prototype outlines the advantages but also the disadvantages of automatic model generation via AML. Up to now simple conveyor systems from straight, curved conveyor belts, rotary tables and single stations from the visualization tool Tarakos – taraVR can modelled automatically in the simulation tool Siemens – Plant Simulation and construction tool Autodesk – AutoCAD.

Keywords: automated and semi-automated model generation, AutomationML (AML)

1. INTRODUCTION
The versatility, speed and flexibility of the product creation process is increasing due to the ever-shorter product life cycles. This development has a direct impact on the participating logistics and production processes (Schenk 2014). The increasing digitization and automation accelerates the development of the continuity of available data models for the digital production and logistics. The use of digital tools in the areas of simulation, visualization and construction improve the quality of planning, increase efficiency and shorten the product development and launch (Schenk 2014; Daft 2016; Klepper 1996; Lüder and Schmidt, 2015). These benefits can only be fully used if it is enabled to translate all relevant and so far isolated digital methods and tools into an integrated planning system (Faltinski 2011; Schreiber and Zimmermann, 2011). Currently used tools cover only specific functional areas within the product lifecycle

management (PLM). This focus on individual areas of application of PLM enables a high level of specialization. The tools include the possibility of simulation, visualization and construction. However, the significance of these functions is limited with respect to the specialization of the users (Figure 1). This limited solution spaces are created by a lack of interface integration. An exchange of data in a heterogeneous system environment is therefore limited (Faltinski et al. 2012; Rawolle et al. 2002). In practice, therefore, no holistic, neutral and IT-based approach has been developed for the integrated digital planning and control of intralogistics systems and production areas. It lacks a neutral exchange format for continuous availability and mutual availability of simulation data, geometric design points and visualization elements.

Figure 1: Ability of visualization, construction and simulation in the different tools

The ongoing implementation phase discusses the concept and tool and first results of establishing continuous modeling between simulation, visualization and construction tools. The further course discusses the concept and tool for continuous model building on the example of the exchange direction of visualization to simulation and visualization to construction.

2. STATE OF THE ART AND SCIENCE
The standardization of systems, processes and their components is an essential element in the mastering of complexity as well as control and structuring of future, digital challenges (Drath 2010; Eigner and Stelzer 2009). The previous focus was the combined planning phases on a shortening in the time-to-market of products through the integration of product, process and production system planning, entitlement to use today is

at an early stage of reliable data from the product development process (Schenk 2014; Schenk and Schumann 2008). The idea of combined planning phases is described for years by modern and powerful tools (Dangelmaier 2013). These special cases are made of individually designed solutions. It lacks the holistic planning approach to take part in processes to promote an integration of all. First approaches for generating layout-based models can be found at Lorenz and Schulze (1995). First approaches to use structural as well as system data from a production planning and control system (PPS) to create models can be found at Splanemann et al. (1995). A first classification of automatic model generation approaches was provided by Eckardt (2002). The author distinguishes between parametric, structural, and hybrid approaches. Another way to classify approaches of the model generation makes the classification of model generation approaches to Straßburger et al. (2010). Early developments through digital planning and control was designed by the enterprise application integration (EAI) and service-oriented architectures (SOA) (Aier 2006; Kaib 2004; Bieberstein 2008). The EAI represents integrated business processes along the value chain. Enterprise applications of different generations and system architectures can interact in this framework over a common network. The SOA describes a method that encapsulates from existing IT components and coordinates. This existing services will be consolidated and summarized to a higher service. The objectives of EAI as also SOA are reducing costs in the development of production processes and increase the flexibility of business processes in the long term. The reason for the low acceptance and development of the methods is the high demands on insecure systems, data security, continuity of the tool development and product development process (Fay 2006; Raupricht et al. 2002). The interaction of different digital planning tools within the product life cycle is summarized often under the term "Digital Factory". The term describes a comprehensive network of digital methods and models, including simulation and 3D visualization. Its purpose is the integrated planning, management and implementation, as well as a steady improvement in all key factory processes and resources (VDI 2008; Wenzel et al. 2003). Also here is a link of different planning tools. However, the use of continuous planning tools is missing. Weigert (2015) describes the first opportunity of an implemented middleware for the automatic data exchange among different tools of the digital factory by AML. The motivation for the use of a common concept and tool can be reducing costs for the planning, control and operation and maintenance of plants and factories. First options for the automatic generation of the model are described. However, no procedure is known to reach the current level which combines three essential tools for the digital factory by an open interface.

2.1. Storage and Exchange format

For the semi- and fully automated model generation is the origin and use of the data and information utmost importance (Bergmann 2014). The currently most popular standards for the automatic model generation are the data formats simulation data exchange format (SDX) (Sly and Moorthy 2001) and Core Manufacturing Simulation Data (CMDS) (Lee 2015; Bergmann et al. 2010). The hierarchically structured SDX format is used to exclusively provide layout information. With the open-source, XML-based format of CMDS can both layout - as also process-related information transmitted. The problem of implementation of comprehensive control and routing strategies and complex system behavior is not completely solvable but manageable with these data formats (Bergmann, 2014; Bergmann et al. 2010). By AutomationML (AML), an open source, free available, object-oriented, XML-based storage and exchange format is being developed. After initial evaluations of different exchange formats, Daimler AG initiated the development and standardization of AutomationML as an intermediate format of the digital factory together with ABB, KUKA, Rockwell Automation, Siemens, netAllied and Zühlke, as well as the University of Karlsruhe and Otto von Guericke University Magdeburg in October 2006. In 2009, the previously closed industrial consortium opened by establishing an association. The first new member was the Fraunhofer IOSB. It becomes clear that the efforts to use AML are driven by the industrial and scientific location of Germany. AML owns the technical requirements for the modeling of production, intralogistics and robotic systems that is used but so far mainly in the area of virtual commissioning. Fundamentally AML linking role profiles (Hoernicke et al. 2016; Hundt et al. 2009; Lüder and Schmidt 2015). So far, topology, geometry, kinematics and behavior of system components can be described with AML. The hierarchical picture of the topology of the subject of the planning is carried out by means of Computer Aided Engineering Exchange (CAEX). The CAEX library concept includes three types of library (Drath 2010):

- The **SystemUnitClass** library is a catalog of concrete physical or logical system objects or their combination. Attributes, interfaces and nested internal elements and their compounds are assigned to the elements
- The **RoleClass** library defined abstract physical or logical system objects, regardless of the actual technical realization. Roles describe the functioning of investment properties
- The **InterfaceClass** library describes the kind of interfaces between the system objects. The relations between investment objects are mapped

Geometry and kinematics can be associated with individual system components through COLLADA files. The control is defined by PLCopenXML and describes the system behavior. AML is adaptable and flexible, it offers the possibility to include other XML formats (Hundt et al. 2009). In addition, the AML format has an inherent distributed data structure. The information is instead of a monolithic XML document, saved as separate documents. The reusability of individual system components and the development of element libraries will be easier (Lüder and Schmidt 2015).

3. APPLICATION GOALS AND CONCEPTUAL DESIGN

A continuous model generation between simulation, visualization and construction tools for the integrated planning of production and intralogistics systems describes the linking of the three digital tools. The expertise in dealing with the tools than the substantive complexity of each tool is highly classified. The presented concept and tool describes the development of application-oriented and open source middleware. The goal is to develop of uniform, digital planning methods and tools for a consistent design. The lossless and accelerated conversion and modelling within the various tools is in the focus. On reason of the different core functions of each tool, as well as their proprietary interfaces, the goal is to develop an open source automated import and export solutions (Figure 2).

Following advantages arise from the use of an automatic exchange system:

- Existing simulation, visualization, and construction tools in the company remain in place, preventing costly new investments
- Productivity and cost reduction can be achieved by the use of development, combining the individual benefits of the tools
- Visualization, modeling, and simulation of real-world intralogistics systems accelerates, reducing the largely manual and costly effort in creating a new model

At present, many software solutions have only limited access to the ability to offer planning data in a heterogeneous system landscape.

This fact is due to the lack of conformity of the interfaces and the resulting lack of integration of all necessary planning data. The reasons for these so-called insellations have been and are the efforts of machine manufacturers to exclude competing products by means of ever new proprietary interfaces in favor of their own "complete solutions". A circumstance that today is hardly accepted by customers. Furthermore, the lack of networking between the planners and partners involved in the process contributes to the existing problem. In order to support the definition of the mapping rules, a graphical interface is developed which contains the following features:

- Visualization of the data model of the source system in a tree structure
- Selection of pre-made rule templates for repetitive application forms:
- *1:1 mapping* (Due to a type attribute; On the basis of a checked attribute value)
- *m:1 mapping* (Illustration of a group with the same or similar attribute value; Illustration of a group by means of object relations (parent-child, Sibling)
- *0:n mapping* (Creating new node with no equivalent in the source model)
- Memory function for the created rule sets with additional information:
  - Author
  - Version
  - Description
  - Source-Tool
  - Target-Tool

The exchange system is not limited to the generation of exchange files but can also be used for exchanging data between current network instances if the tools involved permit such an integration. The concept and tool for the continuous and comprehensive model building is appropriate on the example of the simulation tool Plant Simulation – Siemens and construction tool AutoCAD – Autodesk. The visualization tool taraVR – tarakos serves as a starting point for the investigations.
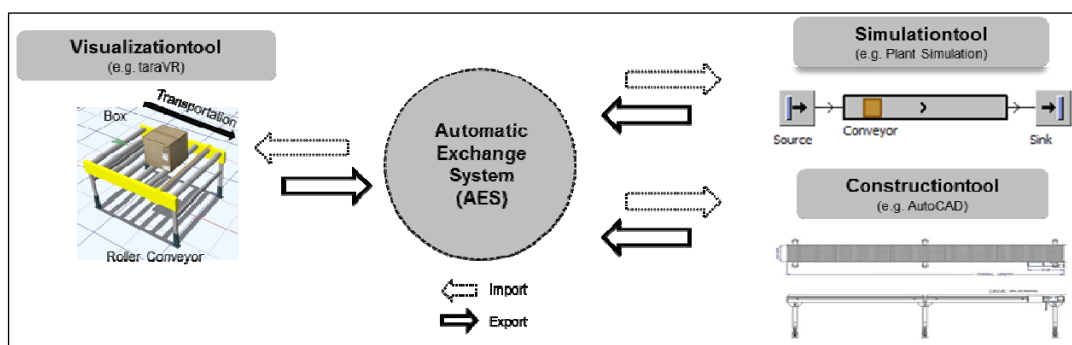

Figure 2: Overall concept of development

4. CONTINUOUS MODELING BETWEEN SIMULATION, VISUALIZATION AND CONSTRUCTION TOOLS

The Automatic Exchange System (AES) forms the basis for the common data and information exchange. The AES is defined by a system of mapping rules and an associated user interface. The model item get its role profiles from the developed AutomationML roll library within the Exchange System. In addition to the transmission of model elements it is important to transfer their attributes such as capable of taking over transport direction, speed, and more state descriptions. Furthermore, the relationships between the elements must be transferred. The approach of an open source standards about the AML function narrows the existing interface packages and fee-based libraries in the commercial product suites for the exchange of data. Goal is to realize independent modeling of closed software packages. Based on the classification ability of model generation approaches to Straßburger et al. (2010) the represented development can be classified as follows:

- Application: parallel (tactical) planning
- Focus: conveyor systems
- Degree of automation: semi-automated
- Approach: direct generic structure
- Support creation of model
- Interfaces: Text - and XML-based

Before models can be automatically generated it is necessary that the modules and elements of different tools are mapped. Data mapping is the process of mapping data elements between different data models and thus constitutes a fundamental step in the information and data integration (Alexe et al. 2008). A so-called rule interpreter is designed to enable this mapping. From the respective visualization, construction, simulation tools the libraries with all modules and their parameters are exported and mapped. This manner defines rules for the transfer. If current rules exist, the import file for the target tool can be generated by the rule interpreter emanating from an export file of source tool automatically.

The import file is loaded using the created interfaces or corresponding data interpretation in the target tool and interpreted. To integrate the relevant elements of the model and to interpret is a systematization of elements of importance for further processing. After the reduction for the respective tools it is important to define the requirements for the interfaces. The needs and situations of the interfaces between the designed EAS, the simulation tools as well as design tool have been analyzed and defined.

4.1. Automated model generation – Construction

The Autodesk AutoCAD tool has considerably more modification options in the area of the interface creation. Through the import of other CAD tools with the formats 3D Studio (.3ds), Autodesk Inventor (.ipt; .iam), SolidWorks (.prt; .sldprt; .asm; .sldasm) and step (.stp; .ste; .step) is there quite a broad base. For the development and integration of and into the EAS is the variety but not of importance. The neutral exchange format AML forms the basis of the development. For this purpose, separate commands are developed for import and export. The goal of consistency to the visualization tool ensures that the interface in the format programmed .net and used. Determined at a minimum for the required parameters of the construction tool for the communication with the AES include:

- Object information: name, geometry
- Layout information: location (connection are not necessary)
- Detailed information on geometry are not needed, can be optional attached in an AML file through the COLLADA format

A direct import AML files is not possible and is not supported by the Autodesk software suite. AutoCAD is used within the development as a two-dimensional representation of intralogistics planning and construction data.
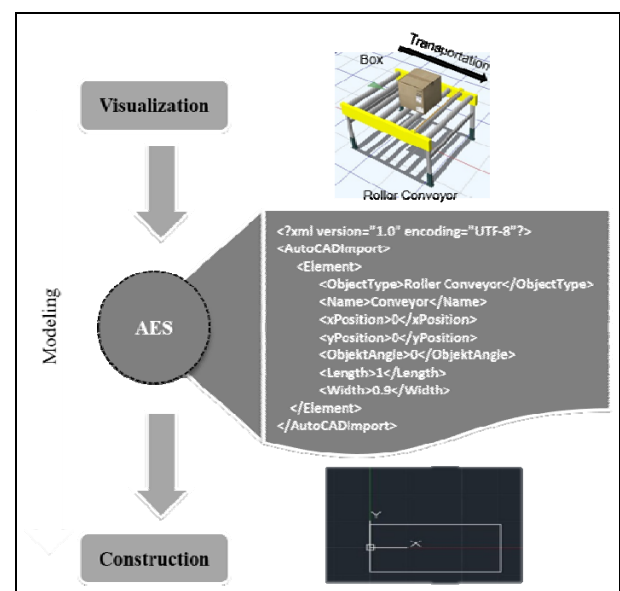

Figure 3: Model generation between visualization and construction tools via AES

Not only static blocks are addressed within the tool. The blocks are predefined two-dimensional body by the user. They are stored in a block table. Hereby the purpose a dynamic adaptation of the body within the construction tool to get.

A pure one-to-one relationship of modules is no longer necessary. Blocks to dynamically build and parameterize the interface with c#.NET was programmed and encapsulated in a Dynamic Link Library (.dll) file. The enclosed program environment is performed as AES.dll.

Algorithm 1: Pseudocode for Modeling – Construction

```
define fcn = rootnode.childnodes and
       scn = FCN.childnodes
open xmlfile
while rootnode has fcn
    read SCN.objecttyp, SCN.x–pos, SCN.y–pos
    create object from blocktable.objecttyp at
            x–pos.,y–pos.
    if object is dynamic
        for i = 1 to SCN.parameter.count
            read SCN.parameter(i) and
                    parameter(i).value
            set object.parameter(i)=parameter(i).value
        next
    end if
end while
close xmlfile
```

After the call of the encapsulated .dll file it is possible to run various commands. The pseudo code describes the reading and processing of the XML file in the construction tool (Algorithm 1). The XML file is created from the AES and read in the target tool. In contrast to the simulation tool, the structure can be an XML file because the tool can work with XML and .dll file to process the models (Figure 3). The file is built in a common XML format. The linking of several objects is ensured by the updating of the source code. By working with dynamic blocks, the parameters can be arbitrarily varied according to the read-in to the construction tool. This is the processing of dynamic blocks. The blocks are user-defined two-dimensional bodies. They are stored in a block table. The aim of the invention is to obtain a dynamic adaptation of the bodies within the construction tool. For a static block, the dimensions are invariable. It is therefore stored with fixed dimensions in the table and can only be generated with these dimensions again. In a dynamic block, the user defines individual elements of the block as variable.

4.2. Automated model generation – Simulation

For the development of the simulation tool, event-oriented simulation approach from Plant Simulation by Siemens has been used. "Plant Simulation" has a variety types of license and fee-based libraries. Costly licenses such as "Professional" in combination with the "Interface Package" have been avoided for the development of a common data interface. The selection would not meet the main goal of an open-source and neutral communication interface. The license type of "Standard" forms the basis for the interface between the

EAS and the simulation tool. An advantage is the upward compatibility of the simple data structure (.txt file). A disadvantage is the complex data processing in the AES before import and export in the simulation tool. The strings must correspond to a given form can be accurately encoded and decoded. As a result is a text file as opposed to a binary file without the use of special programs to read and can be viewed with a simple text editor and edited. Determined at a minimum for the required parameters of the simulation tool for the communication with the AES include:

- Object information: name, type, geometry
- Layout information: location and connection to other elements
- Material flow parameters: time usage, routing

A direct import of AML files is not possible and is not supported by the selected simulation tool. A master file was developed for the simulation tool, which contains all requirements needed for the model creation. The system works according to a hierarchical management of access to a common resource as a master/slave system. The master file allows only the backup model derivatives and contains all the required methods to model creation, export model, library export and import of the modules. A new generation section of source code is generated for each previously mapped module. Sections for identical statements, for example the combination of elements, can be applied. The text file is read from the AES into the simulation tool.
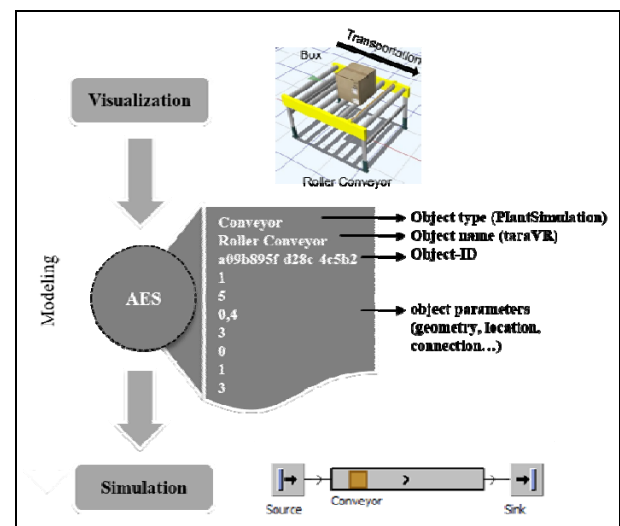

Figure 4 Model generation between visualization and simulation tools via AES

The file contains a hierarchical structure, which allows to process all objects equally. Changes to entire model elements or parts thereof can thus be easily incorporated. The file is integrated by the integrated data interface. This has resulted from the use of general interfaces.

The content of the text file is deliberately simple. The file is divided as follows. The object type is determined by the target tool and is part of the library in the simulation tool. The object name describes the name of the object in the source tool. The identification number of the object is a one-to-one number sequence that identifies the object in the model. Further details describe object parameters such as geometry, location, connection and speed. For multiple objects, the ID also describes the auto-linking points in the model. After conversion from the visualization tool into the AES, a text file is created (Figure 4). This is read in the simulation tool by a simple data interface. The necessary intelligence of the simulation tool is generated by the programmed interpretation of the source code (Algorithm 2). The programmed method reads the text file and checks whether the object exists in the simulation tool library. If the response is positive, the method begins to link the object to the element from the simulation library. After this, the method parameterizes the created object according to the specifications of the text file. Because the entire intralogistics system is present in a text file, the objects are separated by a separator. The unique ID knows all objects predecessors and successors. If the predecessor or successor is missing, the source or sink is created.

Algorithm 2: Pseudocode for Modeling – Simulation

```
open textfile
while not textfile.end
read bibliothek, objecttyp, x-pos., y-pos., name
create object from bibliothek.objecttyp at x-pos.,y-pos.
    if next textfile.line ≠ seperator
        do
        read parameter and parameter.value
        case parameter of
        case "input-id"
            if object.exists(input-id)
            connet object with input-object
            end if
        case "output-id"
            if object.exists(output-id)
                connet object with output-object
            end if
        else
            set object.parameter = parameter.value
        end case
        until next textfile.line ≠ seperator
    end if
endwhile
close textfile
```

The flexibility of the model creation is given by scanning the entire simulation library in the master file at start. This long term causes that even individual building blocks of the end user can be read out and dynamically addressed.

5. SUMMARY AND OUTLOOK

The described concept and tool represents a comprehensive approach for the automatic and open-source modeling and data transmission between different tools. The development describes a possibility to repeatedly use individual AML libraries for simulation, visualization and construction. This can be ensured by an AML role library. First results show that the implementation is successful. Similarly, no proprietary software solutions are used. With the possibility to combine modeling within the different tools it can be possible to produce precise and fast models from a single source. As a result, costly, error-prone and time-intensive re-modeling is already avoided. The development level reached the basis for the transfer of the models into the corresponding target tool, described by simulation and construction. Further steps are the increasing detailing and standardization of the automated exchange system. In the future, the question about the implementation and transmission of the work plans and the accompanying control and the routing of the elements in the simulation tool will be explored. Up to now simple conveyor systems from straight, curve conveyor belts, rotary tables and individual stations from the visualization tool Tarakos - taraVR can be modeled automatically in the simulation tool Siemens Plant Simulation and construction tool Autodesk - AutoCAD. The use of cost-intensive and optional interface libraries can be dispensed. Due to the simplicity of the development it is possible to connect arbitrary tools. To this end, investigations are currently being carried out to prepare the interface to other vendors' tools. Similar advantages are promised. It was necessary to use only existing licenses throughout the development. Only the internal specifications of the tool were used. No license rights of the providers were infringed.

The cooperation with end-users from the practice makes clear what additional requirements on the range of functions the AES must solve in future. The question about the implementation and transmission of the work scheduling and the accompanying control and the routing of the elements in the simulation tool are also to be investigated. The repatriation of information and models of the simulation and construction tool in the visualization tool be made manually in the AES at the current time. Target is to process the data from the construction through renewed data mapping in the AES. Here, the AAS repeatedly assumes the role of the rule interpreter and plausibility auditor. The integration of the event log obtained by the simulation runs is conceived for the simulation. The described and planned scope and procedures are to be simplified and made more dynamic in the future. At the same time, a significant increase in the user-friendliness in handling the tool is planned.

REFERENCES

Aier, S. (ed.): Enterprise application integration: Serviceorientierung und nachhaltige Architekturen. Berlin: GITO-Verl. 2006.

Alexe, B.; Chiticariu, L.; Miller, R.J.; Tan, W.-C.: Muse: Mapping Understanding and deSign by Example. In: IEEE 24th International Conference on Data Engineering, 2008, Cancun, Mexico, 7/4/2008 - 12/4/2008, 2008, p. 10–19.

Bergmann, S.: Automatische Generierung adaptiver Modelle zur Simulation von Produktionssystemen. Ilmenau: Univ.-Verl. 2014.

Bergmann, S.; Fiedler, A.; Straßburger, S.: Generierung und Integration von Simulationsmodellen unter Verwendung des Core Manufacturing Simulation Data (CMSD) Information Model. In: Integrationsaspekte der Simulation: Technik, Organisation und Personal: Generation and integration of simulation models using the Core Manufacturing Simulation Data (CMSD) information model. Karlsruhe: KIT Scientific Publ 2010, p. 461–468.

Bieberstein, N.: Executing SOA: A practical guide for the service-oriented architect. Upper Saddle River, NJ: IBM Press/Pearson plc 2008.

Daft, R.L.: Organization theory & design. Boston, MA: Cengage Learning 2016.

Dangelmaier, W.: Fertigungsplanung: Planung von Aufbau und Ablauf der Fertigung ; Grundlagen, Algorithmen und Beispiele. Berlin [u.a.]: Springer 2013.

Drath, R. (ed.): Datenaustausch in der Anlagenplanung mit AutomationML: Integration von CAEX, PLCopen XML und COLLADA. Berlin: Springer 2010.

Eckardt, F.: Ein Beitrag zu Theorie und Praxis datengetriebener Modellgeneratoren zur Simulation von Produktionssystemen. Aachen: Shaker 2002.

Eigner, M.; Stelzer, R.: Product Lifecycle Management: Ein Leitfaden für Product Development und Life Cycle Management. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg 2009.

Faltinski, S.: A dynamic middleware for real-time automation systems. Lemgo, Hannover: InIT; Technische Informationsbibliothek u. Universitätsbibliothek 2011.

Faltinski, S.; Niggemann, O.; Moriz, N.; Mankowski, A.: AutomationML: From data exchange to system planning and simulation. In: Industrial Technology (ICIT), Athen, 2012, p. 378–383.

Fay, A.: Reduzierung der Engineering-Kosten für Automatisierungssysteme. Industrie Management (2006) 22, p. 29–32.

Hoernicke, M.; Messinger, C.; Arroyo, E.; Fay, A.: Topologiemodelle in AutomationML. Atp-Edition : automatisierungstechnische Praxis ; Organ der GMA (VDI-VDE-Gesellschaft Meß- und Automatisierungstechnik) und der NAMUR (Interessengemeinschaft Automatisierungstechnik der Prozessindustrie) 58 (2016) 5/2, p. 28–41.

Hundt, L.; Lüder, A.; Barth, H.: Anforderungen an das Engineering durch die Verwendung von mechatronischen Einheiten und AutomationML. SPS/IPC/DRIVES 2009 (2009), p. 341–349.

Kaib, M.: Enterprise Application Integration: Grundlagen, Integrationsprodukte, Anwendungsbeispiele. Wiesbaden: Dt. Univ.-Verl. 2004.

Klepper, S.: Entry, exit, growth and innovation over the product life cycle. Estados Unidos: American Economic Review 1996.

Lee, Y.-T.T.: A Journey in Standard Development: The Core Manufacturing Simulation Data (CMSD) Information Model. Journal of research of the National Institute of Standards and Technology 120 (2015), p. 270–279.

Lorenz, P.; Schulze, T.: Layout based model generation. In: Lilegdon, W.R. (ed.): Proceedings of the 27th conference on Winter simulation, Arlington, VA, USA, 3-6 Dec. 1995, 1995, p. 728–735.

Lüder, A.; Schmidt, N.: AutomationML in a Nutshell. In: Handbuch Industrie 4.0 : Produktion, Automatisierung und Logistik. Wiesbaden: Springer Fachmedien Wiesbaden 2015, p. 1–46.

Raupricht, G.; Haus, C.; Ahrens, W.: PLT-CAE-Integration in gewerkeübergreifendes Engineering und PlantMaintenance. atp – Automatisierungstechnische Praxis (2002) 44 (2), p. 50–62.

Rawolle, J.; Ade, J.; Schumann, M.: XML als Integrationstechnologie bei Informationsanbietern im Internet. Wirtschaftsinformatik 44 (2002) 1, p. 19–28.

Schenk, M.: Fabrikplanung und Fabrikbetrieb: Methoden für die wandlungsfähige: Springer Berlin Heidelberg 2014.

Schenk, M.; Schumann, M.: Interoperable Testumgebung für verteilte domänenübergreifende Anwendungen. In: Scholz-Reiter, B. (ed.): Technologiegetriebene Veränderungen der Arbeitswelt. Berlin: GITO-Verl. 2008, p. 155–169.

Schreiber, W.; Zimmermann, P.: Virtuelle Techniken im industriellen Umfeld: Das AVILUS-Projekt - Technologien und Anwendungen: Springer Berlin Heidelberg 2011.

Sly, D.; Moorthy, S.: Simulation data exchange (SDX) implementation and use. In: Peters, B.A. (ed.): Proceedings of the 2001 Winter Simulation Conference, Arlington, VA, USA, 9-12 Dec. 2001, 2001, p. 1473–1477.

Splanemann, R.; Roth, M.; Soravia, S.: Einsatz der Materialflußsimulation zur Planung, Analyse und Optimierung von verfahrenstechnischen Produktionsanlagen. Chemie Ingenieur Technik 67 (1995) 9, p. 1107–1108.

Straßburger, S.; Bergmann, S.; Müller-Sommer, H.: Modellgenerierung im Kontext der Digitalen Fabrik-Stand der Technik und Herausforderungen. Proceedings der 14 ASIM-Fachtagung Simulation in Produktion und Logistik (2010), p. 37–44.

VDI: Digitale Fabrik: Grundlagen ; VDI-Richtlinien ; VDI 4499, Blatt 1. Berlin: Beuth 2008.

Weigert, D.: Automated exchange system between simulation, visualization and construction tools: The 15th International Conference Modeling and Applied Simulation, MAS 2016: p. 112-120 .

Wenzel, S.; Hellmann, A.; Jessen, U.: e-Services - a part of the "Digital Factory". In: Bley, H. (ed.): Proceedings / 36th CIRP International Seminar on Manufacturing Systems, 2003, p. 199–203.

AUTHORS BIOGRAPHY

DAVID WEIGERT studied Industrial Engineering with specialization in Logistics at the Otto-von-Guericke-University Magdeburg. He became a research associate at the Chair Logistical Systems at the Otto-von-Guericke-University Magdeburg and scientific project assistant at the Fraunhofer Institut for Factory Operation and Automation IFF Magdeburg. His areas of competence are the analysis and optimization of logistics processes, as well as modelling, simulation and optimization of logistics systems.

PAUL AURICH holds a bachelor's degree in Industrial Engineering in Logistics. Currently he is studying mechanical engineering at the Master's degree. His research interests include the modeling, simulation and mathematical optimization of logistics systems.

TOBIAS REGGELIN is a research and project manager at Otto von Guericke University Magdeburg and the Fraunhofer Institute for Factory Operation and Automation IFF. He received a doctoral degree in engineering from Otto von Guericke University Magdeburg. His research interests include modeling and simulation of production and logistics systems and the development and application of new modeling and simulation methodologies.

OPTIMIZATION OF THE LOGISTICS PROCESS IN WAREHOUSE OF AUTOMOTIVE COMPANY BASED ON SIMULATION STUDY

Bronislav Chramcov, Milan Jemelka

Tomas Bata University in Zlín, Faculty of Applied Informatics,
Nad Stráněmi 4511, 760 05 Zlín, Czech Republic

chramcov@fai.utb.cz

ABSTRACT

The paper focuses on solving complex warehouse simulation to achieve effective solution. Most attention is mainly paid to optimization of the warehouse operation, especially logistics process of warehouse. The aim of the simulation study is to verify if inbound and outbound deliveries are optimal to cover all the requirements of the warehouse in time and in the right amount. Subsequently, experiment with the simulation helps to formulate recommendation which would improve the processes and reduce total cost. The Witness simulation environment is used for modeling and experimenting. The simulation experiments are evaluated per the total cost for stock transfers and storage costs. Description of the proposed simulation experiments and evaluation of achieved results are presented.

Keywords: warehouse, optimization, modelling, simulation, computer simulation, optimization methods, Witness

1. INTRODUCTION

Warehouses are constantly referred to as cost centers and rarely adding value in the best. Operations efficiency is the key to the success of every company that processes inventories. Management is under constant pressure to reduce the time between customer order and customer delivery. A customer use order-to-delivery time as factor in deciding on a vendor. Therefore, companies must use effective inventory management process to reduce this time to minimum (Curcio and Longo, 2009). When efficiency is low, material may not arrive at customer warehouse on time, orders can get lost, and low stock levels can result in shortage. Companies constantly strive to increase the performance and reduce the costs of production. Here are some recommendations on how to optimize production processes. The functional optimization encounters and accomplishes efficiency and effectiveness of a warehouse process.

The concept of supply chain management entails the consideration and management of logistical processes along the entire supply chain, which includes suppliers, customers, and consumers. Despite the implementation of new enterprise resource planning programs, e-commerce, just-in-time delivery, Kanban, efficient both direction electronic communication aims on shortening the supply chain downtimes. The goal of stock optimization is to harmonize and optimize the processes within the supply chain to reduce the stocks in the entire supply network. Functional optimization inquires to improve efficiency and effectiveness of a warehouse process.

Warehouses are still a typical and central feature in most supply chain due to the partial implementation of lean and agile philosophies. Organizations need to discover ways to effectively manage and perform the operations inside a warehouse with much efficiency and in turn reduce the storage time and costs involved in the storage. There is a need to optimize the technology, operation and the manpower to get good results and high efficiency (Kare et al., 2009). An overview related to warehouse optimization problems presents (Karasek, 2013). Author shows the current state of the art in optimization in three groups of interest in logistic warehouses and distribution centers.

There are many picking, storing, or routing policies. The research (Petersen and Aase, 2004) examines several picking, storing, or routing policies simultaneously to determine which process decisions affect performance the most. Storage policies, which assign stock keeping units (SKUs) to storage locations, generally fall into three broad categories. SKUs may be assigned randomly, grouped into classes with similar SKUs that are placed in the same area of the warehouse, or assigned to a location based on demand or volume.

Inventory classification using ABC analysis is effective way managing materials. ABC classification allows an organization to partition stock keeping units (SKU) into three groups: A, the most important; B, important; and C, the least important (Muppani et al., 2010). The major advantage of ABC analysis is the simple usability of this method. For inventory items, the criterion is frequently the annual cost. In this paper, we are concerned with the ABC classification stock aimed to facilitate inbound deliveries optimally (Bottani et al., 2015).

Recent investigations also reveal that about 33 per cent of logistical costs can be attributed to the costs arising

in inventory management and therefore, a proper investigation of savings that might be achieved within this part of supply chain is necessary and is in many cases profitable (Raidl&Pferschy, 2010). An extensive review on warehouse operation planning problems is presented in (Gu, Goetschalckx, McGinnis, 2007). This paper provides a detailed discussion on warehouse operation-planning methods together with warehouse design, computational systems, and case studies. Moreover, in the work (Baker, Canessa, 2009), the current literature on the overall methodology of warehouse design is explored, together with the tools and techniques used for specific areas of analysis. The output is a general framework of steps, with specific tools and techniques that can be used for each step. This is intended to be a value to practitioners and to assist further research into the development of a more comprehensive methodology for warehouse design.

2. PROBLEM FORMULATION

The aim of the paper is to create a simulation study of the warehouse operation in an automotive company with usage of the computer simulation. Purpose of this simulation is finding the optimal solution of a logistics process control in the warehouse. The warehouse is based on the ABC method. The well-known optimization methods are used for searching the effective solution as quickly as possible.

The simulated model is based on the ABC warehouse management. The ratio 70:20:10 is assumed. The aim is to determine the effective values of inter arrival time of truck loaded part of type A (IAT_A), inter arrival time of truck loaded part of type B (IAT_B) and inter arrival time of truck loaded part type C (IAT_C). It is considered the constant time of unloading units from all parts of the warehouse. Therefore, the inter arrival time of truck loaded part of type A and B can be considered in accordance with the ratio A:B:C (70:20:10) and can be expressed on the base of IAT_C in the form (1).

$$IAT_A = (C \cdot IAT_C)/A$$
$$IAT_B = (C \cdot IAT_C)/B \qquad (1)$$

At the same time, it is necessary to determine the minimum initial number of SKUs in each part of the warehouse to avoid the removal of some parts in the warehouse.

The objective function is total cost of warehouse and this function is possible to define in the form (2). Total cost consists of the storage cost, the transport cost and cost penalty for the removal of some parts in the warehouse.

The storage cost is possible to define in the form (3), where $T_i^{storage}$ is the storage time of i-th SKU. Cost price for storage of one SKU is one CZK per each unit time. The transport cost is possible to express in the form (4), where N_x^{load} is number of unloaded truck with

particular type of part and CR^{truck} is cost rate of used type of truck. The amount of the penalty is considered in the form (5). It is dependent on the number of times - N^{zero} that part of the warehouse has been emptied.

$$Cost^{total} = Cost^{storage} + Cost^{transport} + Cost^{penalty} \qquad (2)$$

$$Cost^{storage} = \sum_{i=1}^{N} T_i^{storage} \qquad (3)$$

$$Cost^{transport} = (N_A^{load} + N_B^{load} + N_C^{load}) \cdot CR^{truck} \qquad (4)$$

$$Cost^{penalty} = 0.1 \cdot N^{zero} \cdot Cost^{storage} + Cost^{transport} \qquad (5)$$

3. MODEL CONSTRUCTION IN WITNESS SIMULATION ENVIRONMENT

For building up the model and subsequent implementation of the proposed experiments, it is possible to use a wide range of simulation programs and systems (Banks 2005). The Witness system environment was available in this case. This system is offered by the Lanner Group and it contains many elements for discrete-part manufacturing. Witness models are based on template elements. These may be customized and combined into module elements and templates for reuse. The standard machine elements can be single, batch, production, assembly, multistation or multicycle. Other discrete modelling elements include multiple types of conveyor, tracks, vehicles, labor and carriers. The behavior of each element is described on a tabbed detail form in the Witness user interface. The models are displayed in a 2-D layout animation with multiple windows and display layers.

The simple model is created for simulation study of our problem. Supply of individual parts of the warehouse is modelled in the Witness environment with help of the element Machine of Production type. The truncated normal distribution is used for modelling of cycle time of each machine. The parameters are set up according to Table 2. The element buffer is used to model individual parts (A, B, C) of the warehouse. Capacity of each buffer is 5000.

The module Witness experimenter is used to search the best solution. In Witness the objective function is set inside the simulation model. The objective function is set according the form (2). This module offers simple analysis of the experiments to determine the variability of typical runs or optional tracking for any other parameters that may be of interest in the results set from the simulation. The experimenter offers a wide choice of options (methods) for experimentation. Some methods are used to experiment in our simulation study. The results according random solution, Min/Mid/Max method, Hill Climb simple algorithm and Adaptive Thermostatistical SA algorithm are compared.

4. DESCRIPTION OF THE SIMULATION EXPERIMENTS

The inter arrival time of truck loaded with part of x-type (IAT_x) is not constant. Time of truck arrival is modeled by truncated normal distribution. This distribution has four parameters (mean, standard deviation, minimum, maximum). Three real situations are considered.

1. The inter arrival times of trucks are kept very strictly.
2. The arrival delay of the truck is not more than 20% of the inter arrival time.
3. The delay of arrival of the truck can reach up to 50% of the inter arrival time.

Setting the truncated normal distribution parameters for experiments defined above is clearly shown in the Table 1. The probability density function of truncated normal distribution with parameters of experiment no.3 and for IAT_x=50 is presented in the Figure 1.

Table 1: Parameters of the truncated normal distribution

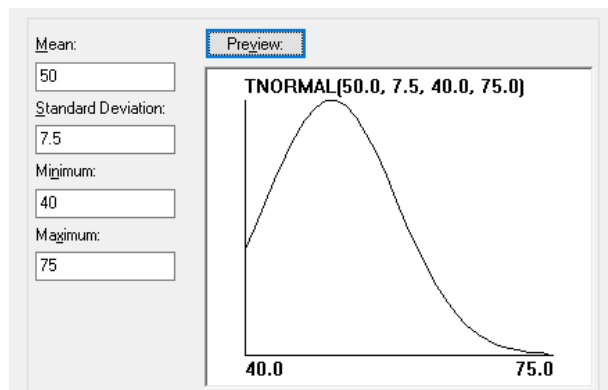| Parameter	Experiment No.1	Experiment No.2	Experiment No.3
Mean	IAT_x	IAT_x	IAT_x
Standard deviation	$0.05*IAT_x$	$0.1*IAT_x$	$0.15*IAT_x$
Minimum	$IAT_x - 0.2*IAT_x$	$IAT_x - 0.2*IAT_x$	$IAT_x - 0.2*IAT_x$
Maximum	$IAT_x + 0.2*IAT_x$	$IAT_x + 0.2*IAT_x$	$IAT_x + 0.5*IAT_x$

Figure 1: Sample of the probability density function of truncated normal distribution

Each of the above experiments is applied to different types of transport. The goal of this simulation study is to select the best means of transport for given conditions. Three different vehicles are assumed in the experiments. There is a small truck, a standard truck and a jumbo truck. Each of these vehicles has a cargo capacity. This capacity is measured by SKU. Each of the vehicles has its own cost rate per kilometer. Cargo capacity and cost rate for assumed type of vehicle are presented in the Table 2. These parameters are specified based on reality. Distance for all the vehicles is determined as an average value of 100 km.

Table 2: Parameters of assumed vehicles

Type of vehicle	Cargo capacity [SKU]	Cost Rate [CZK/km]
Small truck	50	15
Standard Truck	100	18
Jumbo Truck	150	19

5. RESULTS OF SIMULATION EXPERIMENTS

The model created in Witness environment was verified so it could be used for simulation experiments. The module Witness experimenter is used for analysis of the experiments.

The main goal is to find suitable warehouse parameters to minimize warehouse operation costs. The inter arrival time of truck loaded with part of type C (IAT_C) and initial number of SKUs in each part (A, B, C) of the warehouse ($N_A^{initial}$, $N_B^{initial}$, $N_C^{initial}$).

Various optimization methods are used for searching the best solution. The results according Min/Mid/Max method, Hill Climb simple algorithm and Adaptive Thermostatistical SA algorithm are compared. Especially the possibilities of using the random solution algorithm are outlined. This algorithm provides a fast, very good solution that is not optimal, but very close to effective solution. For each method except Min/Mid/Max method, 250 scenarios are generated. The model is running for 5 replications (with 5 different datasets). Model is simulated for stock time units (STU). The stock time units' value is 1000 items. Adapt into real world, its one working month.

The minimal value of the total costs of warehouse is searched during the simulation experiments. This value is calculated according to the formula (2). In addition to the objective function's value, other parameters are further measured. The value of maximum number of SKUs in the warehouse (N^{max}) is measured.

Also, the number of the empty warehouse times (N^{zero}) is measured.

5.1. Simulation experiments for small truck

Firstly, the experiments for small truck are carried out. The cargo capacity of this truck is 50 SKUs. It is considered the constant time (0.1 STU) for unloading one unit from any parts of the warehouse as mentioned above. It means that the 50 SKUs are picked from the warehouse part C for 50 STU while respecting 70:20:10 ratio and maintaining the sufficiency of the given warehouse part. Therefore, the constraint for parameter of IAT_C is set to a very small interval $\langle 48,52 \rangle$ for all experiments with small truck.

The constraints for initial number of SKUs in specific part of the warehouse are set for each experiment separately. The constraint intervals have been set based on simple model experiments. Specific settings for individual experiments are specified in the Table 3.

Table 3: Constraints for initial number of SKUs in specific part of warehouse for small truck

Part of warehouse	Constraints interval		
	Experiment No.1	Experiment No.2	Experiment No.3
A	⟨100,250⟩	⟨150,350⟩	⟨200,500⟩
B	⟨70,150⟩	⟨70,150⟩	⟨70,200⟩
C	⟨50,100⟩	⟨50,100⟩	⟨50,150⟩

The results of experiments defined above are presented in the Table 4, Table 5 and Table 6. The three best results searched by random method and the best result found according the other methods are shown in the tables. The presented results represent the average values of the five replications that are performed within the simulation experiment. Therefore, the values in the last two columns are not integers. The best result for each experiment is highlighted.

Table 4: The results of experiment No.1 for small truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	515 534,41	50	130	82	82	373,8	0
	2	525 115,38	50	132	88	84	383,8	0
	3	537 428,11	50	158	78	80	395,8	0
Hill Climb		523 959,33	50	126	70	60	353,8	0,8
Adaptive Thermostatistical SA		513 848,01	50	136	80	76	371,8	0
Min/Mid/Max		550 303,96	50	174	70	74	397,8	0,2

Table 5: The results of experiment No.2 for small truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	557 918,25	50	198	74	86	451,6	0
	2	592 959,68	50	230	70	80	476,4	0,2
	3	598 447,65	50	194	104	76	473,0	0,4
Hill Climb		594 623,75	50	194	70	60	455,2	1
Adaptive Thermostatistical SA		559 686,46	50	194	74	92	453,6	0
Min/Mid/Max		618 319,48	50	250	70	100	513,6	0

Table 6: The results of experiment No.3 for small truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	598 782,97	49	220	106	104	504,4	0
	2	604 173,34	49	202	100	122	498,4	0,2
	3	623 961,29	49	236	110	110	530,4	0
Hill Climb		629 728,27	48	200	68	70	529,6	0,2
Adaptive Thermostatistical SA		591 224,07	49	218	94	98	484,4	0,2
Min/Mid/Max		672 237,37	48	200	100	100	587,4	0

5.2. Simulation experiments for standard truck

Secondly, the experiments for standard truck are carried out. The cargo capacity of this truck is 100 SKUs. The constraint for parameter of IAT_C is set to a very small interval ⟨95,105⟩ for all experiments with standard truck based on the same considerations as for a small truck. The constraints for initial number of SKUs in specific part of the warehouse are set for each experiment separately. Specific settings for individual experiments are specified in the Table 7. The results of experiments are presented in the Table 8, Table 9 and Table 10.

Table 7: Constraints for initial number of SKUs in specific part of warehouse for standard truck

Part of warehouse	Constraints interval		
	Experiment No.1	Experiment No.2	Experiment No.3
A	⟨150,300⟩	⟨250,500⟩	⟨400,700⟩
B	⟨80,200⟩	⟨80,200⟩	⟨100,200⟩
C	⟨70,150⟩	⟨70,150⟩	⟨100,200⟩

Table 8: The results of experiment No.1 for standard truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	469 753,05	100	182	136	118	542,6	0,8
	2	469 884,19	100	208	132	126	556,4	0,2
	3	472 265,32	101	232	142	138	569,6	0,4
Hill Climb		612 431,54	102	266	124	90	657	3,2
Adaptive Thermostatistical SA		459 707,07	100	198	126	144	551,4	0
Min/Mid/Max		503 856,03	100	224	140	150	597,4	0

Table 9: The results of experiment No.2 for standard truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	559 444,52	100	308	132	126	720,2	0,4
	2	570 107,47	100	312	122	142	692,8	0,4
	3	578 105,01	100	332	118	116	742,8	0,6
Hill Climb		571 932,04	101	368	132	122	766,2	0,4
Adaptive Thermostatistical SA		556 231,28	101	110	312	120	698,4	1,2
Min/Mid/Max		620 783,38	100	374	140	150	780,8	0

Table 10: The results of experiment No.3 for standard truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	663 081,26	100	424	168	154	814	1,2
	2	685 914,67	100	406	194	194	862	1
	3	687 353,25	99	428	198	184	892,2	0
Hill Climb		795 886,91	99	594	170	158	1004,2	0
Adaptive Thermostatistical SA		663 942,21	99	442	176	166	866,2	0
Min/Mid/Max		681 977,55	100	400	200	150	818	1,6

5.3. Simulation experiments for jumbo truck

Finally, the experiments for jumbo truck are carried out. The cargo capacity of this truck is 150 SKUs. The constraint for parameter of IAT_C is set to a very small interval ⟨140,160⟩ for all experiments with jumbo truck based on the same considerations as for a small or standard truck. The constraints for initial number of SKUs in specific part of the warehouse are set for each experiment separately. Specific settings for individual experiments are specified in the Table 11. The results of experiments are presented in the Table 12, Table 13 and Table 14.

Table 11: Constraints for initial number of SKUs in specific part of warehouse for jumbo truck

Part of warehouse	Constraints interval		
	Experiment No.1	Experiment No.2	Experiment No.3
A	⟨150,300⟩	⟨300,600⟩	⟨400,800⟩
B	⟨80,200⟩	⟨100,250⟩	⟨150,400⟩
C	⟨70,200⟩	⟨100,250⟩	⟨150,400⟩

Table 12: The results of experiment No.1 for jumbo truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	490 113,08	151	260	192	182	703,2	0,4
	2	500 195,32	151	288	186	192	735,2	0
	3	510 047,46	149	220	156	198	727,2	0,4
Hill Climb		1 051 368,46	149	238	152	84	1235,6	2,8
Adaptive Thermostatistical SA		468 776,35	149	212	168	178	693,8	0,2
Min/Mid/Max		537 129,02	150	274	200	200	764	0

Table 13: The results of experiment No.2 for jumbo truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	615 326,55	147	314	184	190	929,2	0
	2	644 866,89	149	370	176	228	933,2	0,2
	3	645 275,34	149	374	212	162	940,2	0,4
Hill Climb		722 028,73	149	486	166	182	993,6	0,4
Adaptive Thermostatistical SA		601 607,40	149	360	172	184	875,2	0,4
Min/Mid/Max		673 186,30	150	450	174	174	947,4	0,6

Table 14: The results of experiment No.3 for jumbo truck

Optimization method		$Cost^{total}$ [CZK]	IAT_C [STU]	$N_A^{initial}$	$N_B^{initial}$	$N_C^{initial}$	N^{max}	N^{zero}
Random	1	703 957,89	149	468	222	210	994	1
	2	781 546,69	147	496	280	212	1120	0
	3	783 095,57	148	404	278	328	1118,2	0,4
Hill Climb		1 015 220,44	151	704	218	174	1232,4	2,2
Adaptive Thermostatistical SA		665 739,39	148	438	244	216	1006,2	0
Min/Mid/Max		841 693,35	150	600	274	274	1231,2	0

6. SUMMARY OF RESULTS

The effective solution for three modeled situation (three experiments) is searched by means of four methods. The results presented in the tables above show that The Adaptive Thermostatistical SA algorithm achieves the best results for most experiments. Furthermore, it should be noted that the Hill Climb method and Min/Mid/Max method are inappropriate for this type of simulation study. The Random Solution achieves relatively good results, sometimes even the best result. The advantage of this method is its simplicity and its calculation speed compared to the Adaptive Thermostatistical SA method. The Random Solution is a fast and efficient algorithm for fast decision making. This algorithm provides a fast, very good solution that is not optimal, but very close to effective solution. The Random method error does not exceed 5% compared to the best simulation result.

The summary Table 15 presents the values of objective function (total costs of warehouse) for assumed type of vehicles and for all performed experiments. It is obvious that a standard truck is the best solution for supply the warehouse for most experiments that are performed. If long delay of loaded trucks is assumed, it is better to use a small type of vehicle.

Table 15: The values of objective function for assumed vehicles and experiments

Type of vehicle	Total costs of warehouse [CZK]		
	Exp. No.1	Exp. No.2	Exp. No.3
Small	513 848	557 918	591 224
Standard	459 707	556 231	663 081
Jumbo	468776	601 607	665 739

At the same time, the simulation study also shows how the total warehouse capacity requirements change when the inter arrival times are extended (the arrival of trucks is much delayed). The summary Table 16 presents the values of maximum number of SKUs in the warehouse (N^{max}) for assumed type of vehicles and for all performed experiments. It is clear that if the delay time of loaded vehicles is extended, it will be necessary to increase the capacity of warehouse by up to 50%.

Table 16: The values of maximum number of SKUs in the warehouse for assumed vehicles and experiments

Type of vehicle	The maximum number of SKUs in the warehouse		
	Exp. No.1	Exp. No.2	Exp. No.3
Small	372	452	484
Standard	551	698	814
Jumbo	694	875	1006

7. CONCLUSION

The paper is focused on increasing the efficiency of warehouse logistics using optimization methods. Different methods are used for searching the best solution. The Adaptive Thermostatistical SA algorithm achieves the best results for all experiments. The Random Solution provides a fast, very good solution that is not optimal, but very close to effective solution. The Random method error does not exceed 5% compared to the best simulation result. Sometimes this method achieves the best result.

Three real situations are considered. Firstly, the inter arrival times of trucks are kept very strictly. Secondly, the arrival delay of the truck is not more than 20% of the inter arrival time. Finally, the delay of arrival of the truck can reach up to 50% of the inter arrival time.

Each of the above experiment is applied to three different types of transport. There is a small truck, a standard truck and a jumbo truck. This simulation study shows that a standard truck is the best solution for supplying the warehouse for the most experiments that are performed. If long delay of loaded trucks is assumed, it is better to use a small type of vehicle. At

the same time, the simulation study also shows that if the delay time of loaded vehicles is extended, it will be necessary to increase the capacity of warehouse by up to 50%.

Simulation study ties together the project of controlled warehouse, which is implemented in these processes by the company. The verified model is subsequently used for simulation experiments.

This paper presents the possibilities afforded by using computer simulation for the design, optimization and identification of reserves in warehouse systems. Using concrete examples, it is demonstrated that the use of the Witness simulation environment – not only for suggestions designed to increase the effectivity of existing warehouses, but also in the initial creation and design is valid and effective.

ACKNOWLEDGMENTS

REFERENCES

Banks J. et al., 2005. Discrete-event system simulation. New Jersey: Prentice Hall, 608 p., ISBN: 0-13-144679-7.

Baker P., Canessa M., 2009. Warehouse design: A structured approach. European Journal of Operational Research, 193(2), 425-436.

Bottani E., R. Montanari, M. Rinaldi and G. Vignali (2015). Intelligent algorithms for warehouse management. Intelligent Systems Reference Library, 87, pp. 645-667.

Curcio D., Longo F. (2009). Inventory and internal logistics management as critical factors affecting the Supply Chain performances. International Journal of Simulation and Process Modelling, 5 (4), 278-288.

Gu J., Goetschalckx M., McGinnis L.F., 2007. Research on warehouse operation: A comprehensive review. European Journal of Operational Research 177 (1), 1-21.

Karasek, J., 2013. An Overview of Warehouse Optimization. International Journal Of Advances In Telecommunications, Electrotechnics, Signals And Systems, 2(3), 111-117.

Kare S., Veeramachaneni R.,Rajuldevi M K., 2009. Warehousing in theory and practice. University College of Boras dissertation. Sweden.

Muppani, V.R., Adil, G.K., Bandyopadhyay, A. 2010. A review of methodologies for class-based storage location assignment in a warehouse. International Journal of Advanced Operations Management, 2 (3-4), 274-291.

Petersen C.G., Aase G. (2004). A comparison of picking, storage, and routing policies in manual order picking. International Journal of Production Economics, 92(1), 11-19.

Raidl G., Pferschy U., 2010. Hybrid Optimization Methods for Warehouse Logistics and the Reconstruction of Destroyed Paper Documents. Dissertation paper. Vienna University Of Technology. Austria.

AUTHORS BIOGRAPHY

Bronislav Chramcov was born in Uherské Hradiště, Czech Republic, in 1975. He studied Automatization and control technology at the Faculty of Technology in Zlín of the University of Technology in Brno, and he took his degree in 1998. In 2006 he graduated his doctoral degree from the Faculty of Applied Informatics of Thomas Bata University in Zlín. His doctoral thesis was focused on the utilization of time series prediction for control of technological process. He is working now as an assistant professor at the Faculty of Applied Informatics of Thomas Bata University in Zlín. His research activities are focused on Control Algorithms for District Heating Systems, Time Series Forecast in Energy or Discrete event systems simulation.

VALIDATION OF A COST MODEL FOR THE SUPERSTRUCTURE SERVICE IN JUVENILE PRISONS IN BRAZIL BY MEANS OF THE MONTE CARLO SIMULATION

Camila Isaton [a], Antônio Edésio Jungles [b], Jamil José Salim Neto[c]

[a],[b],[c] Federal University of Santa Catarina - Department of Civil Engineering, Laboratory Construction Management, Florianópolis, Brazil

[a] camila.isaton@pucpr.br, [b] ajungles@gmail.com, [c] jamil@unifap.br

ABSTRACT

The typology of edification studied in this research is limited to Juvenile Prisons. The parametric estimation via Multiple Linear Regression is the method used to create the cost model of the Superstructure service. The model presents response speed for the cost forecast, since the equation generated can be applied through simple information available in the phases Initials of the enterprise. The objective of the article is to present the construction of the cost model for the Superstructure service and its validation is given through the Monte Carlo Simulation. The model presented a power of explanation of costs in the order of 71%. In the Monte Carlo simulation validation it was possible to detect that for scenarios with too small or too big total areas the model presented non-applicable results, that is, validation was performed for the extreme conditions.

Keywords: Estimativa Paramétrica, Prisões Juvenis, Modelo de Custo, Simulação de Monte Carlo.

1. INTRODUCTION

Brazil occupies the 4th place among the countries with the largest prison population (International Center for Prison Studies, 2014). About 7.4% of the Brazilian prison population is composed of young people under the age of 18, and these are distributed in 350 prison units in the country (Secretary of Human Rights of the Presidency of the Republic of Brazil, 2012).

The research approaches the cost estimate for Juvenile Prisons in Brazil. The result of the research is a parametric cost equation that allows the fast calculation of the Superstructure servisse costs during the cost estimation.

The need to estimate costs for juvenile detention in Brazil is due to the lack of visibility of this topic at the academic level in this country, presenting itself in this way as a relevant analysis of social and budgetary importance in the conception of new enterprises.

The parameters that feed the cost model are like geometric characteristics of the buildings. Parisotto (2004), presents that the cost estimates with the analysis of geometric characteristics demand less complex information, that is, as accessible in the initial phases of the project.

The motivation for conducting the research also counts on the need for new Juvenile Prisons in Brazil. The Carta Capital newspaper (2014), through the Access to Information Law, published that among the 148 Socio-Educational Detention Units that make up the socio-educational system of the state of São Paulo (the largest state in population and wealth in the country), 54% of these are overcrowded.

The judicial website JusBrasil (2012) revealed that another Brazilian state, called Minas Gerais, only young perpetrators of serious crimes are interned in the 22 units of the state, young people suspected of homicide, armed robberies and rapes count on the incapacity of the system, which cannot receive them for the fulfillment of the sentences, the Socio-Educational Detention Units of this state show overcrowding in the order of 48.7%.

Brazil coexists with crime increasing and the need to create vacancies for young people to comply with the penalties in a closed regime, starting from this motivation, originates the research. The article is part of a master's degree, where, from a study involving 39 projects, a cost model was developed for the Superstructure Service. This service, chosen for analysis, was chosen because of its representativeness in the ABC curve of Services, for this type of building, representing 18,31% of the total cost.

The generated model is classified as statistical, the methodology used for the construction of the model is the parametric estimation, the method of construction of the model is developed with the help of the software "*Statistica* in the *Ultimate Academic* version",. The validation of the model is performed through Monte Carlo Simulation, with the support of Excel software.

It is expected that with the present research, to assist the public administration (who finances and operates the enterprises of that origin in Brazil) in the speed of response during the process of cost analysis and generation of cost estimate for Juvenile Prisons. Also, the objective is to draw the attention of the academic community to criminal investigations of public security function, which are currently little explored in research, but have a great social relevance, due to its function in meeting the demands and needs of the society.

2. MODEL OF COST FOR THE SUPTRUCTION SERVICE OF JUVENILE PRISONS IN BRAZIL

2.1.1. Objective
Development of a statistical model for an estimation of costs of the Superstructure Service with validation of this model via Monte Carlo Simulation, a service related to the scope of buildings for use of Juvenile Prison in Brazil.

2.1.2 Search Limitations

The cost information considered in the database refers to the contracted values of the project. The price readjustment are defined due to the technological changes that occurred during the execution and were incorporated into the raw data base, thus, the results seek to be compatible with the actual values spent in the execution of works.

The cost model of this research is not applied to construction systems different from the conventional Brazilian system (pillars, beams, slabs and reinforced concrete).

3. REVIEW OF LITERATURE

3.1. Parametric Estimation
The Department of Defense US (2011), describes that parametric cost estimation is an intermediate level classified estimative modality executed when projects are complete in the rage of 10% to 35%.

The parametric estimation can be applied in finalized projects where the number of information is greater. In this manner, in modeling for databases of historical works and project.

Watson and Kwak (2004), report the origin of the parametric estimation, that takes place in the American Air Force, during the World War II, the methodology was used to estimate the costs of airplanes, taking into account the speed, and the reach by the aircraft.

The parametric estimation is the modeling that uses parameters to predict costs in construction, through data from previous projects (Sonmez, 2008; Hyun JI; Park; Soo Lee, 2010).

Mascaró (1985), the precursor of this analysis for buildings in Brazil, cites that from the point of view of geometric analysis, the building behaves like a plane, made up of sets of planes, these being horizontal in intersection with vertical sets, thus forming spaces projected.

Cerea and Premoli (2010), reinforce that the methodology is widely used in complex projects, and is effective in the preliminary identification phase, where the main cost factors are likely to be detected. Cerea and Premoli (2010), present that parametric estimation or parametric modeling is expressed through an analytical function inserted in a set of variables.

Watson and Kwak (2004) state that the characteristics of the project may allow the application of an algorithm that determines the cost approximation, where the characteristics can be: physical attributes or performance specifications.

Martins, Jungles and de Oliveira (2010) contextualize that it is imminent the need of the developers of the products to use models and tools that provide information and options to control the variables that interfere in product development, a practice common to other industry sectors that have focus as on the cost of the product in its initial phase.

Keller, Collopy and Componation (2013), point out that the estimation through parametric models presents statistical limitations, the authors analyze the application of the methodology in the American aerospace industry and criticize the production of mass generalist models that tend to capture all the characteristics of the product, emphasize the prior recognition of cause and effect relations.

3.2. Juvenile Prisons
Juvenile prisons are enterprises funded by public administration in Brazil, and each state of the federation is responsible for building, maintaining and operating them. These undertakings are intended for young people under the age of 21, and therefore have a distinct function of jails, prisons, temporary detention centers and penitentiaries, these last buildings, housing the male and female population over 21 years of age in a closed system.

3.3. Monte Carlo Simulation
The simulation aims to represent a real system, considerations can be made without the need for modifications in the system under analysis (Oliveira, 2008).

The Monte Carlo Method is convenient and has its increasing use for problems involving common simulations, as well as simulations of economic specificity (Di Bernardi, 2012). Gavira (2003), reports that the method allows the resolution of non-probabilistic problems with the use of the simulation through the stochastic process.

4. RESEARCH METHOD
The steps of the research development, according to the flowchart, began with the research question, where the Service of Superstructure shows great impact of the costs of the projects. In this way, it is sought through the parametric estimation to produce a model for forecasting the amount of resources necessary for the execution of the service under analysis. So in a feasibility study, where complete designs are not ready, it is possible to define the cost of the superstructure quickly.
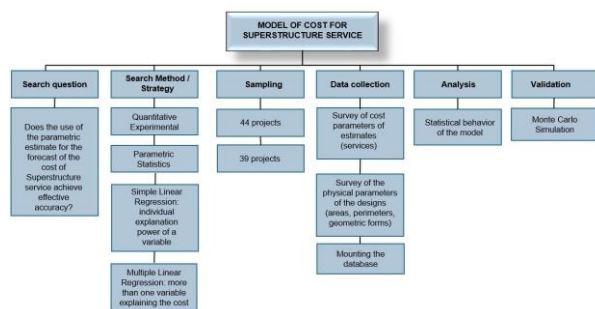
Figure 1: Flowchart of model execution and validation

The construction of the model started with the use of Simple Linear Regression with the measurement of forces between the independent variables and the dependent variable. Thus, a dependent variable such as the cost of the Superstructure was compared to an independent variable, such as Total Area, Total Perimeter etc. It was verified by means of this analysis that the simple linear regression models had no power of cost explanation for the service under study.

Then the model was constructed using multiple linear regression, in this phase, it was considered the hypothesis that more than one variable would influence the cost of the Superstructure.

The initial sampling was of 44 projects as shown in figure 1, suffering reduction for 39 projects, the elimination of 5 projects occurred due to the lack of architectural designs and standardized budgets, characterizing in this way as outliers.

Figure 02: Example of a floor plan of a Brazilian Youth Prison - Form A

Figure 03: Example of a floor plan of a Brazilian Youth Prison - Form B

Figure 04: Example of a floor plan of a Brazilian Youth Prison - Form C
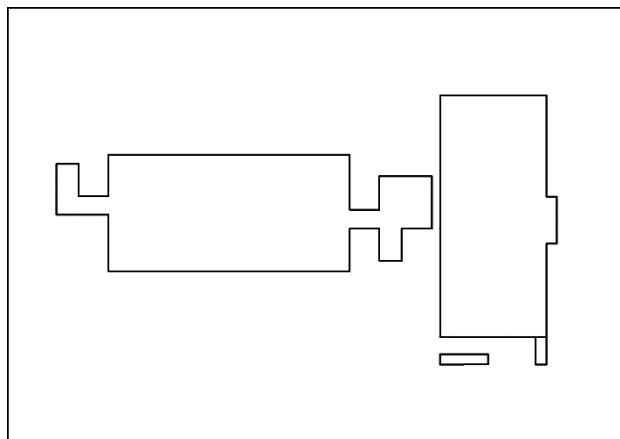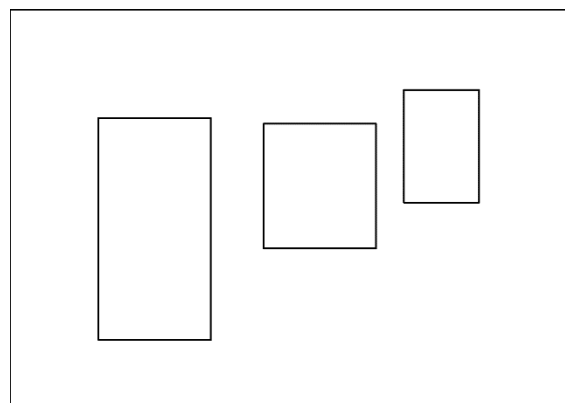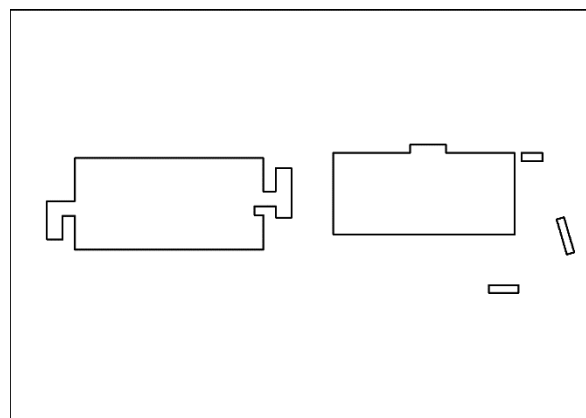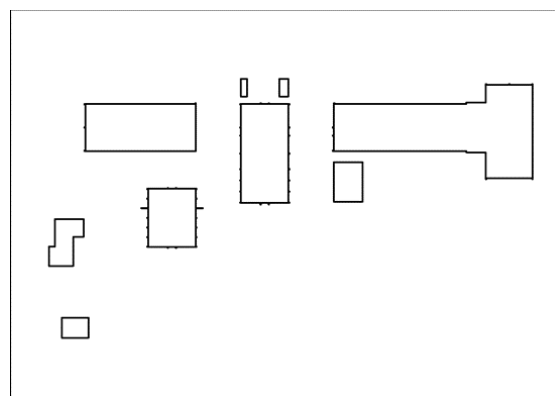

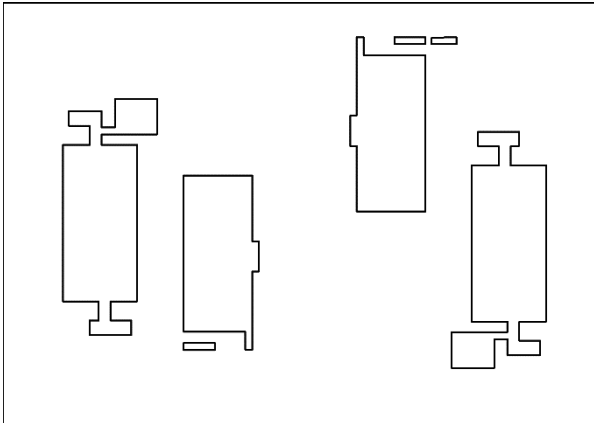Figure 05: Example of a floor plan of a Brazilian Youth Prison - Form D

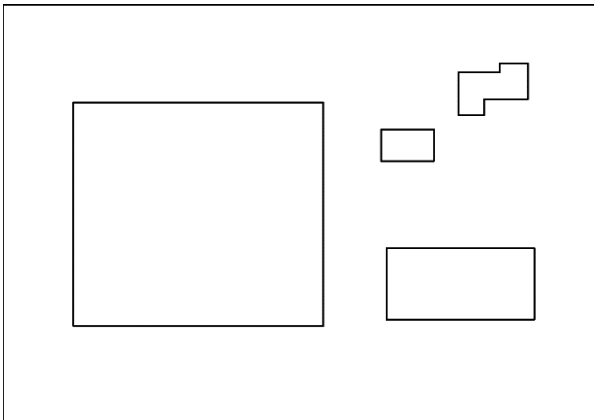Figure 06: Example of a floor plan of a Brazilian Youth Prison - Form E

Figure 07: Example of a floor plan of a Brazilian Youth Prison - Form F

The data collection mentioned in figure 1 refers to the costs of the works removed from the budgets and the geometric parameters taken from the architectural designs.

The figures 02, 03, 04, 05, 06 and 07 are the low levels of architectural design, that is, some of the samples that make up the 39 projects used in the experiment.
The mathematical analysis of the model was done through the analysis of the Test F of Significance, performed the test and ANOVA analysis, also verified if there was behavior of multicollinearity.

The validation of the model was done through the Monte Carlo Simulation, where 1000 scenarios of architectural designs were generated, so that the model for a larger universe could be validated than the sample of the experiment.

4.1. Construction of the cost model for the Juvenile Prisons Superstructure service

The research tool is Parametric Statistics and Monte Carlo Simulation, for the feasibility of the research was used the software Microsoft Excel and Statistica, the latter being in the Ultimate Academic version. The latter software, according to Ogliari (2004), is an integrated program with capacity to manage data analysis and databases, characterizing itself as broad in

the selection of the analytical process, from the basic levels to the advanced levels.

Montgomery (2009), emphasizes that the use of Multiple Linear Regression is recommended when the study presents situations that present more than one regressor.
NASA (2015), recommends as acceptable for cost estimates the coefficient R^2 on the order of 0.80 or higher. Care is needed in order to increase the R^2 value in the search for the best fit of the model, a model with higher R^2 with this addition may not perform better than the previous model (Montgomery, 2009).
The data of the 39 Youth Prisons were elaborated, tendered, contracted and executed from the year 2006 to the year 2015, and refers to a sample of one of the states of Brazil. We chose this time cut in the collection due to the standardization of budgets, descriptive memorials and other documentation of the works, from these, was taken from the geometric parameters that become the research data for the execution of the parametric estimation.
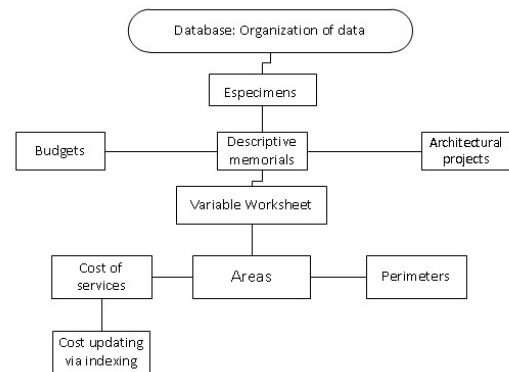
Figure 08: Assembly of the database for parametric estimation

The first step in parametric estimation is to select the samples (projects called Youth Prisons). The sample is composed of detailed budgets, descriptive memorials, architectural designs. These parameters, taken from the samples, form the variable worksheet, where the dependent and independent variables of the estimation will be chosen for later application of the Multiple Linear Regression.
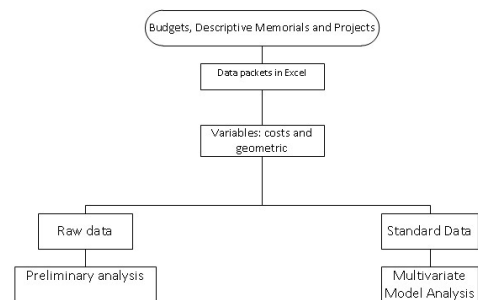
Figure 09: Standardization of data taken from specimens

The preference of the Superstructure service was due to its representativeness of 18.31% of the total amount of the costs of the projects of this nature, as shown by the ABC curve, presented in Table 01.

Description	% Partial	% Accumulated	
Security Frames	25,15	25,15	A
Superstructure	18,31	43,46	A
Roof	9,25	52,71	A
Sanitary facilities	7,52	60,23	A
Hidraulic instalations	7,27	67,50	A
Electrical and telephone installations	5,82	73,31	B
Masonry	5,52	78,83	B
Foundations	4,92	83,75	B
Painting	4,07	87,82	B
Various	3,23	91,05	B
Waterproofing, protection and gasket	2,15	93,20	C
Preliminary Services	2,13	95,34	C
Coatings	1,80	97,13	C
Earthmoving	1,50	98,64	C
Sidewalks, Guides and Gutters	0,49	99,13	C
Landscaping	0,48	99,61	C
Final Cleaning of the Work	0,34	99,94	C
Topography	0,06	100,00	C
TOTAL	100,00	100,00	

Table 1: ABC Curve of Services of Juvenile Prisons

The Suprastructure Service is composed of the following elements as indicated in figure 10: form fitting, forms, steel mounting, concrete pour, concrete cure, shoring of forms and false work.

Figure 10: Superstructure Service

Starting from the analysis of the ABC Curve of Services, it is verified that the Suprastructure service is an eminent driver of costs, justifying its analysis in this research.

For the generation of the Superstructure model recommendations from NASA (2015), were observed, it shows that for cost estimates we have two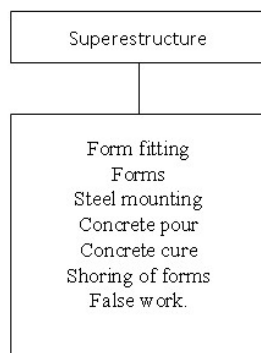 types of uncertainties that may occur, in order to impair accuracy, one of them is related to problems in the method Resulting from the omission of cost variables, poor specification of coefficients and poor mathematical relationships, as well as the lack or inconsistency of historical data used.

Hamaker (1995) apud Watson and Kwak (2004), sustain that most of the estimates are linear in nature, having a single independent variable associated with a cost, citing that inflection points are rare, and reports that changes in costs can be associated with a learning curve that is calculated separately.

Based on the concept above, the model of this study would be mathematically representative through the use of Simple Linear Regression. In this way it can be

combined: the total area with the costs of the service; the total perimeter versus service costs; or the regression between the density of walls and the costs of the Superstructure service.

However, individual models were generated and observed, however, they showed with low explanatory power of costs, leading to the attempt to search for more than one explanatory variable, so as to improve the representativeness of the proposed statistical model.

The correlation was generated with level of significance in the order of $p < 0.05$ measurement presented by "r" (Pearson's correlation coefficient). Correlation provided important information, which tested the parameters that had the best relationship with the dependent variable under study.

Based on the hypothesis, that more than one variable determines the cost of the Superstructure service, an explanatory equation is generated through the Multiple Linear Regression method, using more than one independent variable for the dependent variable.

For the specimens of 44 enterprises, 6 outliers were found, these were excluded from the sample, the coefficient used was 1.5, as recommended by the bibliography.

The parametric relationships can be defined: by the characteristics or properties of the product, according to Valle (2006). Or by the relation of costs with costs, as presented by the Department of Defense of the United States of America (2011). In summary, the quality of the parametric relationships depends on the validity, quality, size of the database and the purpose of the estimation.

In this research the fixed dependent variable was the costs of the Superstructure service, and the independent variables proposed were the parameters of areas and perimeters.

This research can also be performed by relating the cost parameters of other services. However, this requires a consistent database, a well-defined methodology for service definition, and preliminary knowledge of the mathematical correlation between variables (the variables should be related when both are observed on the x and y axis).

The model generated, for the sample of 39 Juvenile Prison units, had R^2 of determination presented in Figure 11 and Table 02.
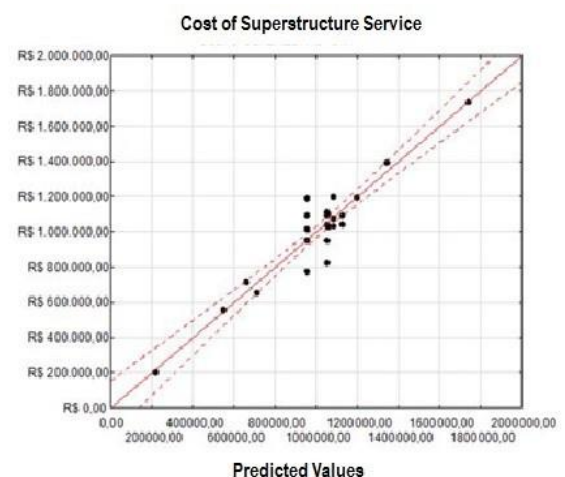
Description	Value
R	0,86
R²	0,74
R² adjusted	0,71
Standard error	110.682,00
Value p (F)	0,00

Table 2: Result of Multiple Linear Regression

The mathematical correlation relates the Cost of Service of Superstructure (CESUP) to the variables: External Perimeter (PE), Internal Perimeter (PI) and Total Area (TA) that are tested as cost drivers and which resulted in the following Equation 1.

CSUP= - 464.150,49 + 2.926,62 * PI + 94, 06 * AT + 177, 43 * PE

The model with the total cost of the Superstructure service as a dependent variable presented the adjusted coefficient of determination in the order of 0.71, that is, the model has explanatory power of 71%, when the parameters of the area and perimeters are available in the Design phase of the building.

The equation can be applied in cost estimates of Juvenile Prisons that have rectangular geometric shapes, built through the conventional constructional system: pillar system, beam and slab with concrete block walls. For the didactic effect of the application of the model that describes the cost of the Superstructure service, some parameters of the architectural design were selected, being: Total Area: 2020.00 m², External Perimeter: 150.73 m and Internal Perimeter: 434.82 m. Applying the Equation 01 produced in the research has the result:

CSUP= - 464.150,49 + 2.926,62 * PI + 94, 06 * AT + 177, 43 * PE
CSUP= - 464.150,49 + 2.926,62 * 150, 73 + 94, 06 * 2020, 00 + 177, 43 * 484, 82
CSUP= 253.001,76

That is, for a building containing these geometric parameters, the cost of the Superstructure service is obtained quickly for the estimation without the need of the structural design.

4.2. Validation of the model through Monte Carlo Simulation

In order to validate the model for projects that are outside the universe of samples, Monte Carlo Simulation was applied.

Monte Carlo Simulation is performed to include uncertainty and variability in the estimate to be performed with the cost model constructed through Multiple Linear Regression.

The Simulation is constructed from the generation of random numbers of probability and from the frequency distributions of each of the input parameters of the model: Total Area, External Perimeter and Internal Perimeter of the buildings.

Simulation of Monte Carlo for this situation and the benefits for diagnosing the effectiveness of the statistical cost model.

The number of 1000 scenarios was explained by Bruni, Fama and Siqueira (1998), where the authors clarify that the stabilization of the mean and standard deviation occurs around the 200th simulation, when performing 1000 simulations.

For Abreu and Stephan (1982, p. 152) apud Di Bernardi (2002), there is no need to repeat the simulation process more than 1,000 times, "since after a certain time the frequencies tend to stabilize and the modifications That occur only marginally affect the probability distribution."

Superstructure

CSUP= - 464.150,49 + 2.926,62 * PI + 94,06 * AT + 177,43 * PE	
Minimum Cost	R$ 80.438,67
Maximum Cost	R$ 2.606.788,61
Average Cost	R$ 1.084.317,22
Standard deviation	R$ 125.275,13

Table 03: Results of the application of the cost model for the scenarios generated through the Monte Carlo Simulation

The Superstructure service had an average cost of R $ 1,084,317.22, with a standard deviation of R$ 125,275.13. It was verified that when the model was applied in scenarios that represent Juvenile Prisons with low Total Area or high Total Area value, it presented values not feasible for the actual execution of an expedited estimate for the enterprise under analysis, when tested with extreme conditions parameters the model does not represent the cost of the Superstructure service effectively.

4.3. Analysis of the results

The cost drivers: Total Area, External Perimeter and Internal Perimeter were shown as explanatory independent variables for the modeling of the Superstructure service, that is, they are cost drivers.

The accuracy of the cost model for the Superstructure Service is compatible with the availability of information in the feasibility phase, in this initial stage of the enterprise, it is possible to estimate the costs with greater accuracy, when compared to simply dividing the overall value of the Works by the total area of the work, where the services are added having all the same weight in the estimate of costs.

In Monte Carlo Simulation, it was verified that when the model was applied in scenarios that represented juvenile prisons with low Total Area and high Total Area values (extreme conditions) it was not applicable in the execution of an expedited estimate of costs for the Superstructure service.

4.4. Final remarks

The model generated for the Superstructure service presented the adjusted coefficient of determination - R², in the order of 0.71, that is, the equation has the explanatory power of 71% of the costs for the service under analysis. It is considered the model of confidence, with moderate power of explanation, since the Department of Defense (2011), recommends indexes with values above 80% of representativeness. However, it is effective for the expedited estimation of Juvenile Prisons when compared to the classical methods of estimation where all services receive the same weight of costs in the project.

The error of the expedited estimate is acceptable, considering the information available in the initial phases of the projects.

The standard deviation generated by Monte Carlo Simulation for the cost model of the Superstructure Service was considered low, that is, the data generated through the random numbers showed homogeneity.

The application of the Monte Carlo Simulation to the model produced was the creation of scenarios (increasing the variability to show the uncertainty behavior of the model) for the application of the model for the Juvenile Prisons, with the generation of random numbers of the values of Total Area, Internal Perimeter, External Perimeter for a universe of different projects of the sample (considering that the data referring to descriptive statistics remained the same: mean and standard deviation, but the generated output values were different).

4.7 References

BRASIL. Sistema Nacional de Atendimento Socioeducativo - SINASE/ Secretaria Especial dos Direitos Humanos – Brasília-DF, 100p. Brasília, 2012.

BRUNI, A FAMA, R. SIQUEIRA, J. Análise do risco na avaliação de projetos de investimento: uma aplicação do método de Monte Carlo. Caderno de Pesquisas em Administração. v.1, n 6. São Paulo, 1998.

CARTA CAPITAL. Disponível em: <http://www.cartacapital.com.br/sociedade/um-em-cada-tres-unidades-da-fundacao-casa-tem-superlotacao-acima-do-permitido-pela-justica-2637.html>. Acesso em 15 set. 2015.

CEREA, A. P; PREMOLI, C. Stima parametrica del costo di costruzione. Individuazione di un metodo di stima, in fase di progettazione preliminare, del costo di costruzione tramite l'uso delle regressioni lineari. Tese (Doutorado em Engenharia Civil). Politecnico di Milano - Facoltà di Ingegneria Edile/Architettura, Milano (Itália), 2010.

DEPARTMENT OF DEFENSE. United States of America. Parametric Cost Estimating Handbook. 116p. Sine loco, Sine nomine, 2011.

DI BERNARDI, P.B. Análise de risco em investimentos imobiliários por simulação. 117p. Dissertação (Mestrado em Engenharia Civil). Universidade Federal de Santa Catarina, Florianópolis (SC), 2002.

GAVIRA. M de O. Simulação computacional como uma ferramenta de aquisição de conhecimento. 163p. Dissertação (mestrado em Engenharia de Produção). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos (SP), 2003.

Internacional Centre for Prison Studies. Word Prision Brief. Available from: http://www.prisonstudies.org/ [accessed 15 July 2016].

JUSBRASIL. Disponível em: <http://amp-mg.jusbrasil.com.br/noticias/3050086/numero-de-infratores-em-centros-socioeducativos-supera-em-48-7-a-quantidade-de-vagas>. Acesso em: 15 set. 2015.

KELLER, S; COLLOPY, P; COMPONATION, P. What is wrong with space system cost models? A survey and assessment of cost estimating approaches. Elsevier – Acta Astronautica, [S. l.], 7p. 2013.

MARTINS, D. das N; JUNGLES, A. E; OLIVEIRA R. de. Avaliação da qualidade geométrica de projetos habitacionais e seu impacto no custo do empreendimento. Ln: Encontro Nacional de Tecnologia do Ambiente Construído, XIII ENTAC, Canela, 2010. Anais... Canela: ENTAC, 2010.

MASCARÓ, Juan Luís. O Custo das Decisões Arquitetônicas. 100p. São Paulo, 1985.

MONTGOMERY, D.D. Estatística aplicada e probabilidade para engenheiros. 493p. 4ª edição. Rio de Janeiro (RJ), 2009.

NASA. Cost Estimating Handbook. Version 4.0 Disponível em http://www1.jsc.nasa.gov/bu2/PCEHHTML/pceh.htm>. Acesso em 10 jul. 2015.

OGLIARI, P.J. Análise Estatística usando o Statistica 6.0. Departamento de Informática e Estatística.130p. Florianópolis (SC), 2004.

OLIVEIRA, Miriam. Caracterização de prédios habitacionais de Porto Alegre através de variáveis geométricas – uma proposta a partir das técnicas de estimativas preliminares de custo. 125p. (Dissertação em Engenharia Civil) Universidade Federal do Rio Grande do Sul. Porto Alegre (RGS), 1990.

PARISOTTO, J. A.; AMARAL, T. G. do; HEINECK, L. F. M. Análise de estimativas paramétricas para formular um modelo de quantificação de serviços,

consumo de mão-de-obra e custos de edificações residenciais estudo de caso para uma empresa construtora. Ln: Encontro Nacional de Tecnologia do Ambiente Construído – X ENTAC, São Paulo, 2004. Anais...São Paulo: ANTAC, 2004.

SONMEZ, R. Parametric Range Estimating of Building Costs Using Regression Models and Bootstrap. Journal of Construction Engineering and Management - ASCE [S. l.], v. 134, n. 12, 6p. Dez, 2008.

VALLE, E. F. Análise de custos paramétricos de edificações não comerciais do oeste de Santa Catarina. Dissertação (Mestrado Profissionalizante em Desempenho de Sistemas Produtivos). Universidade Federal de Santa Catarina, Florianópolis, 2006.

WATSON, R.; KWAK, Y.H. Parametric Estimating in the Knowledge age: Capitalizing on Technological Advances. 2004, IAMOT Internacional Conference on Management of Tecnology. Washington, DC, Apr. 3-7, 2004.

POWER DISTRIBUTION CONTROL ALGORITHM FOR FUEL ECONOMY OPTIMIZATION OF 48V MILD HYBRID VEHICLE

Seongmin Ha [(a)], Taeho Park[(b)], Wonbin Na[(c)], Hyeongcheol Lee[*(d)]

[(a) (b) (c)]Department of Electric Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea
[(d)]Division of Electrical and Biomedical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea

[(a)] haha4100@hanyang.ac.kr, [(b)] koreapow@hanyang.ac.kr, [(c)] nao6114@hanyang.ac.kr, [(d)]hclee@hanyang.ac.kr

ABSTRACT

In this paper, we developed a supervisory control algorithm for fuel economy optimization of 48V MHEV (Mild Hybrid Electric Vehicle). It consists of the driving mode decision algorithm (Driving modes of 48V MHEV: Idle stop & go, EV(EV-launch, sailing), HEV(torque assist, Charge), ICE only, Recuperation) and power distribution algorithm for each driving mode. In particular, power distribution control is a key factor in determining the fuel economy of 48V MHEV. In this paper, a simulation-based analysis is performed to analyze the fuel consumption relevance of the power distribution algorithm. The vehicle model was developed in the Autonomie environment. The optimal power distribution control method was derived by analyzing the fuel consumption simulation results (traveling cycle: FTP 75) for the power distribution control with different tendencies.

Key Words : 48V mild hybrid electric vehicle, Supervisory control, Power distribution control

1. INTRODUCTION

In recent years, OEMs have been working to develop xEVs such as electric vehicles, hybrid electric vehicles, and fuel-cell electric vehicles in accordance with the global fuel economy and CO2 regulations. However, high-voltage, environmentally-friendly vehicles have not satisfied consumers because of the high cost of the vehicle to meet safety requirements. Solving these problems, OEM adopts the 48V system and develops the Mild Hybrid system which has better fuel economy improvement rate. This method can minimize powertrain structural changes, which can reduce the complexity of the vehicle system and reduce the cost. Various configurations (P0 ~ P4) have been proposed according to the electric motor mounting method of the 48V mild hybrid system (Figure 1).
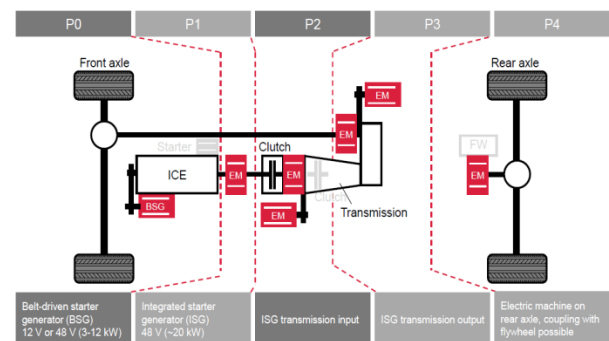
Figure 1: Configuration according to motor position

The P0 configuration replaces the existing belt-driven 12V generator to achieve a 48V system with minimal cost, while the P1 to the P4 configurations can be equipped with a high power motor with high mechanical power transfer efficiency. In addition, the P2-P4 configurations are capable of running in the EV mode, so the fuel efficiency improvement is high. The P4 configuration has the similar shape of e-AWD so that the vehicle dynamics control function can be realized.

In this paper, we study the P0 + P4 mixed configuration. This configuration enables various operations ranging from idle stop & go, EV mode, regenerative braking, charge, and torque assist to high efficiency through the combined operation of the belt drive generator (BSG) and the rear-axle drive motor. Among these supervisory control functions, the tendency of power distribution between regenerative braking, charging and torque assist is a key factor in determining fuel economy improvement. In this paper, we propose a rule - based power distribution algorithm for optimal fuel economy by analyzing the effect of each control on fuel economy. For this purpose, a 48V mild hybrid vehicle model with P0 + P4 configuration is realized using Autonomie and a simulation case for power distribution control with different tendencies is defined. Finally, we derived a rule-based power distribution control method optimized for 48V mild hybrid system through fuel economy simulation in FTP-75 cycle.

2. VEHICLE MODEL AND SIMULATION ENVIRONMENT
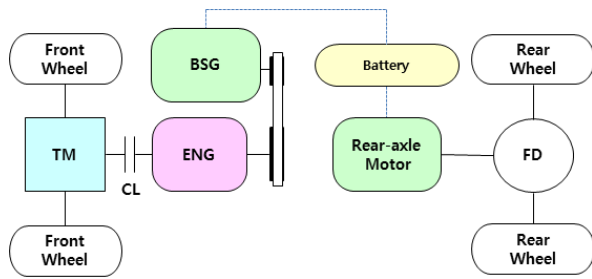
Figure 2: Target Vehicle

As shown in Figure 2, a 48V MHEV vehicle with a P0 + P4 structure is modeled. TM is the transmission, CL is the clutch, ENG is the engine, and FD is the final drive. The front wheel can be driven by the engine and the BSG, and the engine and the BSG are connected by a belt, which allows the torque assist and engine start via the BSG. The rear wheel is driven by a rear-axle motor. The main components information of the target vehicle are shown in Table 1.

Table 1: Vehicle main components

Engine	99kW L gasoline engine
BSG	11kW PM motor
Rear-axle motor	10kW PM motor
Battery	48V/11.5Ah lithium-ion battery

The 48V mild hybrid vehicle with the configuration as shown in the Figure 3 is the composition of the simulation model. The simulation model developed using Autonomie is consists of an upper controller, a driver model, an environmental model, and a powertrain model. Powertrain components consist of an engine, BSG, rear-axle motor, 48V and 12V battery, BDC, LDC, wheel, vehicle dynamics model, etc.
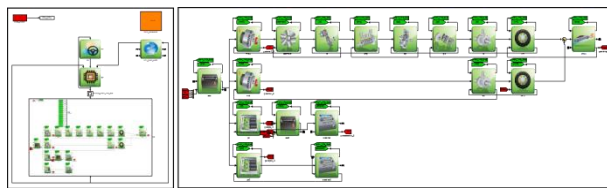
Figure 3: Simulation model and powertrain configuration

Table 2: Vehicle Parameters

Vehicle weight (kg)	1490
Frontal Area (m^2)	2.8
Rolling Coefficient	0.009
Aerodynamic Coefficient	0.37
Air density (kg/m^3)	1.23
Front final drive ratio	4.113
Rear final drive ratio	10.74

3. SUPERVISORY CONTROL ALGORITHM

As shown in the figure 4, the upper control algorithm was developed using Simulink, and consists of the mode decision algorithm and the power distribution algorithm of each mode.
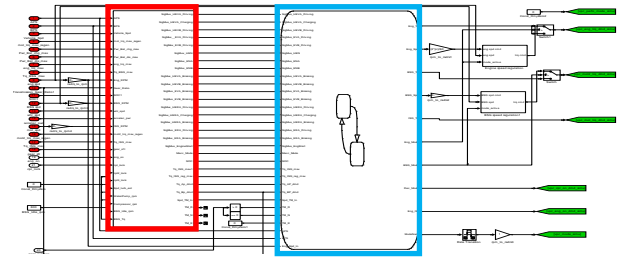
Figure 4:Supervisory control algorithm(Power distribution(red)/Mode decision(blue) algorithm)

3.1. Mode decision algorithm

In the mode decision algorithm, the driving mode is determined according to the driver's request and the vehicle state. The driving modes used in the 48V MHEV are classified as follows.

1. Idle stop & go mode
   In this mode, the engine is turned off at stop to save idle fuel consumption. Since it starts by using BSG, it can operate when the SOC is above a certain level.

2. EV mode
   In this mode, when the driver's acceleration demand torque is below the EV limit torque and the SOC is above a certain level, the vehicle travels using the rear-axle motor. It is used to start with a low demand torque when the vehicle is stationary, or to maintain the vehicle speed while driving.

3. ICE only mode
   In this mode, the vehicle is driven by the engine only when the driver's acceleration demand torque exceeds the assist limit torque or the torque assist of the motor is limited due to low SOC.

4. HEV torque assist mode
   In this mode, the vehicle travels to the engine and the motors when the SOC is above a certain level and the acceleration demand torque is above the EV limit torque and below the Assist limit torque.

5. HEV charge mode
   In this mode, when the charge sustaining is not possible with only recuperation energy, charge using engine and BSG. It is also used to prevent battery over discharge.

6. Recuperation mode

This mode is used to convert the kinetic energy of the vehicle into electrical energy when the deceleration demand torque is occurring.

Table 3: Driving mode according to driver's demand torque

Driver torque demand	Driving mode	
$T_{dmd} > T_{assist\_lim}$	ICE	Charge
$T_{assist\_lim} \geq T_{dmd} > T_{EV\_lim}$	Torque assist	
$T_{dmd} \leq T_{EV\_lim}$	EV	
$T_{dmd} < 0$	Recuperation	

T_{dmd} is the driver demand torque in the engine shaft, T_{assist_lim} is the control parameter that limits the torque assist torque to below the corresponding value. T_{EV_lim} is the control parameter that limits the EV mode to operate below the corresponding value with the EV limit torque.

3.2. Power distribution algorithm

In the power distribution algorithm, the torque command for each part such as engine, BSG and rear-axle motor is calculated for each mode according to the driver's request.

1. ICE only mode
   In this mode, the torque demand is distributed only to the engine. The power distribution formula is as follows.

$$T_{eng} = T_{dmd} \quad where \ T_{dmd} > T_{assist\_limit}$$
$$T_{mot} = 0 \tag{1}$$
$$T_{BSG} = 0$$

where T_{mot} is the rear-axle motor torque, T_{dmd} is the driver demand torque, T_{eng} is the engine torque, T_{BSG} is the BSG torque, T_{assist_limit} is the assist limit torque.

2. EV mode
   In the EV mode, the demand power is satisfied only by the rear-axle motor. The EV limit torque can also be determined.

$$T_{mot} = \frac{R_{tm} \cdot R_{front}}{R_{rear}} \cdot T_{dmd} \quad where \ T_{dmd} \leq T_{EV\_limit}$$
$$T_{eng} = 0 \tag{2}$$
$$T_{BSG} = 0$$

where T_{EV_limit} is the EV limit torque. R_{rear} is the rear final drive ratio, R_{front} is the front final drive ratio, R_{tm} is the gear ratio of transmission.

3. HEV torque assist mode
   In HEV assist mode, the torque for each of the engine, BSG, and rear-axle motor is calculated according to the demand torque. The assist limit torque can also be determined. The torque compensation amount of the motors can be controlled through the control variable f (T). Assist limit torque becomes 0 (f (T)).

$$T_{mot} = \min\left(f(T) \cdot T_{mot\_max}(w), \frac{R_{tm} \cdot R_{front}}{R_{rear}} \cdot T_{dmd}\right)$$

$$T_{BSG} = \min\left(f(T) \cdot T_{BSG\_max}(w), (T_{dmd} - \frac{R_{rear} \cdot T_{mot}}{R_{tm} \cdot R_{front}}) \cdot \frac{1}{R_{pulley}}\right)$$

$$T_{eng} = T_{dmd} - T_{mot} - T_{BSG} \tag{3}$$

$$where \ T_{assist\_limit} \geq T_{dmd} > T_{EV}$$

where $T_{mot\_max}(w)$ is the rear-axle motor maximum torque , $T_{BSG\_max}(w)$ is the BSG maximum torque, R_{pulley} is the belt pulley ratio, $f(T)$ is the control variable for power distribution of motors.

The EV / HEV drive domain is determined by the EV / HEV assist limit torque. Through this, the electric energy consumption of the battery is controlled to perform charge sustaining. Therefore, it is possible to control the use of all the charged electric energy to make a reliable fuel consumption comparison.

4. HEV charge mode
   In this mode, the engine torque is output by adding the torque to be charged to the BSG to the driver's requested torque.

$$T_{BSG} = T_{BSG\_gen\_min}(w)$$
$$T_{eng} = T_{dmd} + T_{BSG} \cdot R_{pulley}$$
$$T_{mot} = 0 \tag{4}$$

where $T_{BSG\_gen\_min}(w)$ is the minimum generating torque of BSG.

5. Recuperation mode
   In Recuperation mode, the torque is distributed to each motor according to the braking torque demand. If the braking torque demand exceeding the maximum regenerative braking torque, the friction brake is used. In this way, distributing the motor torque firstly, the regenerative braking energy can be maximized.

$$T_{mot} = \max\left(T_{mot\_reg\_min}(w), \frac{R_{tm} \cdot R_{front}}{R_{rear}} \cdot T_{dmd}\right)$$

$$T_{eng} = 0 \tag{5}$$

$$T_{BSG} = \max\left(T_{BSG\_reg\_min}(w), (T_{dmd} - \frac{R_{rear} \cdot T_{mot}}{R_{tm} \cdot R_{front}}) \cdot \frac{1}{R_{pulley}}\right)$$

$$where\ T_{dmd} < 0$$

where $T_{mot\_reg\_min}(\omega)$ is the minimum regenerating torque of rear-axle motor, $T_{BSG\_reg\_min}(\omega)$ is the minimum regenerating torque of BSG.

4. SIMULATION

In some cases, the simulation was performed by applying the host controller developed in the simulation model.

4.1. Simulation case

As shown in the table 2, different simulation cases are defined as follows.

Table 4: Simulation case according to driving mode

Simulation case	Driving mode
A	Idle stop & go
B	A + Recuperation
C	B + Torque assist
D	C + EV
E	D + Charge

1. Case A
   Only idle stop & go mode is performed, it is a criteria to determine the degree in improvement of fuel economy.

2. Case B
   Perform only idle stop & go / recuperation mode. It is possible to estimate the electric energy that can be obtained in the driving cycle.

3. Case C
   All electrical energy obtained through regenerative braking is used to perform torque assist mode.

4. Case D
   Set the EV limit torque to the maximum torque of the rear-axle motor to maximize the EV mode and use the extra regenerative energy for torque assist.

5. Case E
   The EV mode is maximized and the extra regenerative energy and forced charged energy are used as torque assist.

4.2. Simulation result
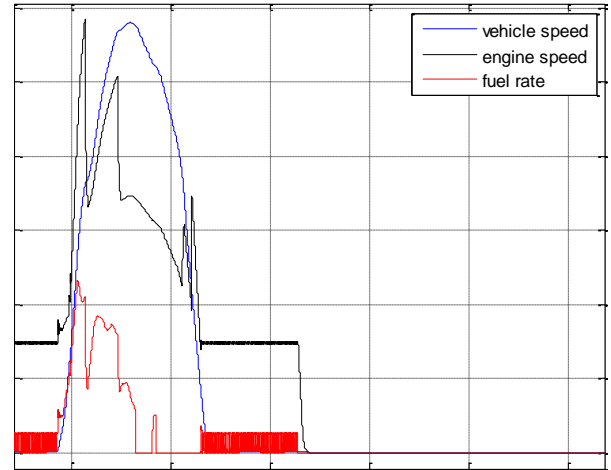
Figure 5: Simulation result of Case A

This is an enlarged view of a part of the simulation result. Each value is scaled for easy viewing because of the scale difference. Idle stop & go The engine keeps idle for a certain period of time after the vehicle is stopped, and the engine is turned off and the fuel consumption is zero. This confirms that idle top & go operation reduces unnecessary fuel consumption.
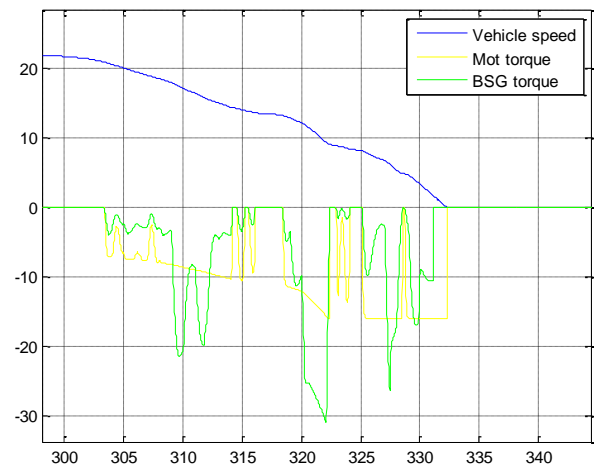
Figure 6: Simulation result of Case B

In the above graph, regenerative braking using two motors can be confirmed when decelerating. It is also possible to confirm that the torque is distributed preferentially to the rear-axle motor.
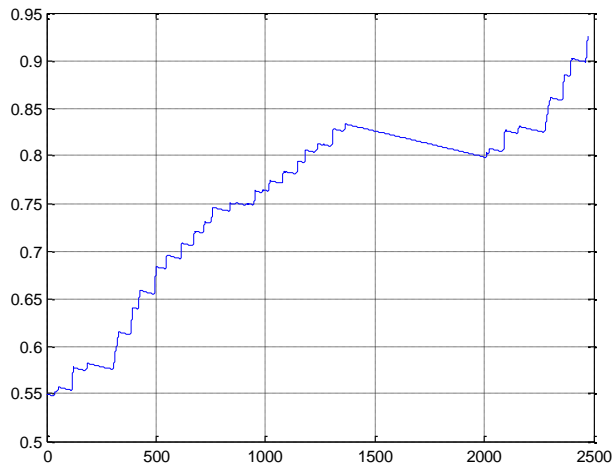
Figure 7: SOC of Case B

Figure 9: SOC of Case C

In the above graph, it can be confirmed that the battery SOC is increased by regenerative braking. It can be confirmed that the initial SOC is 55% and the final SOC is 92.56%, except for the power consumption by the electric component. It is charged 37.56% by regenerative braking, and this energy is used for torque assist and EV mode.

Figure 10: Simulation result of Case D

In the above graph, EV mode starts from about 710.5 seconds. The regenerative braking energy is consumed through the EV mode and the SOC is maintained in the running cycle as shown below.

Figure 8: Simulation result of Case C

This is the beginning of the case. Torque assist mode starts from about 20.5 seconds and ends in about 23 seconds. As shown in the graph above, the energy obtained by regenerative braking is consumed through the torque assist of the motor. In the following figure, it can be confirmed that the starting SOC is equal to the final SOC.
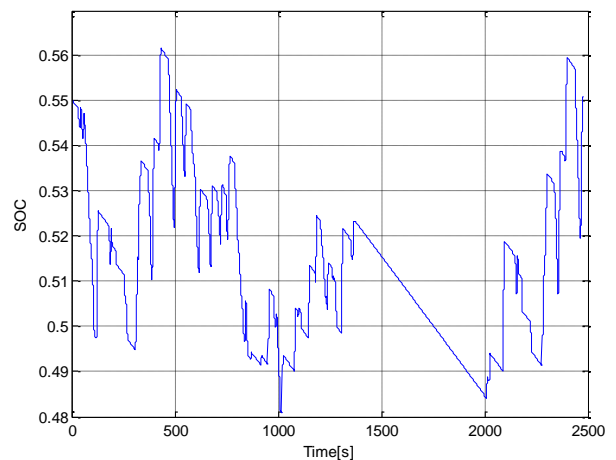
Figure 11: SOC of Case D
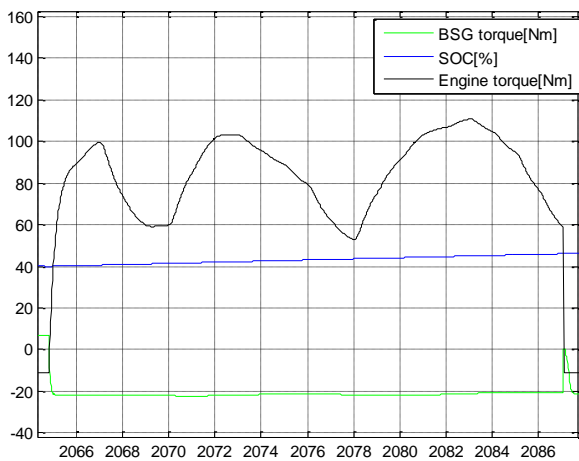
Figure 12: Simulation result of Case E

In the figure 12, it can be seen that the forced charge is working because the SOC drops to less than 40%.

The fuel economy of each simulation case for the driving cycle (FTP 75) is as follows.

Table 5: Fuel economy results

Simulation case	Driving mode	Fuel economy
A	Idle stop & go	14.48
B	A + Recuperation	14.73
C	B + Torque assist	17.03
D	C + EV	18.75
E	D + Charge	18.31

Comparing cases C and D, it can be seen that the EV mode is more effective in improving the fuel economy than the torque assist mode when the regenerative braking energy is the same and the total energy charged in battery is used. While both cases of D and E use EV mode to the maximum, but in the case E, additional torque assist is performed using the electrical energy obtained by the forced charging, which is inefficient. In the C and E cases, Comparing C and E cases, Even if the fuel is used for charging, it can be seen that the fuel efficiency is improved by using the EV mode.

5. CONCLUSION

The fuel economy improvement between C, D and E cases is determined by EV use domain and the torque to assist ratio of the BSG and the rear-axle motor.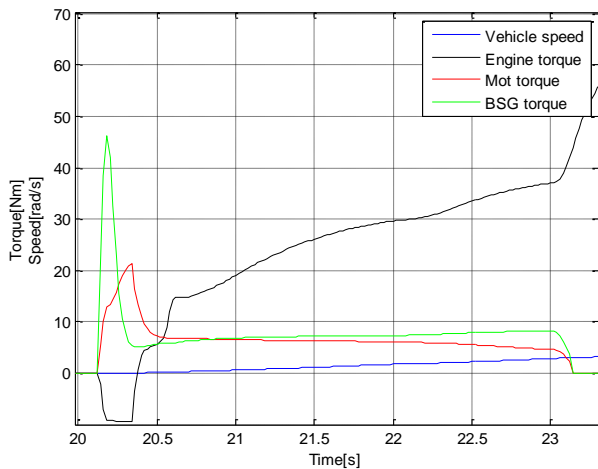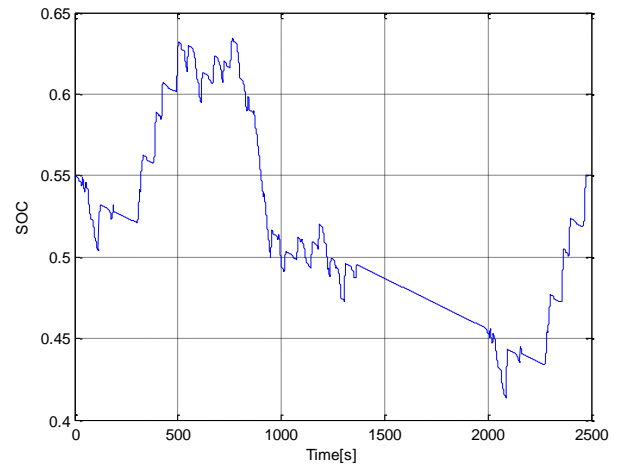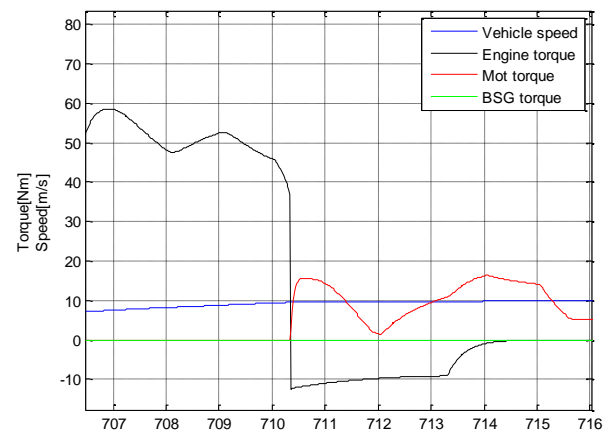 Among them, the EV mode domain contributes the most to fuel efficiency improvement. At presented results, the best fuel economy is shown in case D, but the more detailed study is needed to determine the control tendency to obtain optimum fuel efficiency.

ACKNOWLEDGMENTS

REFERENCES

Bao, R., Avila, V., and Baxter, J., "Effect of 48V Mild Hybrid System Layout on Powertrain System Efficiency and Its Potential of Fuel Economy Improvement," SAE Technical Paper 2017-01-1175, 2017.

German, J. "Hybrid vehicles: Trends in technology development and cost reduction," ICCT, http://www.theicct.org/hybrid-vehicles-trends-technology-development-and-cost-reduction, 2015.

Kuypers, M., "Application of 48 Volt for Mild Hybrid Vehicles and High Power Loads," SAE Technical Paper 2014-01-1790, 2014.

Dixon, G., Steffen, T., and Stobart, R., "A Parallel Hybrid Drive System for Small Vehicles: Architecture and Control Systems," SAE Technical Paper 2016-01-1170, 2016.

Brown, A., Nalbach, M., Kahnt, S., and Korner, A., "CO2 Emissions Reduction via 48V Active Engine-Off Coasting," SAE Int. J. Alt. Power. 5(1):2016.

Kuypers, M., "Application of 48 Volt for Mild Hybrid Vehicles and High Power Loads," SAE Technical Paper2014-01-1790, 2014, doi:10.4271/2014-01-1790.

Kim, S., Park, J., Hong, J., Lee, M. et al., "Transient Control Strategy of Hybrid Electric Vehicle during Mode Change," SAE Technical Paper 2009-01-0228, 2009, doi:10.4271/2009-01-0228.

Piccolo, A., Ippolito, L., Vaccaro, A., 2001. Optimisation of energy flow management in hybrid electric vehicles via genetic algorithms. IEEE/ASME International Conference on Advanced Intelligent Mechatronics Proceedings. Vol.1, pp.434–439.

U.S. ARMY MOBILE AUGMENTED AND VIRTUAL REALITY TRAINING SYSTEMS FOR HANDHELD IED DETECTORS

(a)Dean Reed, (b)Latika Eifert, (c)Shane Reyolds, (d)Travis Hillyer, (e)Clive Hoayun

(a),(c),(e)Institute for Simulation and Training, University of Central Florida
(b),(d) Army Research Laboratory-Human Research and Engineering Directorate Advanced Training and Simulation Division, SFC Paul Ray Smith Center

(a)dreed@ist.ucf.edu, (b)latika.k.eifert.civ@mail.mil, (c)sreynold@ist.ucf.edu,(d)travis.r.hillyer2.civ@mail.mil,
(e)choayun@ist.ucf.edu

ABSTRACT

The University of Central Florida's Institute for Simulation and Training (IST), and the U.S. Army's Research Laboratory (ARL) have collaborated on the creation of a suite of next generation mobile augmented reality (AR) and virtual reality (VR) applications. Our focus for this ongoing effort is full spectrum hand-held, mobile simulation-based training for advanced IED detectors. IST developed game engine based VR trainers capable of fully immersing the Soldier Trainees on low-cost mobile devices. A very advanced handheld AR trainer can convincingly emulate the proper motion required to employ the dual sensor detector in high fidelity virtual environments representing potential real operational environments, even to the representation of soil characteristics. This paper will discuss recent advancements, both hardware and software oriented, that enable the rapid deployment of high-quality end AR/VR training Apps. In addition, this paper discusses real world challenges associated with tackling complex training applications with mobile hardware.

Keywords: augmented reality, virtual reality, mobile embedded training

1. INTRODUCTION

Key enablers such as smart mobile devices with embedded GPUs and HMDs such as the Microsoft HoloLens have made low cost, mobile AR and VR training applications for the U.S. Army achievable. The cross-platform advanced game engines such as Unity have improved the fidelity of the trainee's experience. These tools continue to offer rapid and affordable development environments.

UCF IST has developed AR/VR Apps to support soldier IED training. Some of the Apps are in a ready to deploy state, while others are works in progress. We have achieved a diverse set of implementations that span the gamut of Milgram's Continuum.

2. AUGMENTED REALITY FOR TRAINING

IST and the U.S. Army partners at the Army Research Laboratories have implemented AR applications using several different technologies. Only a few years ago, it was noted that AR wearable HMDs and ancillary tracking systems were not sufficient to provide complete training systems (Stevens, Eifert 2015). The advent of powerful HMD systems such as the Microsoft HoloLens has enabled lightweight, ad-hoc functional on-demand training. IST developed a HoloLens application using a representation of a real life mine/IED device master-trainer. The authors used a motion capture system to collect the variety of critical motions that represent proper employment technique. The avatar can respond to voice queries and demonstrate complex tasks with the IED detector such as proper sweep techniques and can discuss and demonstrate calibration techniques in a variety of soil conditions supported by the trainer (Figure 1).

Figure 1: Microsoft HoloLens AR Instructor Application

A second AR application targeting the same IED sensor uses a distinctly different approach to achieve a very low cost and novel hardware surrogate trainer capable of delivering hands-on detector swing instruction. IST developers employed computer vision algorithms to enable a realistic mixed-reality application that allows trainees to deploy a low cost but convincing surrogate training system.

This AR simulation supports the capability to rapidly change the rendered visual environment. A simple change of the background fiducial image triggers a change in AR environment that also impacts the sensor performance and the simulation behaviour.

2.1. TRAINING COST SAVINGS

Current "real" IED detectors can easily exceed $20,000 per unit, but our AR training solution cost less than $500 per unit. IST combined modern 3D printing capability with high-performance tablet solutions to arrive at a very affordable system with proper weight form factor for the end training. A hybrid solution that adds our training device to a real IED detector was developed, but this only cost-effective when a supply of devices that are beyond repair is available. We removed the physical sensor head to implement the AR sensor head.

Figure 2: Low-Cost Surrogate (Left) Real Device (Right)

Another benefit of offering a low-cost approach is that it increases the accessibility of training. The expensive real/operational IED detectors can be difficult to obtain due to the perceived high dollar loss if a device is lost or damaged during training exercises. Our surrogate AR training system allows the user to practice in soil conditions that only exist in soil conditions from remote locales that would normally only be accessible via time-consuming training excursions or by building expensive infrastructure training ranges.

2.2. AR TRACKING DISCUSSION

The two approaches here, while both using AR, provide very different end applications. The HoloLens approach leverages real-time HMD sensor fusion to determine the correct AR viewing frustum. It uses time of flight sensors, as well as inertial sensors to calculate surfaces available for AR interaction. The onboard sensors are coupled with a form fitting graphics array to provide the rendering of the virtual graphics in the real world.

One identified problem with the HoloLens is the timeliness of the real-time environment mapping. It can take many seconds to acquire a fully realized mapping of the operational area depending on environmental conditions. As an example, we noticed that starting a brand new training exercise introduced in a new area crowded with people presented a challenge for the onboard sensors to realize the local coordinate system.

Conversely, we used a fiducial marker approach on the AR Swing trainer. Marker tracking coupled with internal IMUs available with the mobile device presents an excellent and affordable solution for this domain.

IST evaluated a surveyed a number of fiducial API solutions: Metai SDK, ARMedia, XZIMG, ARMedia, Wikitude, IN2AR, Obvious Engine, Vuforia and NyARToolkit. By trying each solution to compare the initial quality of pose estimation, we determined that the best contenders for our use case are Vuforia and Wikitude. A detailed comparison of the two led us to choose Vuforia over Wikitude because of how mature the Vuforia API is regarding 3D target recognition and independent object tracking. We created our IED swing trainer (Figure 3) with a generic interface to encapsulate the pose estimation software, so that we may rapidly change our solution to take advantage of newer or better solutions.
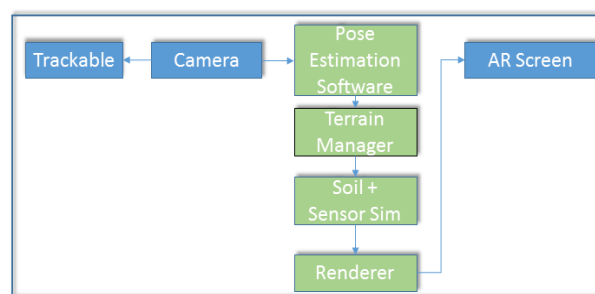
Figure 3: Fiducial Based AR Swing Trainer AR System Block Diagram

Fiducial markers have been successfully used in HMD based AR systems for over 15 years (Kato, Billinghurst 1999). As expected, our fiducial marker system works well within the constraints of the visual marker regions to provide solid pose estimation. Vuforia's extended tracking algorithms offered some improvements to the pure dead reckoning based positioning on sensor values, but we see significant tracking dropouts when we have no tracker images available to the onboard cameras. Sometimes, this can be overcome by cleverly populating our environment with tracking images that seem native to the training area. The non-uniform fiducial markers we employ are geotypical in appearance and offer a benefit in being natural to the end user application. Unfortunately, pre-positioning the markers relative to the physical camera on the mobile device can be error prone. The graph below (Figure 4.) illustrates tracking dropout, and shows a correlation between the distance of the fiducial and the AR camera speed when we lose tracking.
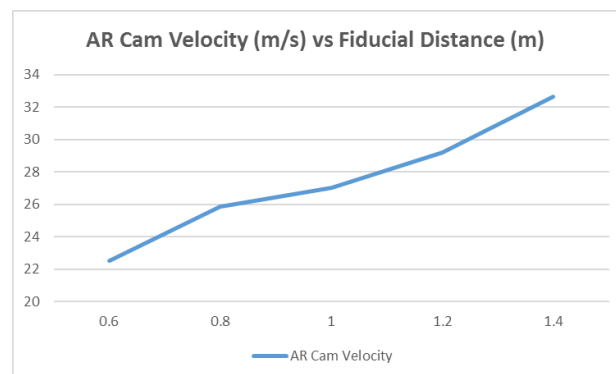
Figure 4: Distance from Fiducial and AR Velocity when tracking is lost.

Our experiment to find the AR pose accuracy leverages a high-quality camera mount with a built in levelling system for all degrees of freedom. We collected data at various distances from the center of our fiducial marker with a built-in grid for Cartesian correlation. A correlated grid incorporated into the 3D virtual environment is used to correlate and calibrate the measurement system. Figure 5 illustrates the experiment setup that we used to measure the Pose Estimation software. All lighting conditions and viewing angles to the fiducial were held constant. We collected a data sets at various heights (the y-axis) from fiducial marker to gauge a single variable change against the quality of position tracking.
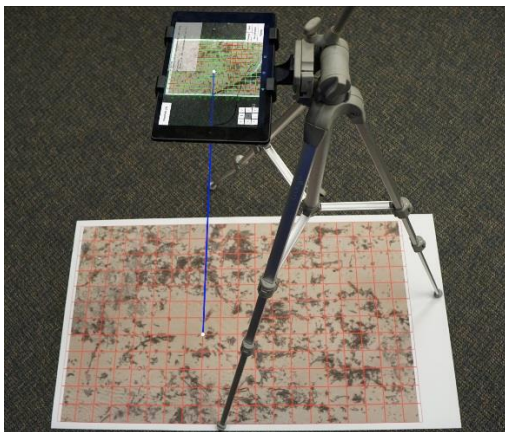
Figure 5. A System to Measure Pose Estimation Accuracy

Distance from Trackable (m)	Number of Samples (N)	Average Pose Estimation Error (cm)	Standard Deviation (cm)
0.8	147	2.7	0.49
1.0	147	3.6	0.81
1.2	147	5.7	1.73
1.4	123	7.3	1.61

Table 1. Average Pose Estimation Error

Table 1 shows that the accuracy of the pose estimation increases with distance from the trackable image. This conclusion can be attributed to per-frame image quality that decays with the focal distance capability of the on-board camera.

Due to the nature of the applicattion, we needed to characterized the tracking quality throughout the entire swing of the IED detector. Understanding the tracking error as it chages as you move relative to the fiducial marker is critical to presenting a quality AR based simulation. Initially we theorized that the fiducial measurement error was also related to the orientation of the marker with respect to the physical camera. The physical camera 16:10 ratio is a landscape field of view, but our marker is presented in a portrait mode. A second variation of data collection was conducted with the fiducial in an aspect ratio consistent with the physical camera. We simply rotated the marker 90 degrees to align the physical camera into proportion with the fiducial

extents. The difference in landscape was only 3mm different than the results in portrait mode. This result speaks to the consistency in the Vuforia API performance.

When we overlay the accuracy over a representation of the features, we noticed the direct relationship with camera location and the pose estimation accuracy outcome. Figure 6 shows a map of pose estimation error relative to the real camera position. The yellow "+" symbols are individual tracking features discovered in the overall image. In our experiment, the Vuforia pose estimation improves when the physical camera is centered with respect to the tracking image.
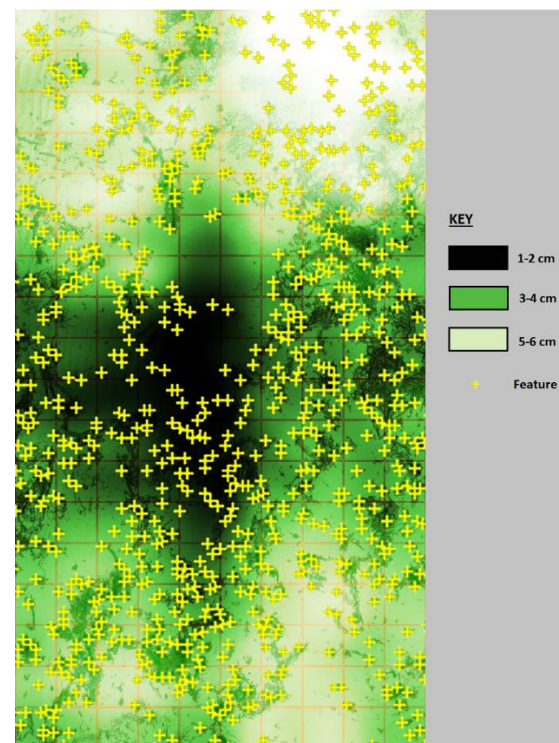
Figure 6. Accuracy Heat Map

For our use case, an important relationship exists between the distance from the trackable image, and the rate in which you can maintain quality tracking. In practice, a Soldier will travel at a suggested 1 m/s while covering a lane of up to 1.5m in width. Since our fiducial tracking system will need to be able to maintain our swing rate requirements while maintaining quality tracking we can see that we will need to maintain dense marker coverage within camera view.

2.3. AR FIDUCIAL FOR SOIL REPRESENTATION

With the AR Swing Trainer application, clever utilization of the Fiducial improves the application experience visually while providing a rapid way to modify the simulation environment. Our trackable images are correlated to geo-specific training environments, meaning that the regional soil conditions are changed with the visual tracker. Employment of

scalable fiducials (Neumann, et al 1999), allows us to create convenient way of providing a "close in" view of the world or quickly change our trackable to a different dimension to achieve a zoomed out effect. We use a straight forward mapping approach to load a new terrain environment as soon as a new marker is detected. Since our markers are integrated using non-uniform high quality dense imagery, we have noted that pose estimation can be less consistent than employing the pre-made markers provided by the Vuforia software.

Soil representation change causes the AR IED simulator to react differently as the user employs the device. We can simulate a variety of soil conditions: mineralization, density and humidity can be parametrically associated with an individual fiducial marker. Using this approach, the individual Soldier can practice in soil conditions that could only be achieved by building expensive ranges or travelling to the remote location where the real environmental conditions exists.

3. VIRTUAL REALITY APPLICATIONS

3.1. VR IED Trainer

We developed open-ended VR game engine-based software that draws on concepts borrowed from game design. Our VR environment is a high quality, third person view with an accurately modeled, real training range from Ft. Benning, GA called the McKenna MOUT facility.

This VR App offers the user multiple training scenarios, an advanced mode, and step-by-step guided tutorial for first-time users. This in-game tutorial style is consistent with modern commercial games that throw the player right into the action and point out the user controls as you progress through the introductory levels. This approach eliminates the need to train the user on how to use the training application. A "game within a game", or mini-game is included to implement precision sensor positioning.

Incorporation of a progressively more challenging set of IEDs that mimic actual training ranges was important to the training subject matter experts (SMEs). These SMEs proofed the simulated training range and offered constructive criticism to increase the training validity and constructive granularity of the overall training scenarios. The VR game is technically challenging, and has a very precise mini-game scenario that is embedded within the main vignette(s) to offer fine-grained sensor control and interaction. The responsiveness, competitive nature and visual quality of the environment contribute to the end users' enjoyment and by extent their willingness to participate in extensive training (Sveistrup, Thornton, Bruant et. al. 2004).

Figure 7. VR Training Scenario deployed on an Android Tablet

3.2. VR Hardware

IST and ARL have evaluated numerous consumer-off-the-shelf (COTS) candidate deployment targets for the VR HMD. We currently deploy the App to a tablet and phone form factor for common Android systems, but for immersion, we have created versions that run in HMDs as well. We have leveraged the Oculus Rift, HTC Vive, and the Samsung Gear VR to test the App in an immersive environment. An important benefit of an untethered VR solutions is the ability to train without an expensive personal computer. At the time of writing, the Google Daydream View and Samsung Gear VR are the most capable and offer the most mature API's for application development. UCF/IST chose to focus on the Samsung Gear VR for development since the Samsung "S" and "Note" lines of phones are the official devices used by large U.S. Army program of records such as Nett Warrior.

We chose the 3rd person perspective for our VR IED trainer as it offers best visual coverage of the simulated range environment. A trainee can observe the sweep speed of the Master Trainer Avatar as it navigates the environment.

The built-in controller on the side of the Samsung VR are assigned to fine tune VR camera controls while the Gear VR controller is assigned to the larger and more in-depth controls for Avatar locomotion and key IED detector functionality.

3.3. VR Game Environment

The real-world motion of sweeping a dual head sensor to reveal an (IED) is complex and precise. The timing of the sweeping motion coupled with the forward velocity are intentionally synchronized to provide optimal lane coverage with the onboard sensors. If the motion of the sensor sweep is too fast, slow, wide or shows abnormal tilting, called bowing or cupping, then it can induce safety issues by providing negative training.

Since the ARL supplied a subject matter expert (SME), we were able to capture and render the Master Trainer avatar in the mine detection training application

performing the proper sweep techniques. This includes the techniques used to determine the outline pattern of a potential IED. This precision animation would take several weeks to animate using key-frames and would be error-prone. IST leveraged a commercial motion capture (MoCap) system and captured the animations with a real world SME, a Sergeant assigned to the ARL. We used a passive MoCap system that allowed to capture the motions of the Soldier and the detector. In this way were able to use the most accurate motion data possible. The motion data from the Soldier also takes into account the weight of the Detector, which is fairly significant, as the Soldier is sweeping with the unit.

We employed photogrammetry to implement an ultra-high fidelity representation of the IED detector. The IED detector complete with individual buttons was captured into a very dense point cloud consisting of over 19,000,000 individual points. Seventy-two individual camera positions were used as we spun the real IED detector to collect the surface mesh and photorealistic textures. Our initial ultra-high resolution model weighs in at over 80,000 polygons.

Figure 8. High Resolution Photogrametric Model

Some modern mobile devices have very capable 3D graphics capabilities. We have targeted devices with high-quality graphics processing such as the NVidia Tegra chipset and the Samsung Exynos SoC solutions. While these mobile solutions offer higher polygonal rendering capabilities than previous generations, we still have to consider optimization techniques to preserve high frame rate and battery life. As an example, we applied a series of decimation methods to reduce rendering requirements for our high resolution IED detector model. Our ultra-high resolution model shown in Figure 8, was reduced from 80,000 to 12,000 polygons to facilitate loading into the desktop implementation of the VR software. Upon initial import, the 12k high resolution model caused our frame rate to drop below 10 fps on the mobile devices. To run efficiently within our game engine on the mobile device, while maintaining a large geospecific training environment, we decimated the IED model to around 2,000 polygons.

The VR training environment is a geospecific implementation of the McKenna MOUT site in Ft. Benning, GA. It represents a complex urban environment with a 2 km surrounding area represented at an average of 0.3 m post spacing resolution. A portion of the geospecific source was derived from airborne LIDAR sensors. Measurement for individual buildings and their accurate interiors were collected on site.

In the VR training game level, one of the important variables left to the trainee is the pacing speed that he selects for the Master Trainer Avatar. In the real world, doctrine calls for 1 m/s total area clearance, which is a slow and steady rate. In real world operation, incorrect swing rates can contribute to missing a concealed IED. Application users have the ability to outpace recommended search rates, and receive either real-time feedback or receive a full after action review (AAR) that offers a correlated speed vs. lane area that was missed during the clearance exercise. A user can replay his motions, the game engine renders the actual lane coverage, and then highlights the regions that were left uncleared. A summary or rollup score of IEDs located vs. the number of total IEDs in the area is displayed for user feedback.

REFERENCES

Stevens, J.A., 2015. Augmented Reality Technology in U.S. Army Training (WIP). 2015. The Society for Modeling and Simulation, Proceedings of the Summer Simulation Conference #88.

Kato, A.H. and Billinghurst M., 1999. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop on

Neumann U., You S., Cho Y., Lee J., Park J., 1999. Augmented Reality Tracking in Natural Environments, Mixed Reality - Merging Real and Virtual Worlds, Ohmsha and Springer-Verlag, 101-130

Sveistrup H., Thornton M., et al. 2004. Outcomes of Intervention Programs using Flatscreen Virtual Reality. Proceedings of the 26th Annual International Conference of the IEEE EMBS.

AUTHORS BIOGRAPHY

Dean Reed (B.S. in Computer Science, UCF 2000) is a Senior Associate for Simulation with the Institute for Simulation and Training (IST) of the University of Central Florida. Mr. Reed is a veteran of the U.S. Army, leads a team of developers at IST. He has worked on a vast array of projects under the auspices of the University including NASA Vision Spaceport. He is currently managing team efforts directed at evolving future training ranges on behalf of the U.S. Army.

Latika (Bonnie) Eifert (M.S. in Computer Engineering, UCF 2003). As Science and Technology Manager at the U.S. Army Research, Development and Engineering Command (RDECOM), Army Research Laboratory, Human Research Engineering Directorate, Simulation and Training Technology Center (ARL-HRED ATSD) located in Orlando, Florida, Ms. Eifert manages several projects associated with simulation and training. She is also supporting the Defense Advanced Research Project Agency (DARPA) by managing research program efforts.

Shane Reynolds is a graduate of UCF in Digital Media. He has specialized in developing compelling 3D content and mobile game engine development for over 10 years. Mr. Reynolds is a veteran of the U.S. Air Force. Shane is a Research Associate at the Institute for Simulation and Training where he has been a faculty member since 2008. His primary activities involve research and integration of modern technologies to train dismounted Soldiers at the squad level. Currently, his focal areas are technologies involving virtual reality, augmented reality, and photogrammetry.

Travis Hillyer is a Science & Technology Manager at the US Army Research Laboratory, Orlando, FL. He is a Mechanical Engineer by training and has been supporting research efforts in Live Training for ARL and DARPA for the past two years. The focus of his efforts with the Army Research Lab have been on characterizing and modeling free-space optical communication systems for improved Live Training realism and advanced ballistic simulation systems.

Clive Hoayun is a Computer Scientist and graduate of the University of Central Florida. While pursuing a Master's degree of Computer Science, he works with the IST working on a variety of U.S. Army projects. He excels at creating novel mobile applications and has interests in evolving compelling AR apps. Clive is a veteran of the U.S. Marine Corps.

METAHEURISTIC AND HYBRID SIMULATION-BASED OPTIMIZATION FOR SOLVING SCHEDULING PROBLEMS WITH MAJOR AND MINOR SETUP TIMES

A. Nahhas(a), P. Aurich (b), S. Bosse (c), T. Reggelin (d), K. Turowski (e)

(a), (c), (e) Magdeburg Research and Competence Cluster for Very Large Business Applications, Faculty of computer science, Otto von Guericke University Magdeburg, Germany
(b), (d) Department of Logistics and Material Handling Systems, Otto von Guericke University Magdeburg, Germany

(a) abdulrahman.nahhas@ovgu.de, (b) paul.aurich@mail.de, (c) sascha.bosse@ovgu.de
(d) tobias.reggelin@ovgu.de, (e) klaus.turowski@ovgu.de

ABSTRACT

This work has been motivated by an industrial case study in the field of printed circuit board's assembly production. Two- and four-stage Hybrid Flow Shop (HFS) scheduling problems with family major and minor sequence-dependent setup times are investigated. The majority of HFS scheduling problems are NP-hard optimization problems. Therefore, in this work, a metaheuristic and two hybrid simulation based optimization approaches will be presented to solve the problems and present a decision-making support tool for setting scheduling policies. Hybrid solution approaches that combine Genetic Algorithms (GA) with a heuristic are presented to solve the problems and compared to the GA. The optimization approaches are integrated into a discrete-event simulation model, which contributes as well as evaluates the quality of the obtained solutions. The formulated optimization problems are based on multi-objective measures to take into consideration the optimization of the system utilization through minimizing the makespan and the total number of major setup times as well as the customer satisfaction through minimizing the total tardiness. The presented solution techniques are evaluated based on real data, which are supported by the enterprise.

Keywords: Simulation-based Optimization, Hybrid Flow Shop Scheduling problem, Genetic Algorithms, Meta-heuristics

1. INTRODUCTION

Scheduling problems have been intensively investigated in the last four decades in different fields of academia due to their essential role in different manufacturing and service sectors as well as the challenging and complexity nature of their formulation (Ruiz and Vázquez-Rodríguez, 2010). In spite of the operative nature of scheduling tasks, they have a critical impact on most of the strategic decision making processes in an enterprise (Pinedo, 2012). This work has been motivated by an industrial case study in the field of printed circuit board's assembly production. The investigated problems are classified under Hybrid Flow Shop (HFS) scheduling problems. The HFS scheduling problems constitute a major class of scheduling problems, which is recently heavily addressed since the majority of assembly industrial production system are classified under HFS production systems (Ribas et al., 2010; Ruiz and Vázquez-Rodríguez, 2010). An HFS production environment consists of k production stages in series. Each production stage comprises m identical parallel machines. Each job j has to be processed on each production stage on one of the identical machines (Pinedo, 2012).

The majority of HFS scheduling problems are NP-hard optimization problems (Lenstra et al., 1977). Therefore, in this work, a metaheuristic and a hybrid simulation based optimization approaches will be presented to solve the problems and support decision-making processes regarding setting scheduling policies in the investigated system. Hybrid solution approaches that combine GA with a heuristic are presented to solve the problems and compared to the GA. The majority of the previous contributions in the field of scheduling problems strive to optimize the makespan (C_{max}), while very few target problems with multi-objectives function (Ribas et al., 2010). Therefore, the formulated optimization problems are based on multi-objective measures to take into consideration the optimization of the system utilization through minimizing the makespan and the total number of major setup times as well as the customer satisfaction through minimizing the total tardiness.

The presented solution techniques are evaluated based on real data, which are supported by the enterprise. The solution approaches are integrated into a simulation model to deliver a decision support tool for setting scheduling policies. This research aims to investigate the performance of the hybrid solution approach against Genetic Algorithms (GA) to solve the two-stage problem and then analyze the impact of expanding the complexity to solve the four-stage problem in terms of optimizing the objective measures and the required computational effort to obtain the solution. In the course of this paper, a literature review is presented in the second section to outline the relevance of the investigated problem and the often adapted solution

methodologies. The definition of the problems and the description of the investigated system are demonstrated in the third section. The fourth section comprises the presented hybrid approaches and implementation of the GA. Followed, in the fifth section, the computational results of the evaluation are demonstrated. Finally, the paper is closed with conclusions and further investigation scopes.

2. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

According to Graham et al. (1979), scheduling problems are formally described based on three fields problem description α | β | γ. In the first field α, the machine environment and configuration is illustrated. The job characteristics and the restrictions are presented in the second field β. The objective functions are then described in the third field γ.

2.1. Machine environment and configuration

The investigated problems are derived from the same production system. The four-stage problem is an expansion of the two-stage one. The system is classified as a Hybrid Flow Shop (HFS) production system. The first production stage contains four identical Surface Mounting Devices (SMD) machines. The second stage comprises five identical Automated Optical Inspection (AOI) machines. All jobs must be processed on the first and second production stages. Some jobs undergo the third and/or the fourth processing stages. The third stage contains a single Selective Soldering (SS) machine. Similarly, on the fourth stage, a Conformal Coating (CC) machine is available.
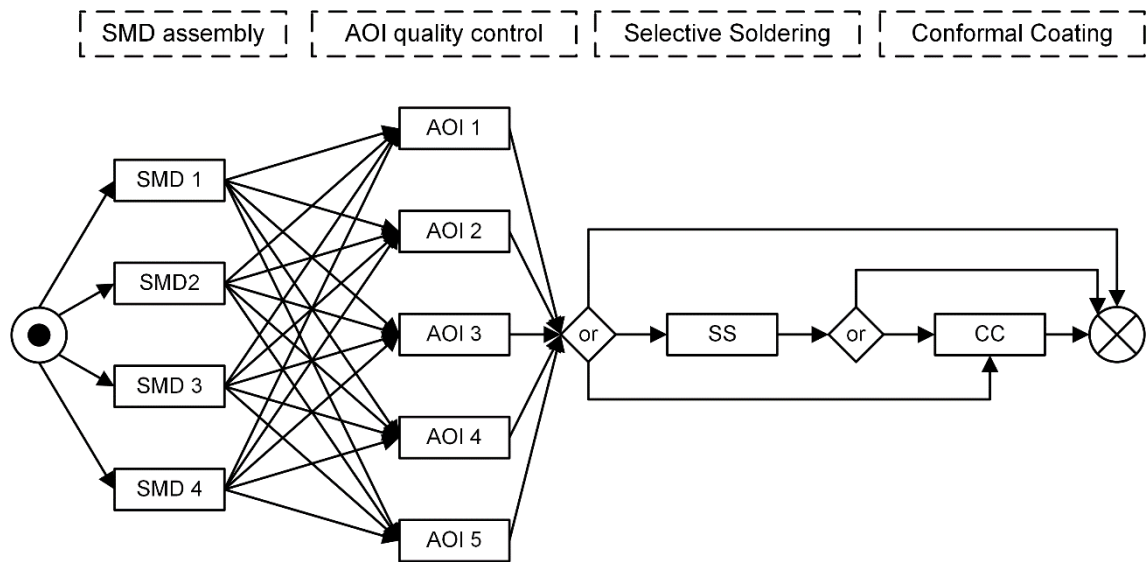
Figure 1: The investigated production system.

2.2. Job characteristics

Jobs of the analyzed HFS scheduling problems can be characterized as follows:

- The number of jobs in a certain time period and the number of products per job are known and fix.
- Part types are very heterogeneous.
- The family type of a job depends on the used raw materials.
- The processing time $p_{j,m,s}$ of each job j on the machine m of stage s is known and fix.
- The priority of a job represents the delivery date to the customer.
- The major setup time $s_{g,h}$ is the required time to configure a machine that was processing jobs from family g to process jobs from family h.
- Machine breakdowns are aggregated and subtracted from the production capacity.
- Buffer size between production stages is unlimited.

In the first production stage (SMD), jobs are scheduled with family major and minor setup times on the machines. In the second production stage (AOI), jobs are scheduled incurring sequence-independent setup times on the machines. The concept of major and minor setup time was introduced by Wittrock (1990) as well as by Kut C. So, (1990) to describe sequence-dependency. Jobs which share common raw materials are grouped into families. A minor setup time is inquired if the machine switches from one part type to another inside the same family. On the other hand, a major setup time is encountered, if the machine switches from one part type to another from a different family.

2.3. Objective functions

Accomplishing a balance between production system efficiency and the job's due date is a trade-off decision. For this reason, tardiness has been frequently used as a major supplementary performance criterion along with the makespan (Lenstra et al., 1977). The objective functions of the analyzed HFS problems are to

minimize the makespan C_{max}, the total tardiness T and the total number of major setup times to process all jobs as demonstrated in. The makespan is the necessary time to complete all released jobs (Wittrock, 1990) as demonstrated in $C_{max} = \max C_i \; \forall \; i = 1, \; ..., \; n$

(**1**. To minimize C_{max} it is important to minimize the number of major setups. Tardiness is the difference between the completion time of a job C_j and its due date d_j as shown in

$$T = \sum_{i=1}^{n} T_i \; , \; T_i = \max(C_i - d_i, \; 0)$$

(**2**.

$$C_{max} = \max C_i \; \forall \; i = 1, \; ..., \; n \qquad (1)$$

$$T = \sum_{i=1}^{n} T_i \; , \; T_i = \max(C_i - d_i, \; 0) \qquad (2)$$

3. LITERATURE REVIEW

Generally speaking, the solution methodologies that are often used to deal with scheduling problems can be classified according to their solution quality and method of implementation into two main groups. The first group contains analytical methods and exact optimization techniques, while the second group consists of heuristics and improvement approaches. The exact approaches usually guarantee optimal solutions or bounded optimal ones using some approximation scheme (Baker and Trietsch, 2009). Dynamic programming (Held and Karp, 1961) and branch and bound (Kis and Pesch, 2005) are typical exact optimization approaches, which are often employed to deal with scheduling problems. Those approaches are usually adapted to solve small- and moderate-size scheduling problem. One of the prior works in the flow shop is scheduling is the contribution of Johnson (1954). The author investigated a two-stage Flow shop (F) scheduling problem $F(1,1) \| C_{max}$ and presented an exact algorithm to minimize the makespan in polynomial time.

Kis and Pesch (2005) presented a comprehensive review of the adapted exact approaches for solving HFS scheduling problems, in which the branch and bound optimization technique was the dominant exact solution approach. Although exact approaches grantee optimal solution, they are computationally expensive when the problem domain gets more complex. In addition, the conducted implementations of exact approaches in the field of operation research maintain often a scientific nature of problem formulation, which usually involves many simplifications and rough assumptions to reduce the complexity of a considered problem. This fact propagated a gap between the research conducted on scheduling theory and scheduling activities in practice (Maccarthy and Liu, 1993, p. 59).

In the industry, scheduling policies are often managed based on some intuitive rules and procedures. Academia classified those procedures under the so-called Priority Dispatching Rules (PDRs). They constitute the simplest form of heuristic procedures. These rules are widely used in practice to manage production planning. By definition, PDRs are some simple constructive procedures used to prioritize a set of released jobs for scheduling and production (Hunsucker and Shah, 1994). The Earliest-Due-Date (EDD) and the Shortest-Processing-Time (SPT) are typical examples of PDRs, which are often used to set a sequencing strategy to prioritize and measure the significant degree of jobs (Andersson et al., 2008). Hunsucker and Shah (1994) presented a profound analysis on PDRs and their performance. They analyzed six PDRs on different flow shop scheduling problems $HFS_m \| C_{max}, F_{max}, \overline{F}$ that are subject to the minimization of various objective functions such as the makespan, the mean flow time and the maximum flow time (F_{max}).

However, when the quality of the obtained solution become more essential, PDRs are not anymore applicable. This statement can be explained through observing an inverse correlation between their solution quality and the complexity of an investigated problem. Therefore, PDRs have been recently investigated in combination with metaheuristic approaches as for instance genetic algorithms, in which a simulation study is conducted to evaluate the quality of the obtained solution as presented by Geiger et al. (2006). A Similar concept has been investigated by Andersson et al. (2008), who conducted a simulation study, in which genetic algorithms are employed to identify the appropriate combination of priority dispatching rules for solving scheduling problem. The encoding strategy of the metaheuristic in such implementation usually targets the problem indirectly, through passing the different rules to the optimization techniques instead of approaching the problem directly. More sophisticated methods have been also applied to construct PDRs using some machine learning mechanisms as for instance neural networks (Wang et al., 2005).

Improvement heuristics and metaheuristics can be anticipated as a middle ground between PDRs and exact methods in terms of the solution quality and the computational effort required to solve scheduling problems. Improvement heuristics are conceptually more sophisticated heuristic procedures than the constructive ones since the construction of a production schedule is the first step in the internal functionality of an improvement heuristic. Thereafter, based on the initial constructed schedule, an improvement heuristic seeks to conduct single or several changes on the schedule, which yield to a better investigated objective function (Pinedo, 2012, p. 382). A particular class of improvement procedures is local search algorithms. A local search algorithm functions in a similar manner to an improvement procedure, except that the modification procedures on an investigated solution must be well

defined and identically conducted on all candidate solutions (Pirlot, 1996).

Local search algorithms attempt to find the optimum of an investigated region or 'neighborhood' of the solution space. Two solution candidates are neighbors if conducting a predefined change on a schedule generates the other neighbor schedule. The modification process is iteratively conducted on the investigated solutions until the optimum of that region in the solution space is found or a breaking criterion is met (Orlin et al., 2004, p. 588). The modification procedure in this sense defines the complexity of an improvement algorithm since it shapes the size of the investigated neighborhood in the solution space. Wittrock (1990) investigated a simplified form of the considered problems. He presented an improvement heuristic approach based on binary search tree to solve the identical parallel machines problem (P_m), which is a stage of a hybrid flow shop production system $P_m \mid s_{g,h} \mid C_{max}$. The problem was investigated under major and minor family sequence-dependent setup time ($s_{g,h}$), which is subject to minimize the makespan. The author proved that the complexity of the problem is NP-hard. The performance of the presented approach was finally compared against a lower bound on the problem, which was demonstrated by the author.

Gupta, (1988) addressed a sophisticated form of the identical parallel machine problem $HFS_2(P_m, 1) \parallel C_{max}$, in which the second stage with a single machine has been taken into consideration to form an HFS problem. The problem is NP-hard in a strong sense. Gupta treated the problem by dealing with the sequencing part and the allocation part independently. He assumed that only two-stage flow shop scheduling problem has to be solved with the objective function to minimize the makespan $F(1,1) \parallel C_{max}$. He adapted the algorithm, which was presented by Johnson (1954) to deal with the sequencing part of the problem since this algorithm solves the problem optimally. Independently, he presented a heuristic to deal with the allocation part of the problem on the first stage with the objective function to minimize the total idle time on the second stage. He reported near optimal results of the makespan with three to five percent deviation from his calculated lower bound in almost all generated problems.

Voß (1993) addressed a similar problem to the one investigated by Gupta, in which sequence-dependent setup times is considered $HFS_2(P_m, 1) \mid s_{j,k} \mid C_{max}$. The author adopted the same solution strategy presented by Gupta to solve the problem. His contribution lied in integrating different improvement procedures to the solution strategy presented by Gupta and a new setup method. He also applied Tabu Search to obtain a local optimum from the solution space of the generated initial solution. The author reported improved results for all problem instances in comparison to the ones generated by Gupta. Li (1997) also treated a two-stage hybrid flow shop scheduling problem. However, his problem

formulation was literally opposite in term of machines to the one addressed by Voß and Gupta, in which only one machine on the first stage and parallel identical machines on the second stage have been taken into consideration $HFS_2(1, P_m) \mid s_{j,k} \mid C_{max}$. The problem was investigated under family major and minor sequence-dependent setup time constraint using different heuristic methods.

Although improvement and among them local search algorithms point out good results, their conceptual design and functionality are based on searching only in the neighborhood of an initial solution to achieve better ones. This implies that the optimum in this region of the solution space will be identified (Ross, 2005, pp. 529-530). In essence, a rough assumption has to be made to define the modification strategy as mentioned earlier that automatically discard many feasible solutions, which might lead to finding a global optimum for a problem. This major drawback gave a solid motivation to academia to address the problem of the local optimum. As a result, many sophisticated optimization methodologies under the name metaheuristics have been presented to solve very difficult combinatorial optimization problems (Glover and Kochenberger, 2003). Metaheuristics are guided local search algorithms, which consist of two main fundamental elements: A local search algorithm and an overall optimization or control strategy. The control strategy is used to guide and control the local search algorithm (Ross, 2005, p. 530). Simulated Annealing (SA) (Kirkpatrick et al., 1983), Tabu Search (TS) (Glover, 1989), Genetic Algorithms (GA) (Goldberg, 1989; Holland, 1975) are some examples of often adapted metaheuristics to solve HFS scheduling problems (Aurich et al., 2016; Mirsanei et al., 2011; Nahhas et al., 2016; Reeves, 1995; Ruiz and Maroto, 2006).

Reeves (1995) presented one of the first implementation of genetic algorithms to solve scheduling problems. The author addressed the flow shop scheduling problems with the objective function to minimize the makespan $F_m \mid permu \mid C_{max}$. The problem is NP-hard.

The author investigated a special form of flow shop, in which the permutation of schedule is maintained after the sequencing process on the first machine. A permutation flow shop is a special case of the flow shop except that after building the sequence on the first, the First-In-First-Out (FIFO) discipline is used to further dispatch jobs from the in-production queues. The computational results comprised a comparison between the presented a GA and SA approaches and a local search algorithm. For all investigated problem instances, both GA and SA outperformed the local search algorithm. Similarly, Zheng and Wang (2003) presented a hybrid implementation to deal the same problem. The authors adopted GA to address the problems and further incorporated the heuristic presented by Nawaz et al. (1983), which is known with the name NEH heuristic to generate the initial population before triggering the GA. Their main idea is

to size the advantage of a high-quality initial population of solution candidates, which helps GA to systematically search in the regions of high-quality solutions by the beginning of the first iteration.

One of the first contributions addressing HFS problems using GA was presented by Serifoğlu and Ulusoy (2004). The authors presented a comprehensive analysis of their implementation with the objective function to minimize the makespan $HFS_m \| C_{max}$. The problem is NP-hard with a simple reduction on the problem treated by Gupta (1988), in which a single machine one the second stage has been taken into consideration. The authors encoded the problem based on the permutation of jobs on the first production stage. The genetic algorithms are employed to solve the sequencing problem in the first stage, whereby Last-In-First-Out (LIFO) discipline has been applied to dispatch jobs on the other production stages. The experiments were conducted on a benchmark datasets, in which up to 100 jobs have been taken into consideration. Very similarly implementation was presented by Oğuz and Ercan (2005) to deal with the same problem.

Ruiz and Vázquez-Rodríguez (2010) presented a comprehensive literature review, which involved the analysis of over two hundred contributions in the HFS research between the early seventies and 2010. The presented results showed that over sixty percent of the contributions targeted the minimization of the makespan as an objective function. A similar review of the HFS research has been presented by Ribas et al., (2010). The authors, however, restricted their review timeline between the middle nineties and 2010. The findings of the review also stressed on the unrealistic problem formulations. The authors also explicitly pointed out the lack of contributions, in which the investigated problems are formulated based on real-world experiences with real data used for the evaluation of the solution approaches.

Nahhas et al. (2016) presented one of the few contributions, which involved the minimization of the makespan as well as the total tardiness. The authors adapted TS and SA metaheuristics and a heuristic approach named ISBO to deal with a two-stage hybrid flow shop scheduling problem with identical parallel machines on each stage $HFS_2(P4, P5) | s_{g,h} | C_{max}, \sum T_j$. The problem was investigated under family major and minor setup times constraint on the first production stage. The authors evaluated their solution approaches based on an industrial use case using real data. This work will extend the conducted analysis and present two hybrid solution approaches that combine the ISBO with GA to solve the two-stage problem and then analyze the performance of the approaches and the impact of the complexity for solving the four-stage problem $HFS_4(P4, P5, 1, 1) | s_{g,h} | C_{max}, \sum T_j$.

4. SOLUTION APPROACHES

The problem to minimize the makespan of a two-stage flow shop is NP-hard (Gupta, 1988). Similarly, the four-stage problem is also NP-hard with a simple reduction on the two-stage one. The development of a polynomial algorithm, which guarantees an optimal solution in a reasonable computational time, is unlikely possible. Thus, dealing with the allocation and the sequencing parts of the problem independently can be a key element to obtain near optimal solution in reasonable computational time. This implies that first a new solution for the allocation is obtained through the GA or a heuristic. Consequently, four single machine problems with family major and minor setup times emerge on the first production stage and five single machine problems with sequence-independent setup times arise on the second production stage. Finally, a single machine problem on the third and the fourth stages. For dealing with the sequencing part of the problem, the sequencing algorithm presented by Nahhas et al. (2016) is adapted. For solving the allocation part of the problem, two hybrid approaches are presented and compared to GA. The first hybrid solution strategy presented is the Improved Integrated Simulation Based Optimization (IISBO). The second solution strategy involves another combination of the ISBO and GA (ISBO & GA). Finally, those approaches will be compared to GA.

4.1. Improved Integrated Simulation Based Optimization (IISBO)

This solution approach is based on the ISBO solution strategy. The conducted analysis in Nahhas et al., (2016) revealed a considerable potential of the ISBO in terms of minimizing the makespan and the total number of major setup times in comparison to the Tabu Search (TS) and Simulated Annealing (SA). In the context of this work, some sensitive parameters of the ISBO are identified and further passed to the Genetic Algorithms (GA). Thus to present a hybrid approach that size the advantages of the metaheuristic (high-quality solution, robustness, etc.) and propose quick solutions likewise the heuristic. This implies that the GA deals with the investigated problems indirectly through optimizing the performance of the heuristic.

Briefly, the ISBO is a constructive heuristic, in which the production schedule is obtained in a single simulation run using an integrated sequencing and allocation algorithms in a simulation model. The ISBO conceptual design is based on a periodical balancing of the production load of the high priority jobs between the available machines during the simulation. The reallocation processes are conducted based on a static predefined reallocation event-list by the end of each simulated working day during the scheduling period. However, this static behavior might lead in some instances to violations in the delivery dates of jobs, if an inappropriate reallocation process is triggered. Furthermore, the production load is balanced between the machines taking into consideration a static balancing indicator of four working days, which might not be optimum for all cases. The allocation algorithm reallocates the families based on the balancing indicator, in which the next four highest priorities are

roughly equally distributed between the machines to avoid unnecessary major setups and sustain a balance in the production load between the machines. Accordingly, in spite of the outperformance of the ISBO in terms of minimizing the makespan and the total number of major setup times, some violations in the delivery date of jobs have been encountered. The IISBO is a combination of GA and the ISBO, in which the GA are employed to optimize the reallocation event-list and the balancing period as demonstrated in the conceptual design of the IISBO in Figure 2. Those parameters are encoded in a genetic representation to optimize the performance of the ISBO using the GA.
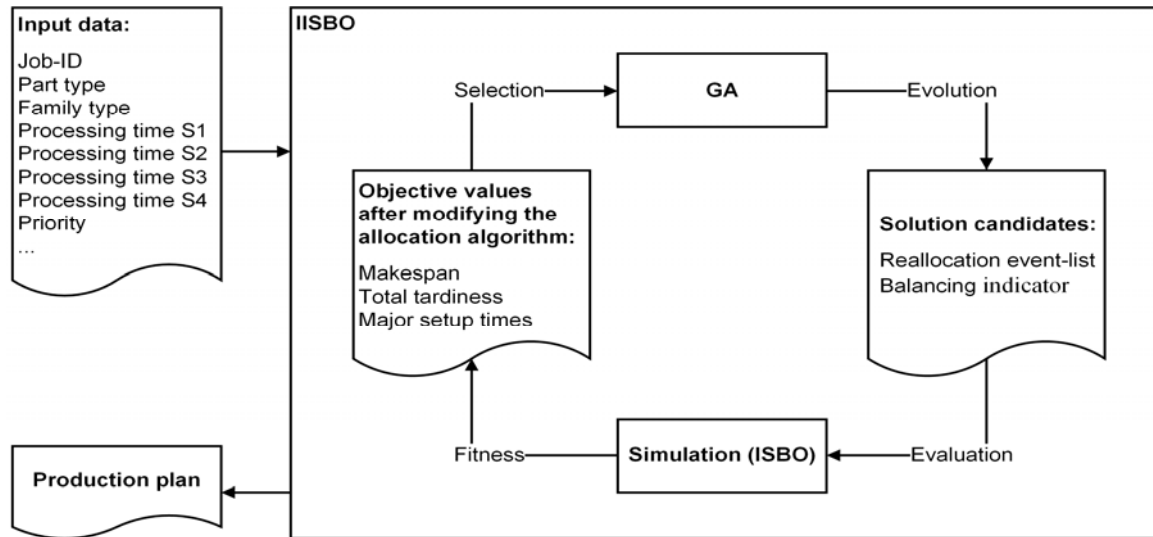

Figure 2: The conceptual design of the IISBO.

The optimization process starts by randomly generating a population of different possible sets of the parameter by the GA for the ISBO. This implies that a genome of the GA is a set of the required parameters of the ISBO. Then the solution candidates or the different parameter sets are passed to the ISBO simulation model to be evaluated, in which the sequencing algorithm presented by Nahhas et al. (2016) is implemented to build the sequence of jobs on each SMD machine. For the dispatching on the Automatic Optical Inspection (AOI) processing stage and the Selective Soldering (SS) processing stage, the Earliest Due Date (EDD) priority dispatching rule is used to optimize the total tardiness taking into consideration the required processing time of jobs on the third and fourth processing stages. Moreover, the sequencing algorithm is used to dispatch jobs on the Conformal Coating (CC) processing stage. The allocation algorithm access the current solution candidate of the GA to trigger the reallocation process during the simulation. In addition, the families are balanced between the machines based on the balancing indicator, which is also generated by the GA.

After evaluating the solution candidates based on the objective function and assigning fitness values, tournament selection strategy is used to select parents for evolving a new generation of solution candidates. For a comprehensive discussion about tournament selection strategy for GA, one can refer to the contribution of Miller and Goldberg (1995). One of the main advantages of this selection strategy is the opportunity to select considerably low-quality solution candidates, which contribute in maintaining a higher

diversity among the solution candidates to avoid being trapped in local optimum caused by false convergence. From the selected solution candidates, a uniform crossover operator is used to generate the genes of the offspring solution candidates. Thereafter, based on a random mutation rate the genes of some individuals are mutated to maintain diversity in the population of solution candidates. Moreover, elitism strategy is implemented in this GA. Elitism strategy simply ensures that the best solution candidate so far found in the search process will survive to the next generation (Konak et al., 2006, p. 1001). After creating the new generation, the solution candidates are evaluated using the simulation model to assign fitness values before starting the selection process again, if the GA did not converge or the maximum number of generations is reached. The convergence function is based on calculating the relative difference between the best and the worst solution candidates in the population using the mean and/or the median of their fitness values.

4.2. Combination of GA and ISBO (GA & ISBO)

The second hybrid approach is also a combination between the same GA and the ISBO. However, the GA deals with the problems directly in this solution approach. The encoding strategy of the GA is targeting the allocation part of the problem on the first processing stage (SMD). This implies that a genome of the GA represents the allocation of families to the SMD machines. The sequencing part of the problem is treated using the same sequencing algorithm identically to the previous approach. However, for evaluating the solution

candidates of the GA, the ISBO simulation model is used, in which the allocation of the GA is then manipulated during the simulation through the allocation algorithm of the ISBO. In this solution strategy, the GA propose in essence the initial allocation for the ISBO before starting the simulation. A conceptual representation of the solution strategy is represented in Figure 3.
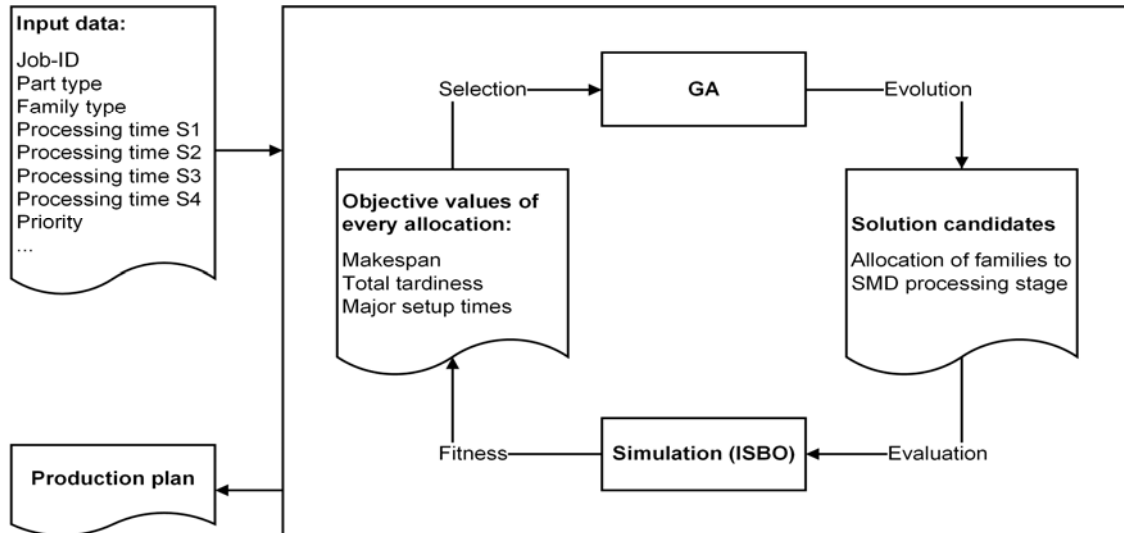
The third solution approach is based on GA. The solution candidates of the GA are encoded to deal with the allocation part of the problem on the SMD processing stage. The rest of the logic is identical to the other solution approaches. In addition, the solution candidates are evaluated using a normal simulation model that describes the behavior of the considered production system.

Figure 3: The conceptual design of the (GA & ISBO).

5. COMPUTATIONAL RESULTS

The presented approaches are evaluated by solving four problem instances of the two- and the four-stage problems. The experiments are designed to investigate the quality of the obtained schedules based on the objective functions as well as the required computational effort to obtain the solutions. The solution approaches are implemented and integrated into a simulation model that is built using the ExtendSim 9.1 simulation package. A descriptive information of the used datasets is presented in Table 1. The processing times are given in minutes.

Table 1: Descriptive information of the input data.

	Number of jobs	Number of families	Accumulated SMD processing time	Accumulated AOI processing time	Accumulated SS processing time	Accumulated CC processing time
Dataset 1	164	41	54,685	72,528	-	-
Dataset 2	170	37	62,345	88,702	-	-
Dataset 3	175	36	61,274	74,738	-	-
Dataset 4	143	35	56,250	79,294	-	-
Dataset 5	181	21	55,344	68,952	22,803	21,155
Dataset 6	179	23	65,470	81,601	21,661	17,749
Dataset 7	194	24	44,270	55,576	22,040	20,439
Dataset 8	170	29	55,457	68,585	23,791	20,442

In order to facilitate the comparison between the solution approaches, a convergence function is used to break the optimization process in all approaches if they converge to 99 %. The convergence function is based on calculating the relative difference between the best and the worst solution candidates of the GA population using the mean of their fitness values. To overcome the stochastic nature of the presented approaches, ten optimization runs have been recorded for solving each problem instance for all approaches. The population size used in the IISBO, the (GA&ISBO) and the GA are 15, 25 and 50 respectively. The applied maximum number of generation is 1000. The used mutation rate is 0.4. Those values have been empirically obtained based on an intensive analysis. A weighted sum approach has been adopted to formulate the objective function. The makespan, the total number of major setup times and the total tardiness are weighted with 0.4, 0.2 and 0.4 respectively. The considered scheduling period is four weeks with three shifts operating model. We reduced

the overall production capacity in 10 % to considered machine breakdowns in an aggregated form.

The computational results for solving the problems are presented in Figure 4. The experiments are conducted on a computer with the following characteristics: CPU 4 x 2.6 GHz, RAM 8 GB and Windows operating system. As demonstrated, a clear domination by a single solution technique cannot be concluded. However, the IISBO and the (GA&ISBO) share a clear domination over the GA and the ISBO in terms of minimizing the makespan and the total number of major setup times. More precisely, the (GA&ISBO) deliver the best makespan on three occasions for solving the two-stage problem, whereby the IISBO also reports the best

makespan in three occasions for solving the four-stage problem. However, the IISBO clearly outperforms the other hybrid approach in terms of the required computational effort for solving all problem instances. In the same context, both approaches deliver optimized production schedules in considerably less computational time in comparison to the GA. Some conclusions on the total tardiness can be drawn. Clearly, the ISBO performs worst in comparison to the other solution approaches since seven penalties are indicated in all datasets. On the contrast, the GA deliver optimized production schedules for all datasets without reporting violations in the delivery dates.
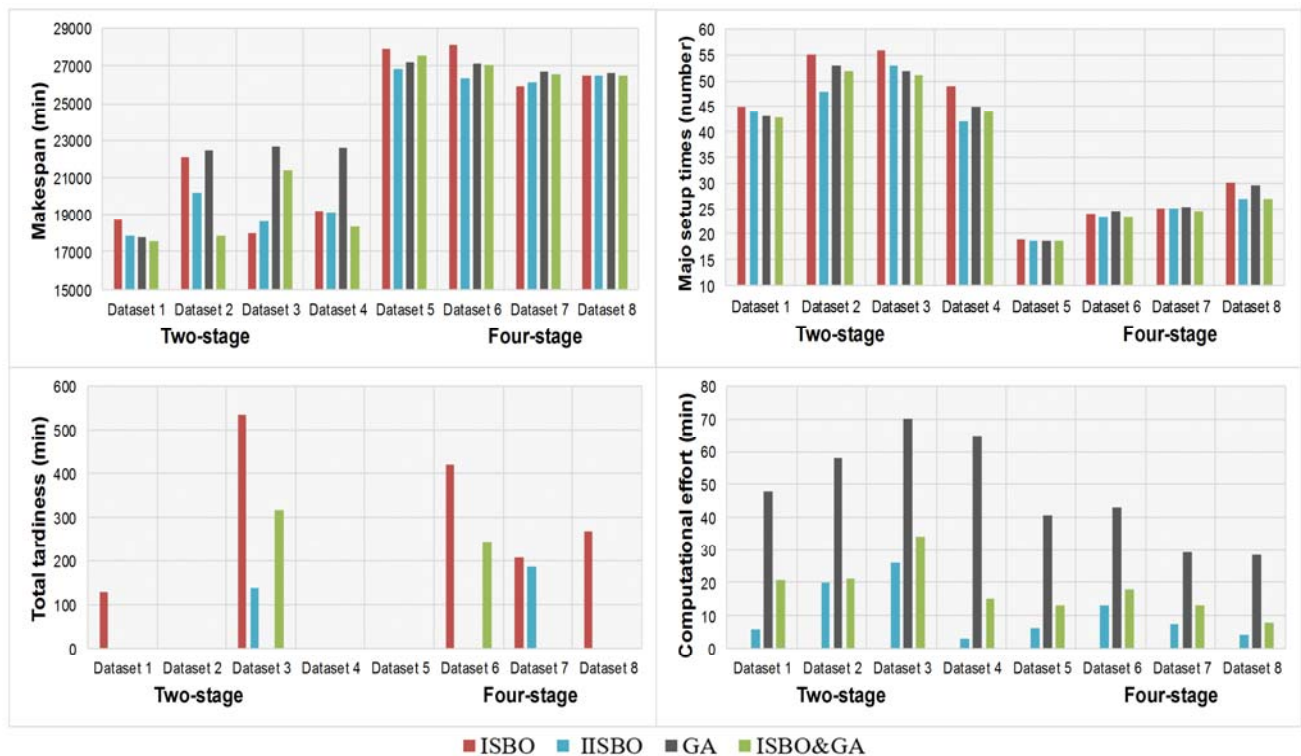

Figure 4: The computational results of the solution approaches for solving the problems.

6. CONCLUSION AND FUTURE WORK

In this paper, two hybrid solution approaches: improved integrated simulation-based optimization (IISBO) and the combination of the GA and ISBO (GA&ISBIO) are evaluated against the genetic algorithm (GA) and the (ISBO) for solving hybrid flow shop (HFS) scheduling problems. The performance of the IISBO and the hybrid approach is pretty similar in minimizing the makespan, the number of major setup times and the total tardiness. The IISBO outperforms the hybrid approach in terms of the required computational time to obtain the solutions. In average, the IISBO obtains the solutions in roughly 50 % of the required computational time of the hybrid approach. Furthermore, the hybrid approach clearly dominates the GA in optimizing the objective values as well as in the required computational time. In addition, the designed improvement strategy for the ISBO did not only overcome the drawback of the total tardiness but

also reports significant improvements in terms of optimizing the makespan and the total number of major setup times.

The performance of the GA is strongly impacted by the random initialization of the solution candidates at the beginning of the optimization, which requires GA longer computational time to overcome the very low-quality initial population. This can be explained by a strongly unbalanced allocation of families on the SMD processing stage, which accordingly leads to an increase in the makespan and number of major setup times. Therefore, developing a simple heuristic or adapting some introduced heuristic in the literature might be a matter of further research to generate the start population of the GA to overcome the computational effort drawback.

We conducted ten optimization experiments on four problem instances of the two and the four stages

problems, which allows us to conclude that the quality of presented solution approaches is stable for solving many problem instances. The increase of the complexity for solving the four-stage problem leads to a decrease in the optimization margin, which the solution approaches strive to achieve. This statement can be supported by observing the performance of the solution approaches in the last four datasets and specifically for minimizing the makespan and the number of major setup time objective values. We can notice that the difference in the performance of the solution approaches in the first four datasets witnessed stronger deviation, whereby in the last four datasets, slight deviations can be noticed. In addition, the required computational time did not increase for solving the last four problem instances with the increase of the complexity in comparison to the first four problem instances. In fact, it rather decreases due to the less number of considered families in the four-stage problem instances.

REFERENCES

Andersson, M., A. H. C. Ng, and H. Grimm. 2008. "Simulation Optimization for Industrial Scheduling Using Hybrid Genetic Representation." In *Proceedings of the 40th Conference on Winter Simulation,* 2004–2011: Winter Simulation Conference.

Aurich, P., A. Nahhas, T. Reggelin, and J. Tolujew. 2016. "Simulation-based Optimization for Solving a Hybrid Flow Shop Scheduling Problem." In *Proceedings of the 2016 Winter Simulation Conference,* 2809–2819, Piscataway, NJ, USA: IEEE Press.

Baker, K. R., and D. Trietsch. 2009, Principles of Sequencing and Scheduling, Wiley-Blackwell: Oxford.

Geiger, C. D., R. Uzsoy, and H. Aytuğ. 2006. "Rapid Modeling and Discovery of Priority Dispatching Rules: An Autonomous Learning Approach." Journal of Scheduling, 9 (1), 7–34.

Glover, F. 1989. "Tabu search—part I." ORSA Journal on computing, 1 3, 190–206.

2003, Handbook of Metaheuristics, Springer Science & Business Media. Glover, F., and G. A. Kochenberger, eds.

Goldberg, D. E. 1989, Genetic algorithms in search, optimization, and machine learning, Addison-Wesley: Reading, Mass., Wokingham.

Graham, R. L., E. L. Lawler, J. K. Lenstra, and Rinnooy Kan, A. H. G. 1979. "Optimization and Approximation in Deterministic Sequencing and Scheduling: a Survey." In *Discrete Optimization II Proceedings of the Advanced Research Institute on Discrete Optimization and Systems Applications of the Systems Science Panel of NATO and of the Discrete Optimization Symposium co-sponsored by IBM Canada and SIAM Banff, Aha. and Vancouver,* edited by E. J. P.L. Hammer and B.H. Korte, 287–326: Elsevier.

Gupta, J. N. D. 1988. "Two-Stage, Hybrid Flowshop Scheduling Problem." The Journal of the Operational Research Society, 39 4, 359.

Held, M., and R. M. Karp. 1961. "A Dynamic Programming Approach to Sequencing Problems." In *Proceedings of the 1961 16th ACM National Meeting,* 196–210, New York, NY, USA: ACM.

Holland, J. H. 1975, Adaptation in natural and artificial systems. *An introductory analysis with applications to biology, control, and artificial intelligence / by John H.Holland*, University of Michigan Press: Ann Arbor, Mich.

Hunsucker, J. L., and J. R. Shah. 1994. "Comparative performance analysis of priority rules in a constrained flow shop with multiple processors environment." European Journal of Operational Research, 72 (1), 102–114.

Johnson, S. M. 1954. "Optimal two- and three-stage production schedules with setup times included." Naval Research Logistics Quarterly, 1 (1), 61–68.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. "Optimization by simulated annealing." Science (New York, N.Y.), 220 4598, 671–680.

Kis, T., and E. Pesch. 2005. "A review of exact solution methods for the non-preemptive multiprocessor flowshop problem." European Journal of Operational Research, 164 3, 592–608.

Konak, A., D. W. Coit, and A. E. Smith. 2006. "Multi-objective optimization using genetic algorithms: A tutorial." Reliability Engineering & System Safety, 91 9, 992–1007.

Kut C. So. 1990. "Some Heuristics for Scheduling Jobs on Parallel Machines with Setups." Management Science, 36 4, 467–475.

Lenstra, J. K., A. Rinnooy Kan, and P. Brucker. 1977. "Complexity of Machine Scheduling Problems." In *Studies in Integer Programming,* edited by P.L. Hammer, E.L. Johnson, B.H. Korte, and G.L. Nemhauser, 343–362: Elsevier.

Li, S. 1997. "A hybrid two-stage flowshop with part family, batch production, major and minor set-ups." European Journal of Operational Research, 102 (1), 142–156.

Maccarthy, B. L., and J. Liu. 1993. "Addressing the gap in scheduling research: a review of optimization and heuristic methods in production scheduling." International Journal of Production Research, 31 (1), 59–79.

Miller, B. L., and D. E. Goldberg. 1995. "Genetic algorithms, tournament selection, and the effects of noise." Complex systems, 9 3, 193–212.

Mirsanei, H. S., M. Zandieh, M. J. Moayed, and M. R. Khabbazi. 2011. "A simulated annealing algorithm approach to hybrid flow shop scheduling with sequence-dependent setup times." Journal of Intelligent Manufacturing, 22 6, 965–978.

Nahhas, A., P. Aurich, T. Reggelin, and J. Tolujew. 2016. "Heuristic and Metaheuristic Simulation-Based Optimization for Solving a Hybrid Flow Shop Scheduling Problem." In *The 15th*

International Conference on Modeling and Applied Simulation, edited by A. G. Bruzzone, et al., 95–103, RENDE (CS), ITALY.

Nawaz, M., E. E. Enscore, and I. Ham. 1983. "A heuristic algorithm for the m-machine, n-job flow-shop sequencing problem." Omega, 11 (1), 91–95.

Oğuz, C., and M. F. Ercan. 2005. "A Genetic Algorithm for Hybrid Flow-shop Scheduling with Multiprocessor Tasks." Journal of Scheduling, 8 4, 323–351.

Orlin, J. B., A. P. Punnen, and A. S. Schulz. 2004. "Approximate Local Search in Combinatorial Optimization." In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms,* 587–596, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Pinedo, M. L. 2012, Scheduling: Theory, Algorithms, and Systems, Springer New York.

Pirlot, M. 1996. "General local search methods." European Journal of Operational Research, 92 3, 493–511.

Reeves, C. R. 1995. "A genetic algorithm for flowshop sequencing." Computers & Operations Research, 22 (1), 5–13.

Ribas, I., R. Leisten, and J. M. Framiñan. 2010. "Review and classification of hybrid flow shop scheduling problems from a production system and a solutions procedure perspective." Computers & Operations Research, 37 8, 1439–1454.

Ross, P. 2005. "Hyper-Heuristics." In *Search Methodologies,* edited by E. K. Burke and G. Kendall, 529–556, Boston, MA: Springer US.

Ruiz, R., and C. Maroto. 2006. "A genetic algorithm for hybrid flowshops with sequence dependent setup times and machine eligibility." European Journal of Operational Research, 169 3, 781–800.

Ruiz, R., and J. A. Vázquez-Rodríguez. 2010. "The hybrid flow shop scheduling problem." European Journal of Operational Research, 205 (1), 1–18.

şerifoğlu, F. S., and G. Ulusoy. 2004. "Multiprocessor task scheduling in multistage hybrid flow-shops: a genetic algorithm approach." Journal of the Operational Research Society, 55 5, 504–512.

Voß, S. 1993. "The Two — Stage Hybrid — Flowshop Scheduling Problem with Sequence — Dependent Setup Times." In *Operations Research in Production Planning and Control,* edited by G. Fandel, et al., 336–352: Springer Berlin Heidelberg.

Wang, K.-J., J. C. Chen, and Y.-S. Lin. 2005. "A hybrid knowledge discovery model using decision tree and neural network for selecting dispatching rules of a semiconductor final testing factory." Production Planning & Control, 16 7, 665–680.

Wittrock, R. J. 1990. "Scheduling parallel machines with major and minor setup times." International Journal of Flexible Manufacturing Systems, 2 4, 329–341.

Zheng, D.-Z., and L. Wang. 2003. "An Effective Hybrid Heuristic for Flow Shop Scheduling." The International Journal of Advanced Manufacturing Technology, 21 (1), 38–44.

AUTHORS BIOGRAPHY

ABDULRAHMAN NAHHAS is a research associate in the Very Large Business Application Lab at the Otto von Guericke University Magdeburg. His main research and work interests are IT-system landscape engineering, scheduling problems and simulation-based optimization solution strategies. Abdulrahman Nahhas holds a bachelor and a master degree in Business Informatics from Otto von Guericke University Magdeburg. His email address is abdulrahman.nahhas@ovgu.de.

PAUL AURICH is a master student of Industrial Engineering with specialization in production systems at the Otto von Guericke University Magdeburg. His main work interests are scheduling problems and simulation-based optimization solution strategies. Paul Aurich holds a bachelor degree in Industrial Engineering with specializing in logistics from Otto von Guericke University Magdeburg. His email address is paul.aurich@mail.de.

SASCHA BOSSE is a research coordinator and a project manager at the Very Large Business Application Lab. His professional field of interests includes IT service management, modeling, simulation, and optimization. Sascha Bosse received a doctoral degree in engineering from the Otto von University Magdeburg. Furthermore, he holds a bachelor and a master degree in Business Informatics from Otto von Guericke University Magdeburg. His email address is sascha.bosse@ovgu.de

TOBIAS REGGELIN is a project manager at the Fraunhofer Institute for Factory Operation and Automation and Otto von Guericke University Magdeburg. His main research and work interests are logistics system modeling and simulation and the development and conduction of logistics management games. Tobias Reggelin received a doctoral degree in engineering from Otto von University Magdeburg. Furthermore, he holds a diploma degree in industrial engineering in logistics from Otto von Guericke University Magdeburg and a master's degree in Engineering Management from Rose-Hulman Institute of Technology in Terre Haute, IN. His email address is tobias.reggelin@ovgu.de.

Klaus Turowski is heading the Chair of Business Informatics (AG WI) at the Otto von Guericke University Magdeburg, the Very Large Business Applications Lab and the world's largest SAP University Competence Center. His professional field of interests includes IT service management, IT system landscape engineering, SAP systems, and optimization. Klaus Turowski achieved his doctorate at the Institute for Business Informatics at the University of Münster

and habilitated in Business Informatics at the Faculty of computer science at the Otto von Guericke University. His email address is klaus.turowski@ovgu.de.

IMMERSIVE, INTEROPERABLE AND INTUITIVE MIXED REALITY FOR SERVICE IN INDUSTRIAL PLANTS

Agostino G. Bruzzone [a], Marina Massei [b], Letizia Nicoletti [c],
Kirill Sinelshchikov [d], Riccardo di Matteo [e], Marina Cardelli [f], Praveen Kuman Kandunuri [g]

[a] [b] [c] Simulation Team, DIME University of Genoa, [c] Simulation Team, CAL-TEK Srl,
[c] Simulation Team, Liophant Simulation, [d] Simulation Team, SIM4Future, [f] [g] Simulation Team, MAST

[a] agostino.bruzzone@simulationteam.com, [b] marina.massei@simulationteam.com,
[c] letizia.nicoletti@simulationteam.com, [d] kirill.sinelshchikov@simulationteam.com,
[e] riccardo.dimatteo@simulationteam.com, [f] marina.cardelli@simulationteam.com,
[g] praveen.kuman.kandunuri@simulationteam.com

[a] [b] [c] [d] [e] [f] [g] www.simulationteam.com

ABSTRACT

The authors propose an innovative Mixed Reality solution representing an immersive intuitive and interoperable environment to support service in industrial plants. These methodologies are related to concepts of Industry 4.0. Solutions based on a mix of VR and AR (Virtual and Augmented Reality) with special attention to the maintenance of industrial machines; indeed the authors propose an overview of this approach and other synergistic techniques. Moreover, alternative instruments are presented and their specific advantages and disadvantages are described.

Particularly, the approach is based on the SPIDER, an advanced interoperable interactive CAVE developed by the authors which supports cooperative work of several users involved in training, troubleshooting and supervision are proposed.

Last but not least, an overview of projects using same techniques in other fields, such as construction, risk assessment, Virtual Prototyping and Simulation Based Design is presented.

Keywords: Augmented Reality, Virtual Reality, Mixed Reality, Industrial Plants, Maintenance & Operation, Plant, Service, Interoperable Simulation

1 INTRODUCTION

The use of Virtual and Augmented Reality is pending as explosive technology since several years; it is still very fresh in our memory the promise never accomplished of Google Glasses as well as the Oculus Rift new version that was downsized after acquisition by Samsung and promotion of low cost smartphone head mounted displays (Chafkin 2015; Scudellari 2016); despite these limited achievement respect commercial expectations for 2015 and 2016, it is evident the big growth of VR & AR (Virtual & Augmented Reality) applications and their potential (Wagner 2016). The very interesting aspects about the advance in this area is the possibility to recombine different technology, reduce costs, increase capabilities and ramp up reliability in order to create solutions able to be usable in a wide spectrum of applications (Bruzzone et al., 2016a).

Fig.1 –Guided Procedures and Troubleshooting Processes of an Industrial Skid within MR SPIDER

Considering these aspects, the potential for industry is great and especially in training and education, service, maintenance, remote tutoring and troubleshooting; (Spanu, 2016). For instance a peculiar application could be related to safety training and emergency planning/simulation in chemical plants involving batch and semi-batch reactors processing reactive chemicals and relevant storage of flammable liquid and gases (Fabiano et al. 2015; Palazzi et al. 2017). This paper proposes researches derived from industrial projects that have been extend over a specific CAVE (Cave Automatic Virtual Environment)where it is possible to combine AR & VR to develop new Mixed Reality (MR) solutions for Service of Plants and Skids.

2 SIMULATION, AR & VR FOR INDUSTRIAL MAINTENANCE

The authors have been extensively involved in application of M&S and VR to maintenance in industrial Plants; in facts in recent years a specific project named SISOM (SImulation SOlution based on virtual and augmented reality for Maintenance) allowed to develop new solutions and to demonstrate the potential of modern VR (Virtual Reality) and AR (Augmented Reality) for industrial maintenance (Bruzzone, 2016a). In this field several solutions and technologies are available such as Laptop, Head Mounted Display, Glasses, CAVE, Smartphone.

Fig.2 – Cooperating Environment: Local Review on CAVE,Virtual Distributed Service between Site & Office

Fig.3 – Guided Start Up Procedure by Mixed Reality within the Interactive Interoperable Immersive CAVE

In facts in SISOM and other projects the authors evaluated the specific potential and shortfalls of different solutions in order to identify best configurations for specific applications. In facts it is very important to find evidence of their strongholds to properly develop immersive virtual ambient that could be reliable and suitable for industrial applications (Bruzzone et al., 2016a). During SISOM project the Simulation Team, with special cooperation among members from Genoa and Cosenza developed solutions able to run over multi-plaforms: in this way it was possible to operate from very light and compact platform by based on tablets, smart phones, HDV and virtual glasses, as well as from more immersive and interactive framework such as CAVE presented in figure 1. It resulted that the CAVE approach could be very interesting in case the use involve cooperative work among more people that could easily enter in it and interact directly with the virtual world and among their selves in an intuitive way. This solution is obviously more expensive, but it is pretty good in case of remote supervision for troubleshooting and industrial service and in facts this mode allows participation of multiple users maintaining at the same time an high level of interactivity. In this paper it is summarized this approach as well as the capabilities of these technologies and instruments along with opportunities provided by integration of the new solutions in Industry 4.0 which foresee connection of IoT (Internet of Things), Data Farms and Smart Devices in the field of support of industrial maintenance are presented.

3 INDUSTRY 4.0 & MIXED REALITY SERVICE

It's evident that adopting Industry 4.0 assumptions to create innovative solutions for Industrial Service based on Mixed Reality and Simulation is a guarantee to support supervision and assistance to personnel in case of problems and during of execution of complex procedures. Hence these new services are expected to allow a sensible reduction on the number of interventions of high-qualified personnel on site; this will turn to be possible by resolving significant amount of critical problems operating from offices and providing support in cooperative way utilizing opportunities provided by VR and AR.

Obviously this approach allows to improve drastically safety and efficiency of operations with heavy savings. Furthermore the skilled people in charge of commissioning and on-site service are usually pretty rare. So this approach multiples their availability providing them the opportunity to work from the main office; this reduces their traveling around the world and makes it possible to serve customers in higher numbers as well as plant facilities that lay on not accessible locations.

4 TECHNOLOGIES AND APPLICATIONS

Improving availability and reliability of industrial plants as well as safety by using solutions based on VR and AR is one of most promising applications in this field (Tatić et al. 2017). In facts many R&D projects have been developed in this area and specific networks and consortiums have been created in Europe to address these issues (Quero et al. 2012; Pérez-Ara et al. 2013).

One of the ways VR and AR could assist mentioned activities is to guarantee "telepresence", which allows to move virtually high qualified personnel in the place where presence of its knowledge and professional capacity is required, in the same time it's possible to provide training different from the traditional face to face (Alzahrani et al. 2014; Zhang et al. 2014; Safir et al. 2015). In fact last years there were performed researches in this field (Peña-Rios et al. 2016), one of them is dedicated to airplane maintenance (Gonzalez-Franco et al. 2017). There are several interesting applications of cooperation and remote assistance in industrial field utilizing VR based on CAVE, for example virtual environments which could be connected to share data (Bruzzone et al., 2011b; Dai 2012).

VR and AR have been applied also in mining industry, with scope to increase productivity and safety of miners and operators, which control machines used in extraction and transportation of raw materials, using HMD (Head Mounted Display) with scope of training and orientation inside mine (Benes & Kodym 2014). Some applications have been developed also in the field of construction, which is still influenced by risks caused by lack of training of personnel (Le et al. 2015, Perlman et al. 2014).

Fig.4 – Interactive with CAVE Walls while applying a procedure on Driver Rack through Mixed Reality

By the way, another improvement of productivity and safety in this field could be achieved by remote training based on above mentioned concepts.

Immersive VR demonstrated its potential also as instrument of instruction for operators and students, creating non physical training classes, that could be worldwide distributed and could provide quite comparable results respect face-to-face lessons' (Bower et al. 2016).

In this sense there are still concerns about the capability to operate by VR in remote classes with equivalent results, therefore it is evident that technology evolution is expected to overpass some of this shortfalls, while cost and time saving will force to further extend these applications (Seidel & Chatelier 2013).

5   MIXED REALITY CHARACTERISTICS

Mixed Reality (MR) deals with the capability to create Virtual Environments where Augmented information and interfaces could be used to multiply usability, visibility, efficiency and many other indicators (Lindgren et al. 2016).

Indeed to improve usability of MR to multiple applications and by multiple users, it is very useful to guarantee compatibility of different technologies in proposed solutions: devices, MR and their connection; in some specific cases Virtual Environments are generated to allow several users to act individually and in collaboration through their Avatar representation Avatars (Biocca & Levy 2013).

Another important aspect which must be taken into account is related to immersion capability of the MR solution; indeed this aspect guarantees a quite complete engagement of the users, changing them from passive watchers of movies/animation to active participants of interactive activities within the virtual world (Sherman & Craig 2003).

In facts within VR applications, the user must interact with world, becoming "actively immersed" (Nakatsu & Tosa 2000).

In facts this status creates user's temporal estrangement from the world outside of the Virtual Environment even if virtual world is not perfectly real or even realistic and caused mainly by deep involvement of the user (Pimentel & Teixeira 1993).

As already mentioned, the interaction is a fundamental aspect of correct expression of MR, and it is guaranteed by sensors of different kind such as positioning sensors for eyes, fingers or muscles, otherwise motion capture, quick response graphics, touch devices, voice recognition systems and joysticks (Bowman & Hodges 1999).

6   THE CAVE: COLLABORATIVE & IMMERSIVE ENVIRONMENT

While it is already stated the importance to operate over multiple solutions, in this paper it is focused the attention on the use of frameworks able to guarantee multiple user engagement and collaboration. In industrial plants, during troubleshooting this is a quite important aspect dealing with the connection between the people on site and that one the office as proposed in figure 2. The collaborative environment between industrial site and central service center office should rely on an distributed virtual environment that could benefit of web simulation services and interoperable simulation (Bednarz et al. 2015). Vice versa the local collaboration in the main service office and in other overseas supervision sites it should rely on solutions able to simplify interpersonal relationship while interacting with the virtual world; due to these reasons the authors propose to adopt innovative new generation CAVEs able to interoperate with Simulation and MR. CAVE systems have been introduced since long time; first appearance are dating since in the beginning of 90s by Illinois University researchers and it is not surprising that even in that case they were focused on providing the capability to visualize a virtual environment for cooperative use (Cruz-Neira et al. 1992). CAVE have been used in many different fields, from military training to medicine to visualize parts of a body giving so an opportunity to prepare for operation in shared environment (Hale et al. 2014). In addition to these fields there are CAVE used in Universities and Industries as virtual show rooms, or in Museum for the reproduction of natural or past environment. In fact, some applications of CAVE are specifically related to entertainment sector (Jacobson & Hwang 2002).

In facts, a CAVEs contains usually a limited space where virtual world is reproduced, but it allow the users to enter and eventually, by most modern solutions, to interact with it (Hale et al. 2014; Bruzzone et al. 2016a). Therefore traditional CAVE have a cubic shape, with images, creating visual part of virtual environment, shown of its sides; screens could be done using different types of material, for example white plastic or mirrors (Hereld et al. 2000). Images could be created using classic direct view otherwise rear projection, which reduce drastically number of components inside CAVE, hence improving its virtual immersion, however this solution is not so comfortable in exploitation and requires bigger external volume and space occupancy of the whole equipment (Hale et al. 2014).

Fig.5 – Reviewing Virtual Handbook while interacting dynamically with the Plant Simulator

In facts the authors developed a special CAVE, named SPIDER (Simulation Practical Immersive Dynamic Environment for Reengineering) designed especially to provide top performance at very low costs with high versatility an interoperability levels.

SPIDER utilize direct view created by Super Short Throw Projectors which are capable to project image on 2 meters width screens from distances around 30cm with particular angles which allows to reduce significantly shadows (Bruzzone et al. 2016c). Much more expensive and large solutions require usually curved screens, up to completely spherical surface (Kenyon et al. 2014).

In any case the CAVE could be effectively integrated with many other I/O devices, for example in SPIDER it could be placed a motion capture system as well as motion platform. In facts it is possible to install a CAVE over a large platform as it is widely done for commercial flight simulators (Muhanna 2015).

6 SPIDER: INNOVATIVE CAVE

Simulation Team has used its SPIDER for several projects including SISOM in order to test the related versatility and performance; from this point of view, the SPIDER represent a innovative compact and movable solution which allows to evaluate different combinations of real equipment and immersive virtual environment (Bruzzone et al. 2016b). The SPIDER is a compact innovative interoperable and scalable thanks to its design based on modules suitable to be installed in a standard High Cube Container, in fact main SPIDER dimensions are 2m x 2m x 2.6m to fit in these kind of box and shelters.

SPIDER is designed to work in a federation of simulators with other interoperable simulators by most advanced standards. Another interesting SPIDER feature is represented by his touch screen technology: indeed the SPIDER screen surfaces are interacting with user using direct touch captured by laser scanners and/or by tracking electronic pencils. The interactivity on the multiple big screens (2m width each and up to 4 horizontally plus up to two more for top and/or bottom) is extremely intuitive due to the fact that users could touch elements in the MR corresponding to simulation objects; in this way it becomes possible to assign tasks, to give orders, to use them or just to require specific

information easily within complex environment. All these functions turn to be immediately available to CAVE guests just by a direct touch by their fingers or by special electronic pens. These aspects guarantees an high level involvement for users and it could be further combined with other elements such as, for example, sounds, visual effects, vibrations, accelerations and touches. SPIDER solution has been used in R&D applications in fields of defense, marine, transportation and logistics (Bruzzone et al. 2016b). In the last case the BBBUS Simulator (Box Bull Bus Simulator) developed in collaboration with Central Labs of the University of Cagliari for training purposes has demonstrated its capacity to assist training, also procedures, using the same VR as other simulators of cranes (Bruzzone et al. 2011a). In facts, SPIDER is a CAVE which emphasizes interactivity and interoperability characteristics by being compatible with many HLA Simulators (Bruzzone et al. 2016c). In addition, this system could be integrated with several biometric devices (e.g.cardiofrequenzimeter, oculometer, muscle tone meter) to measure human experience as well as with devices such as other motion platforms o motion capture systems (Gonzalez et al. 2016).

In the same time a lot of devices could be connected to the system, for example various I/O systems (audio, motion) which are capable to make user experience inside SPIDER deeply immersive, allowing efficient supervision and analysis to bigger amount of persons, which dealing with dynamic virtual presence in VR activating different functions and subsystems.

Even more, the interoperability of the SPIDER guarantees the possibility to connect this environment to simulators of different kind using federation, which allows to represent different parts of machines (electric, mechanic, hydraulic and control parts).

This proposed solutions is very effective in terms of allowing the Subject Matter Expert (SME) to supervise operations remotely, in fact SME become available to control remotely distributed systems without leaving office, hence optimizing availability of limited high qualified human resources. In this way it becomes possible not only to supervise and/or control even geographically distributed systems, but also to provide training both in office or on field sites through internet in centralized way. In fact the applications of SPIDER represent an interesting example of new generation CAVEs able to provide new efficient ways of collaboration over the web and on site.

In fact in case of supervision by groups of 4 or 5 persons, the SPIDER allows to all of them to remain inside the CAVE and, at the same time, to move them and interact with machines and operators that are on the real plant on the field by reading data from various sensors, receiving video and audio feeds from remote cameras, etc. It is important to state the use of simulation models allow to combine this synthetic environment with functionalities of plants, skids and machines, to make possible to carry out virtual experimentations as well as to test in advance

procedures (Whisker et al. 2003); vice versa the use of interoperable simulation based on HLA Standard keeps open the possibility to federate the SPIDER also with real equipment to act as an element of a much complex simulation or emulation (Liu & Theodoropoulos 2014). Feedback could be provided using MR, by combining VR and AR; for example a remote assistants could send data through his tablet to the CAVE where experts review the situation and provide suggestions and directions back to him by using AR. For instance it could be possible to superimpose images from the tablet's camera and indications from SME (Subject Matter Experts), such as markers on control panel with description of sequence of procedures to be done, received from the remote assistant (See figures 3-5). Another potential application is related to the possibility to demonstrate on screen the procedures and operations; for instance the virtual models of a skid could be augmented with explanations of required actions and images related to this specific case and overlapped within the cave and on the field tables in a synchronized way based on data obtained from camera. In fact this approach, now it could be possible to address multiple problems in different areas, such as service, maintenance, safety, E&T. Combined use of VR and AR for remote assistance has already demonstrated interesting results and its efficiency in industry, at the same time the potential for training was experienced in SISOM Project (Bruzzone et al. 2016b); in these case experiments have been conducted by involving students and researchers in reference to industrial case studies in different sites.

8   CONCLUSIONS

Nowadays evolution of AR & VR allows to produce new MR solutions with high potential, for these reasons the authors are further developing the capabilities of their SPIDER to address maintenance in Industrial Plants. In this way it is possible to develop new support systems for service and maintenance that enable new distributed collaborative work procedures. This approach is efficient in cases of training of operators and remote control and support. The authors are currently conducting experiences with different user groups to collect important data and to define guidelines for introducing the new MR solutions in industry. In facts several experiments and examinations have been already conducted as well as comparison among many different possible approaches. In this way the MR approach maximizes the impact of multiple innovative M&S techniques into the Industry 4.0 concepts.

In facts, it is as important to extend analysis and theoretical studies by development and experimentation of new conceptual models, simulation architectures and software solutions to evaluate their real benefits on the field. This approach allows to refine the design and to identify new concepts and new solutions utilizing M&S, AR & VR. In facts, based on current researches and experiments it validated and quantified the capability to assist operators both on site and remotely through the SPIDER as example of innovative MR solution.

REFERENCES

Alzahrani, A., Callaghan, V., & Gardner, M. (2014) "Towards the Physical Instantiation of Virtual People and Components in Physical Mixed-Reality Tele-Presence Environments", Proc. of Intelligent Environments Workshops, pp. 285-294, July

Bednarz, T., James, C., Widzyk-Capehart, E., Caris, C., & Alem, L. (2015). "Distributed collaborative immersive virtual reality framework for the mining industry" in *Machine Vision and Mechatronics in Practice.* Springer Berlin Heidelberg. pp. 39-48

Benes, F., & Kodym, O. (2014) "Application of Augmented Reality in Mining Industry", Proc. 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Vol. 1, 1-35, June 19-25

Biocca, F., & Levy, M. R. (2013) "Communication in the Age of Virtual Reality", Routledge, London

Bower, M., Lee, M. J., & Dalgarno, B. (2016) "Collaborative learning across physical and virtual worlds: Factors supporting and constraining learners in a blended reality environment", British Journal of Educational Technology

Bowman, D., Hodges, L. (1999) "Formalizing the design, evaluation and application of interaction techniques for immersive virtual environments", J. Visual Lang. Comput., 37–53

Bruzzone A., Longo F., Nicoletti L., Vetrano M., Bruno L., Chiurco A., Fusto C., Vignali G. (2016a). Augmented reality and mobile technologies for maintenance, security and operations in industrial facilities. 28th European Modeling and Simulation Symposium, EMSS 2016, pp. 355.

Bruzzone A.G., Massei M., Maglione G., Agresta M., Franzinetti G., Padovano A. (2016b) "Virtual and Augmented Reality as Enablers for Improving the Service on Distributed Assets", Proc. of I3M, Larnaca, Cyprus, September

Bruzzone A.G., Massei M., Maglione G.L., Di Matteo R., Franzinetti G. (2016c) "Simulation of Manned & Autonomous Systems for Critical Infrastructure Protection", Proc. of DHSS, Larnaca, Cypurs, September

Bruzzone A.G., Fadda P, Fancello G., Massei M., Bocca E., Tremori A., Tarone F., D'Errico G. (2011a) "Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes", Simulation October 2011, vol. 87 no. 10, p. 857-874, ISSN: 857-874

Bruzzone A., Massei M., Longo F., Madeo F., (2011b). Modeling and simulation as support for decisions makers in petrochemical logistics. Proceedings of the 2011 Summer Computer Simulation Conference, pp. 130.

Chafkin M. (2015) "Why Facebook's $2 Billion Bet on Oculus Rift might one day connect Everyone on Earth", Vanity Fair Hive, October

Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., Hart, J.C. (1992) "The CAVE: audio visual experience automatic virtual environment", ACM Commun., 35 (6), 64–72

Dai, F. (Ed.). (2012) "Virtual reality for industrial applications", Springer Science & Business Media

Fabiano, B., Parentini, I., Ferraiolo, A., Pastorino, R. (1995) "A century of accidents in the Italian industry", Safety Science, 21, 65-74

Gonzalez, D. S., Moro, A. D., Quintero, C., & Sarmiento, W. J. (2016, August). Fear levels in virtual environments, an approach to detection and experimental user stimuli sensation", *Proc. of XXI IEEE Symposium on Signal Processing, Images and Artificial Vision (STSIVA),* pp. 1-6

Gonzalez-Franco, M., Pizarro, R., Cermeron, J., Li, K., Thorn, J., Hutabarat, W. & Bermell-Garcia, P. (2017) "Immersive Mixed reality for Manufacturing Training" Frontiers, 4(3), 1

Hale, K. S., & Stanney, K. M. (2014). "Handbook of virtual environments", CRC Press, Boca Raton, FL

Hereld, M., Judson, I. R., & Stevens, R. L. (2000). "Introduction to building projection-based tiled display systems". *IEEE Computer Graphics and Applications*, *20.* pp. 22-28.

Jacobson J.,Hwang Z.(2002) "Unreal tournament for immersive interactive theater", ACM Comm., 45 (1), 39–42

Kenyon, A., Van Rosendale, J., Fulcomer, S., & Laidlaw, D. (2014). "The design of a retinal resolution fully immersive VR display", *Virtual Reality (VR), IEEE,* pp. 89-90

Le, Q. T., Pedro, A. K. E. E. M., Lim, C. R., Park, H. T., Park, C. S., & Kim, H. K. (2015) "A framework for using mobile based virtual reality and augmented reality for experiential construction safety education", International Journal of Engineering Education, 31(3), 713-725

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016) "Enhancing learning and engagement through embodied interaction within a mixed reality simulation". *Computers & Education*, *95*, 174-187

Liu E., Theodoropoulos G.(2014)"Space-time matching algorithms for interest management in distributed virtual environments", *ACM TOMACS*, vol.24

Muhanna, M. A. (2015) "Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions", Journal of King Saud University-Computer and Information Sciences, 27(3), 344-361

Nakatsu R., Tosa, N.(2000)"Active Immersion: the goal of Communications with interactive agents", Proc. of Int.Conf. on Knowledge-Based Intelligent Engine

Palazzi, E., Caviglione, C., Reverberi, A.P., Fabiano, B. (2017) "A short-cut analytical model of hydrocarbon pool fire of different geometries, with enhanced view factor evaluation", Process Safety and Environmental Protection, August

Peña-Rios, A., Hagras, H., Gardner, M., & Owusu, G. (2016) "A Fuzzy Logic based system for Mixed Reality assistance of remote workforce", Proc. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July, pp. 408-415

Pérez-Ara, M. A., Quero, S., Navarro, M. V., Botella, C., & Baños, R. M. (2013) "Augmented Reality for Safety at Work: Needs Analysis of the Priority Risks for Safety", Proc. of Health Context in Spain, INTED, pp. 812-817

Perlman, A., Sacks, R., & Barak, R. (2014). "Hazard recognition and risk perception in construction". *Safety science*, *64*, pp. 22-31.

Pimentel, K., Teixeira, K. (1992) "Virtual reality: through the New Looking Glass", TAB Books, NYC

Quero, S., Botella, C., Pérez-Ara, M. A., Navarro, M., Baños, R. M., Maciá, M. L., & Rodríguez, E. (2012). The use of Augmented Reality for safety in health: The European Project ANGELS. In ICERI2012 Proceedings (pp. 215-218). IATED.

Safir, I. J., Shrewsberry, A. B., Issa, I. M., Ogan, K., Ritenour, C. W., Sullivan, J., & Issa, M. M. (2015). Impact of remote monitoring and supervision on resident training using new ACGME milestone criteria. Can J Urol, 22, 7959-7964

Scudellari M. (2016) "Google Glass Gets a Second Life in the ER", IEEE Specturm, May 25

Seidel, R. J., & Chatelier, P. R. (Eds.). (2013). "Virtual reality, training's future? Perspectives on virtual reality and related emerging technologies". Springer Science & Business Media.

Sherman, W. R., & Craig, A. B. (2003) "Understanding Virtual Reality—Interface, Application, and Design. Presence", Morgan Kaufmann Publisher, SF, 12(4)

Spanu S., Bertolini M., Bottani E., Vignali G., Di Donato L., Ferraro A., Longo F., (2016). Feasibility study of an augmented reality application to enhance the operators' safety in the usage of a fruit extractor. International Food Operations and Processing Simulation Workshop, FoodOPS 2016, pp. 70.

Tatić, D., & Tešić, B. (2017) "The application of augmented reality technologies for the improvement of occupational safety in an industrial environment", Computers in Industry, 85, 1-1

Wagner K. (2016) "Two Years Later: Facebook's Oculus Acquisition Has Changed Virtual Reality Forever", Recode, March 24

Whisker, V. E., Baratta, A. J., Yerrapathruni, S., Messner, J. I., Shaw, T. S., Warren, M. E., Rotthoff E.S., Winters J.W., Clelland J.A. & Johnson, F. T. (2003). "Using immersive virtual environments to develop and visualize construction schedules for advanced nuclear power plants". In *Proceedings of ICAPP* Vol. 3, pp. 4-7.

Zhang, J., Hou, H. T., & Chang, K. E. (2014) "UARE: Using reality-virtually-reality (RVR) models to construct Ubiquitous AR environment for e-Learning context", Proc. of IEEE Science and Information Conference (SAI), August, pp. 1007-1010

A STRATEGIC SERIOUS GAME ADDRESSING SYSTEM OF SYSTEMS ENGINEERING

Agostino G. Bruzzone [(a)], **Marina Massei** [(b)],
Giovanni Luca Maglione [(c)], **Kirill Sinelshchikov** [(d)], **Riccardo di Matteo** [(e)]

[(a) (b)] Simulation Team, DIME University of Genoa,
[(c) (d)] Simulation Team, Liophant Simulation,
[(e)] Simulation Team, SIM4Future

[(a)] agostino.bruzzone@simulationteam.com, [(b)] marina.massei@simulationteam.com,
[(c)] giovanni.luca.maglione@simulationteam.com, [(d)] kirill.sinelshchikov@simulationteam.com,
[(e)] riccardo.dimatteo@simulationteam.com

[(a) (b) (c) (d) (e)] www.simulationteam.com

ABSTRACT

Serious Games are currently extending their capabilities to strategic Education and Training by innovative approaches and new technological solutions. In this paper, the authors propose a new Serious Game devoted to address such aspects with special focus on System of Systems Engineering (SoSE). The proposed case uses a challenging framework related to the development of an innovative System of Systems for defense and homeland security that could be used by users to acquire the fundamental concepts of SoSE. The scenario allows to investigate alternative interoperable solutions among different platforms, sensors, infrastructures and doctrines respect evolving threats in relation to an air defense solution based on airborne radars.

Keywords: Serious Game, System of System Engineering, Stochastic Simulation, Web Applications, Homeland Security, Airborne Radar

INTRODUCTION

System of Systems Engineering (SoSE) represents a complex sector addressing the development of new solutions that overpass the complexity of the single system to become a common approach created by a really integrated and interactive approach to a challenging problem.

In facts, it is mostly impossible to create labs or exercises on SoSE without using M&S, so the authors decided to proceed by applying MS2G paradigm (Modeling, interoperable Simulation and Serious Game) to prepare a serious game devoted to Education and Training (E&T).

The proposed case study is quite challenging and the use of the simulation engine allows the students to test directly the concepts on a virtual project observing the effectiveness of different techniques and also the need to act in coordinated way with different project stakeholders. The proposed Serious Game, by its simulation engine, allows to estimate the impact of engineering and operational alternatives in terms of effectiveness and efficiency on the overall mission environment.

The authors have already conducted some tests with classes of industrial and academic students and are proceeding to further develop the simulator to be used for this purpose.

2 EDUCATION ON STRATEGIES RELATED TO SoSE

SoSE is a interdisciplinary approach that is popular in reference to development of new large and complex SoS (System of Systems), with the main goal to support a performance evolution based on properly defined requirements able to guarantee empowerment of the overall capabilities (Sousa-Poza et al.2008). Indeed SoSE is quite popular in aerospace and defense, but is in use also for large plants and other kind of SoS such as a new Pool of Power Plants over a region and their grid (Giribone et al.1996; Jamshidi 2011). In these sectors usually the complexity is due to the high number of interactions among different systems and the relative requirements makes pretty difficult to finalize an effective design able to balance efficiency, effectiveness and flexibility, so SoSE represent an important concept (Keating et al. 2005). Indeed the original motivation for introducing SoSE was strongly related to the capability of identifying proper requirements and configurations since the early phase of the project and let them evolve consistently with scenario evolution along the new SoS project life cycle (Jamshidi 2008). Indeed SoS are used to be large and complex and their projects involve big quantities of money and significant durations, so it is crucial to support early design and engineering in order to succeed, especially considering that these phases are affected by the strongest impact on maximizing the SoS performance with lowest efforts (Rhodes et al. 2009).

However, considering the nature of large programs, it is evident that these projects usually involve multiple players cooperating with final users and, quite autonomously, developing, managing and finalizing engineering of their specific systems that are elements of one or more systems; from this point of view, it is necessary to develop an approach able to guarantee that these activities are coordinated without losing the overall picture of the SoS and its performance in a wide spectrum of boundary conditions (Keating et al.2003).
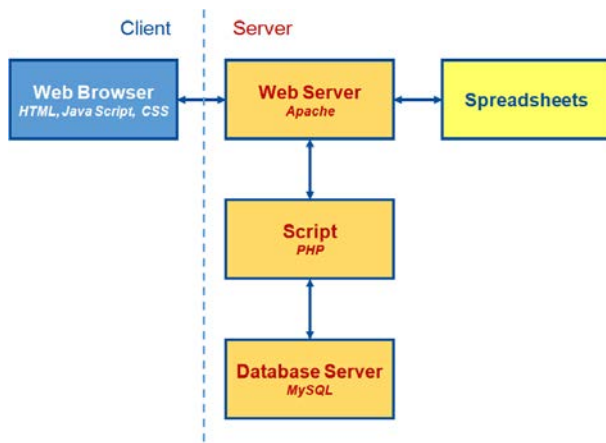
Figure 1 - MISCHIEF Architecture Client and Server

In general sense, the SoSE has to deal with many boundary conditions that affect the scenario including among the others: technical elements & engineering, operational issues, regulations and laws, finance, politics, social aspects, organizational factors; for instance considering the multiple players involved in these projects usually it is crucial to be able to couple the different perspectives of the stakeholders as well as their mutual relationships.

So, it necessary to support the objective and requirement changes along the project life cycle, not only in terms of engineering, but also considering functional and operational needs. In this way it becomes possible to dynamically adapt to the evolving functional requirements and capabilities of the overall SoS. Due to these reasons it is evident the fundamental role of Modeling and Simulation to support this approach and SoSE have been applied to several cases by the author with special attention to defense and aerospace sector, but also operating in industrial plants (Giribone et al.. 1996; Bruzzone & Bocca 2008).

3 SERIOUS GAMES AND STRATEGIC EDUCATION AND TRAINING FOR SoSE

SoSE as emerging approach for complex SoS is pretty interesting, therefore it is often not easy to transfer the capability to apply these concepts moving from a *generic declarative call for improvements* to an *effective renewed engineering approach*. In facts, in SoSE one of the crucial word is "Engineering": *why we need Engineering in SoS? Because these are very Engineering Intensive Systems... and their Combination is very complex... requiring a lot of Ingenium to illuminate us on understanding and designing the SoS (Bruzzone & Maglione 2016).* Due to the above mentioned reasons, it is evident the importance to develop capabilities in understanding and applying SoSE. The authors decided that in order to succeed in this direction it is necessary to create virtual frameworks devote to provide a direct SoSE experience to trainees. It should be outlined that the training audience for these techniques is pretty articulated including young engineers as well as project engineers, program managers, company executives, public

authority managers as well as other stakeholders (Ncube 2011). Some of these individuals have very limited time and in any case the complexity of a real case related to SoSE could make almost impossible to experience it within a class, even if virtual, due to the number of details to be acquired before to get the whole picture.

In facts, Modeling and Simulation represents a great opportunity for SoS Engineering considering that:

- Physical experiments are typically infeasible in SoSE
  - o Computer simulation is required to reproduce this context
  - o Computer Simulation are expected to be quite computationally intensive and time consuming to address SoSE
  - o Verification and Validation is challenging due to the high number of objects and variables
- SoS are complex
  - o Special models are required to address each element
  - o Many subsystems and variables to be considered also by meta-models
- SoS have a broad and articulated configuration space
  - o Very large number of alternative configurations
  - o Need to speed up simulation for extensive experimentation and data farming
  - o Results could be hard to be understood, visualized and shared among stakeholders
- SoS have very high stochastic components
  - o For a given set of inputs, it is required to define the uncertainty and their expected distribution in real operations to be simulated
  - o It is required to adopt ANOVA and confidence band analysis for determine the output data distributions
- SoS include Intelligence as element for interoperability among different systems
  - o It is required to include behavioral models
  - o Behavior models should be modular to be able to combined under different conditions

In facts, M&S should be adopted taking into considerations these challenges and using consolidated methodologies (Amico et al. 2000; Mongomery 2000; Mittal et al. 2008). The state of the art review clearly reveals the potentials of M&S (in different domains) to come up with solutions able to take into account many of the issues mentioned above; from this point of view different review articles and specific applications can be found in Harvey and Stanton (2014), Davis et al., (2016); Longo et al. (2015).

Due to these reasons, the authors decided to develop an ad hoc scenario that provides a real challenge in terms of SoS and that is suitable for applying SoSE in short time, by an interactive and intuitive simulation environments to play with. The authors propose a strategic, dynamic stochastic simulation based serious game to be used during classes enabling fast time and distributed simulation; indeed by this approach it becomes possible to train people in different sites and

providing them interactive experience with other players over the web. The Serious Game could be designed to operate as web service in order to be usable in physical and virtual distributed classes from pc, laptop or even smartphone (Keegam 2005; Bruzzone et al.2014a).
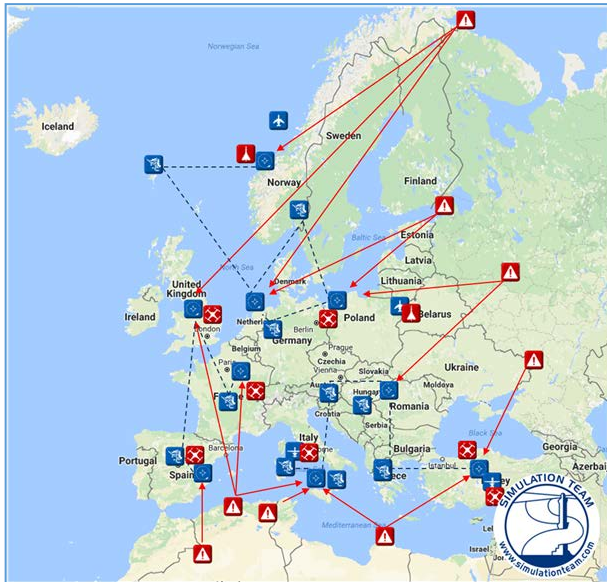

Figure 2 - MISCHIEF Scenario, Objects & Threats

The use of MS2G approach provide additional benefits by supporting interoperability and distributed simulation, so in this case the proposed solution adopt this paradigm (Bruzzone et al., 2014b) and it is named MISCHIEF (Multiple Interoperable Systems for joint Control of Hybrid threats through Intelligent Extended Fusion). The distributed use of MISCHIEF is guaranteed by the general client server architecture summarized in figure 1. Indeed MISCHIEF has been successfully experimented with industrial and academic users and deals with defense and homeland security respect air threats (Bruzzone and Maglione, 2016).

4 STRATEGIC SERIOUS GAMING

The proposed context deals usually with distributed operations respect a very broad spectrum of solicitations; so it is required that the players of the Serious Game should develop solutions with high level of interoperability ready to face all the potential challenges. Obviously another big challenge is represented by the uncertainty about the new different system real performance as well as on the real detailed characteristics of the scenario to be investigated. In facts in these very complex problems usually the real mission environment is not really well know and the design of new solutions relies strongly on hypotheses and assumptions; this image of the real mission environment, as it is supposed by developers, matches with "Simuland" concept proposed by McLeod. Indeed it is fundamental, while playing with the SoSE scenario, to check the consistency of the initial assumptions and

hypothesis respect the emerging new information and knowledge (McLeod 1968; Amico et al.2000).

Obviously the SoSE should be applied by players keeping clear in mind the multi-target goals addressing operational performance, reliability and cost effectiveness (Bruzzone et al.2006). In this way the trainee, by acting as players in the MISCHIEF Serious Game, experiences interactively that fundamentals decisions, strongly affecting costs, are taken during the Early Phase of Project and learn how to proceed properly in system design during architecture definition. MISCHIEF goal is not only training, but also educating users on Capability Oriented Design, evaluating both the acquisition of additional capabilities as well as in an effective interoperability enhancing the overall capabilities rather than the single system performance, as suggested by SoSE in order to be able to address the diverse challenge (Bruzzone & Maglione 2016).

5 MISCHIEF CONTEXT BACKGROUND SCENARIO

The MISCHIEF scenario is related to defense and homeland security with special attention to aerial threats (see figure 2), so it address a complex SoS that involves ground installations, electronic systems, power engineering, multiple platforms, operational modes, etc. The system is inspired by a real context addressing air space defense based on airborne radars; indeed from the end of the first half of XX century radars have been widely used in aerospace control in both civil and military fields. Depending on purpose they can be installed on fixed basement or mobile carrier such as truck, ship or plane. For example in civil air-traffic control typically used ground based radars are installed over tall infrastructures and towers (Nolan 2010).

However, the ground based radars have several limitations caused by curvature of the Earth and obstacles, such as trees, wind turbines and/or heels, create radar shadows hence limiting their capability to detect targets. This constraint was known since the beginning of World War II and was widely used by German aviation to reduce efficiency of British early warning radar system called Chain Home (CH) approaching to the coastline at low altitude (Brown, 1999). The only efficient solutions was installing as many radars as high as possible from the ground to reduce the radar's dead zones. Indeed thanks to the advances in electronics, soon new solutions were found installing the radar directly over special airplane devoted to carry out such task as in the case of AWACS (Airborne Warning and Control Systems) implemented over a large number of fixed wing airplanes (e.g. A-50, E-1, E-2C, E-3 Sentry, KJ-500, KJ-2000, PB-1W, , Tu-126 Moss) and also on helicopters (EH-101A, KA-31, SH-3H, Sea King AEW) and blimps (Good Year ZPG-2, ZPG-3).

In this sense, a very efficient solution is the installation of the sensors on flying platforms and today the autonomous systems are providing additional

opportunities in this field to add to planes also unmanned aerial vehicle (UAV) or unmanned LTAV (Lighter Than Air Vehicle). Nowadays it's impossible to imagine zone of military conflict or even movement of modern military forces without UAV, Airborne Warning and Control System (AWACS) or another reliable surveillance instrument with high mobility (Dorn, 2014). Even more, in many cases small targets, such as UAV themselves, flying at low altitude, are almost undetectable from the ground and the only reliable solution is based on airborne radars.

Obviously such systems are also used in civil fields, for example ground-penetrating radars (GPR) installed on UAV or helicopters are used in cartography of floods (National Research Council, 2004).

For any of mentioned above application simulation could be effectively used to support decision making as well as for training. For example, nowadays in air-traffic control simulators are used to support flow management (Tumer and Agogino, 2007; Shah et al. 2005), risk assessment (Stroeve et al., 2009) and even estimate airspace capacity (Majumdar and Polak, 2001).

In general, one of main tasks of defense systems is to detect and prevent airspace violation by civil and military flying objects; it should be said that in many cases, military flying objects are airborne surveillance systems performing gathering of information including capabilities of defense systems.

As mentioned above airborne surveillance systems are key components of defense and information support, hence are widely used in zones of military conflicts or during movement of military forces such as carrier strike groups (CSG), therefore current scenarios involving asymmetric threats and hybrid warfare introduce many more insidious menaces (e.g. drones) that require innovative defensive solutions.

However, airborne surveillance systems have some limitations, for example in terms of autonomy and patrolling time due to engine consumption and maximum carried fuel amount.

This means that engineering solutions should be strongly connected with the policies for using these assets and with efficient planning of operations; this represent a perfect example of System of Systems and the development of a breakthrough solution by introducing new enabling technologies such as UAV and other platforms and sensors represent a great case for a case study on SoSE. Due to these reasons the proposal for a new hypothetical airborne surveillance systems is the basis of the MISCHIEF serious game; in facts, as anticipated, the aim of this research is to develop an innovative tool for education and training to be used for experimenting it within SoSE and to explore the capabilities of new simulation solutions to find best configuration and support proper strategic decision making in this field.

6   MISCHIEF SIMULATION SOLUTION

Simulator contains several databases: targets (e.g. type, maximum speed and range etc), sensors (e.g. max range, mass, power consumption, MTTR, MTBF etc), generators (e.g. power, space, weight), mobile platforms on which sensors are installed (e.g. max duration of flight, max mass of load, costs etc.), ground installations at which platforms are assigned to (e.g. coordinates, capabilities), areas of origin of targets (e.g. coordinates, probability of departure, number of false alarms, etc.).

Table I– Example of Game Data Set on Radar Equipment

Radar Name	Range Max [km]	Acq.Cost [MEuro/year]	Op.Cost [MEuro/year]	Resolution [m2]	Detection Reliability	MTBF [h]	MTTR [h]	Weight [tons]	Volume [m3]
RDK1	500	100	25	1.00	95.0%	592.00	48.00	48.000	30.000
RDK2	100	10	2	1.00	94.0%	552.00	48.00	48.000	20.000
RDK3	100	10	3	1.00	93.0%	516.71	48.00	28.000	1.000
RDK4	50	2	0.5	1.00	94.0%	276.00	24.00	1.000	0.500
RDK5	120	25	7.5	0.25	98.0%	180.57	48.00	16.000	0.500
RDK6	50	16	4	0.50	96.0%	352.00	48.00	22.000	0.500
RDK7	100	14	2.8	10.00	80.0%	204.57	24.00	6.000	0.800
RDK8	40	4	1.6	5.00	88.0%	318.86	24.00	1.000	0.200
RDK9	35	2	0.4	0.20	97.0%	133.13	48.00	0.500	0.250
RDK10	60	0.5	0.125	7.00	84.0%	251.41	24.00	0.500	0.250

It is important to outline the need to develop an easy playable Serious Game: both GUI and scenario generators have to be designed to minimize the efforts of the users and to automate as much as possible the operations: MISCHIEF self generates the scenario based on reference data stochastically changing in terms of threats, system performance, etc. In this way multiple games could be easily set up in a quick and effective way. Another important aspect is that even just the simple definition of the disposal of assets and their initial conditions to finalize a configuration could be quite explosive in terms of decisions to be taken, resulting time consuming and characterized by low added value for trainees in terms of E&T in SoSE. Due to these reasons at the beginning of the simulation, after players have chosen sensors, platforms and other main systems, as well as their initial locations, the IA (Intelligent Agents) driving the different assets are self organizing their planning in a basic yet consistent way, in order to make the scenario immediately playable. The Sensors and platform should be chosen considering their cost, volume, mass, power consumption and other parameters, and the fine tuning of the engineering solutions is easily defined by players through MISCHIEF GUI. The Serious Game includes preliminary checks about the consistency of the configuration, in order to ensure a feasible combinations of sensors, generators, platforms and ground stations. For example if a platform have a payload exceeding its capabilities in terms of weight or volume, the Serious Game downsizes this configuration dismissing some equipment and consequently downsizing the system capabilities. This actions generate alerts that enable the trainees to identify and correct rising issues . When all mentioned above criteria are satisfied, simulation is initiated and experimental results are collected to further improve the SoSE. In facts, as soon as a configuration is tested by the simulation, more correct estimations respect the *a priori* performance value are provided to the player. In this way they have the possibility to review and adapt the engineering in order

to fill up the gaps. Each change made by the players is characterized by a project development cost that increase along the project timeline and it is computed as a mix of fixed elements and comparative difference respect previous configuration, in order to provide the feeling of the experimentation and prototyping costs. Targets and threats are placed in predefined zone and IA drive them based on behavioral algorithms; also in this case new sources and new capabilities and penetration strategies could be introduced along the game to check the resilience of the proposed SoSE and the capabilities of the players to adapt the solution to the emerging new challenges.

As anticipate MISCHIEF is implemented as a web service, with the simulator running on the Server even to avoid possibility for trainee to cheat on their client side (see figure 2). MISCHIEF is based on a connection the specific game Database generated to create the scenario at the beginning of the Game based on PHP and related server-side scripting, while it is used Apache Web Server to manage the client/server communication through HTTP (Hypertext Transfer Protocol) and/or HTTPS (HyperText Transfer Protocol over Secure Socket Layer),.

The architecture is based on a MISCHIEF client implemented in JavaScript using jQuery additional libraries executed within the web browsers and HTM HTML and CSS for GUI definition; this approach allows to operate as platform independent application and to be usable even from smartphones for mobile training. The simulation provide also dynamic graphics of the scenario evolution to simplify the problem understanding by using web-socket server; indeed this solution overpass the limitation of most browsers allowing to guarantee continuous update of the tactical situations and of the overall strategic variable. The MISCHIEF simulation engine is implemented by using Python, Autobahn and Twisted networking engine.

This solution have a big potential being able to support communications mostly in real-time and in the future it could be possible to proceed with further developments for operational use instead than just E&T.

MISCHIEF target function are measured through analysis of simulation results for the whole SoS:

$$Tre = t_x / \frac{1}{2}\left[1 + \frac{2}{\sqrt{\pi}}\int_0^{\frac{t_x - mtre}{\sqrt{2}\cdot stre}} e^{\frac{-t}{2}} dt\right] \leq \lambda$$

$$mtre = \sum_{i=1}^{m} \frac{t_i}{i}$$

$$stre = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(t_i - mtre)^2}$$

Tre   Responsiveness
t_x Reference time required to detect, classify and engage a target in the simulation
mtre   Mean time to classify and engage targets
stre   Standard deviation to classify and engage targets
m   Number of generated targets
λ Threshold level for responsiveness (e.g. 80%)

The MISCHIEF Game could be played in standalone mode as well as in teams cooperating to find a best solution while they compete with other teams.

The authors are currently using MISCHIEF with different classes achieving very interesting results and being able to validate and verify the game rules and logic.

7 CONCLUSION

This paper present an early stage development of a new generation solution for addressing the challenging topic of Education and Training for engineers and decision makers engaged in SoSE. In facts, the SoSE is a pretty challenging environment and the preliminary results achieved through experiments with students as well as during professional courses for industry are pretty promising. Currently the authors are working to organize distributed exercises mixing different classes in order to evaluate the specific characteristics of players with different technical and cultural background respect the SoSE concepts.

REFERENCES

Amico Vince, Bruzzone A.G., Guha R. (2000) "Critical Issues in Simulation", Proceedings of Summer Computer Simulation Conference, Vancouver, July

Brown, L. (1999). "A Radar History of World War II". Institute of Physics Publishing, Bristol

Bruzzone A.G., E. Bocca (2008) "Introducing Pooling by using Artificial Intelligence supported by Simulation", Proc.of SCSC2008, Edinbourgh, UK

Bruzzone A.G., Maglione G.L. (2016) "Complex Systems & Engineering Approaches", Simulation Team Technical Report, Genoa

Bruzzone, A. G. et al. (2006) "Simulation and Optimization as Decision Support System in Relation to Life Cycle Cost of New Aircraft Carriers", Proceedings of Modelling Simulation and Optimization, Gaborone, Botswana.

Bruzzone, A. G., Massei, M., Tremori, A., Longo, F., Nicoletti, L., Poggi, S., Bartolucci C., Picco E. & Poggio, G. (2014b) "MS2G: simulation as a service for data mining and crowd sourcing in vulnerability Reduction" Proc. of WAMS, Istanbul, September.

Bruzzone, A. G., Massei, M., Tremori, A., Poggi, S., Nicoletti, L., & Baisini, C. (2014a) "Simulation as enabling technologies for agile thinking: training and education aids for decision makers" International Journal of Simulation and Process Modelling 9, 9(1-2), 113-127

Davis M., Proctor M., Shageer B., (2016). A Systems-Of-Systems Conceptual Model and Live Virtual

Constructive Simulation Framework for Improved Nuclear Disaster Emergency Preparedness, Response, and Mitigation. Journal of Homeland Security and Emergency Management, vol. 13, no. 3, pp. 367-394.

Dorn A.W. (2014) "Aerial Surveillance: Eyes in the Sky" in Air Power in UN Operations: Wings for Peace (A. Walter Dorn, Ed.), Ashgate Publishing, Farnham, UK, pp. 119-134

Giribone P., Bruzzone A.G. & Tenti M. (1996) "Local Area Service System (LASS): Simulation Based Power Plant Service Engineering & Management", Proc. of XIII Simulators International Conference, New Orleans LA, April 8-11

Harvey C., Stanton N.A., (2014). Safety in System-of-Systems: Ten key challenges, Safety Science, vol. 70, pp. 358-366.

Jamshidi, M. (2008) "Introduction to system of systems. System of Systems Engineering", CRC Press, NY, pp. 1-43

Jamshidi, M. (2011) "System of systems engineering", Innovations for the twenty-first century, vol 58, John Wiley & Sons, NYC

Keating, C. B., Sousa-Poza, A., & Kovacic, S. (2005) "Complex system transformation: a system of systems engineering (SoSE) perspective", Proc. of 26th ASEM National Conference, pp. 200-207

Keating, C., Rogers, R., Unal, R., Dryer, D., Sousa-Poza, A., Safford, R., Peterson W. & Rabadi, G. (2003) "System of systems engineering", Engineering Management Journal, 15(3), 36-45

Keegan, D. (2005) "The incorporation of mobile learning into mainstream education and training", Proc. of World Conference on Mobile Learning, Cape Town, October.

Longo F., Chiurco A., Musmanno R., Nicoletti L., (2015). Operative and procedural cooperative training in marine ports. Journal of Computational Science, vol. 10, pp. 97-107.

Majumdar, A., Polak, J. (2001). "Estimating capacity of Europe's airspace using a simulation model of air traffic controller workload". Journal of the Transportation Research Board, (1744), 30-43.

McLeod, J. (1968) "Simulation: the Dynamic Modeling of Ideas And Systems with Computers", McGraw-Hill, NYC

Mittal, S., Zeigler, B. P., Martín, J. L. R., Sahin, F., & Jamshidi, M. (2008) "Modeling and Simulation for Systems of Systems Engineering", in Systems of Systems Innovation for 21th Century, Wiley & Sons, NYC

Montgomery D.C. (2000) "Design and Analysis of Experiments", John Wiley & Sons, New York

National Research Council. (2004) "Assessing the national streamflow information program", National Academies Press, Washington DC, USA

Ncube, C. (2011) "On the Engineering of Systems of Systems: key challenges for the requirements engineering community", Proc. of IEEE Workshop on Requirements Engineering for Systems, Services and Systems-of-Systems (RESS), August, pp. 70-73

Nolan, M. (2010). "Fundamentals of air traffic control", Delmar, Boston, MA

Rhodes, D. H., Valerdi, R., & Roedler, G. J. (2009) "Systems engineering leading indicators for assessing program and technical effectiveness", Systems Engineering, 12(1), 21-35

Shah, A. P., et al (2005, June). "Analyzing air traffic management systems using agent based modeling and simulation" In Proceedings of the 6th USA/Europe Seminar on Air Traffic Management Research and Development.

Sousa-Poza A., Kovacic S., Keating C. (2008) "SoSE: An Emerging Multidiscipline", Int.Journal of Systems Engineering, Vol.1, Nos.1/2

Stroeve, S. H., Blom, H. A., Bakker, G. B. (2009). "Systemic accident risk assessment in air traffic by Monte Carlo simulation", Safety science, 47(2), 238-249.

Tumer, K., Agogino, A. (2007, May). "Distributed agent-based air traffic flow management" In Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems (p. 255). ACM.

Authors' Index