THE 18TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION

SEPTEMBER 18 – 20, 2019
LISBON, PORTUGAL

EDITED BY
AGOSTINO G. BRUZZONE
FABIO DE FELICE
LUÍS MIGUEL SILVA DIAS
MARINA MASSEI
ADRIANO SOLIS

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2019

ISBN 978-88-85741-30-0 (Paperback)
ISBN 978-88-85741-29-4 (PDF)

ISBN 978-88-85741-30-0 (Paperback)
ISBN 978-88-85741-29-4 (PDF)

THE 18TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION, MAS 2019

SEPTEMBER 18 – 20, 2019

LISBON, PORTUGAL

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

HUNGARIAN ACADEMY OF SCIENCES CENTRE FOR ENERGY RESEARCH

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

ISBN 978-88-85741-29-4; Bruzzone, De Felice, Dias, Massei and Solis, Eds

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

University of Minho

Universidade do
Minho

I3M 2019 Industrial Sponsors

CAL-TEK SRL

LIOTECH LTD

MAST SRL

SIM-4-FUTURE

I3M 2019 MEDIA PARTNERS

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SIMULATION AND PROCESS MODELING

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SERVICE AND COMPUTING ORIENTED MANUFACTURING

IGI GLOBAL – INTERNATIONAL JOURNAL OF PRIVACY AND HEALTH INFORMATION MANAGEMENT (IJPHIM)

HALLDALE MEDIA GROUP: THE 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

MDPI - SUSTAINABILITY

EUROMERCI: THE ITALIAN MONTHLY LOGISTICS JOURNAL

EDITORS

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

LUÍS MIGUEL SILVA DIAS
UNIVERSITY OF MINHO
lsd@dps.uminho.pt

FABIO DE FELICE
UNIVERSITY OF CASSINO, ITALY
defelice@unicas.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 2019

GENERAL CO-CHAIRS

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

PROGRAM CO-CHAIRS

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

THE 18TH INTERNATIONAL CONFERENCE ON MODELING AND APPLIED SIMULATION, MAS 2019

GENERAL CO-CHAIRS

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

PROGRAM CO-CHAIRS

FABIO DE FELICE, *UNIVERSITY OF CASSINO, ITALY*
LUÍS MIGUEL SILVA DIAS, *UNIVERSITY OF MINHO, PORTUGAL*

ISBN 978-88-85741-29-4; Bruzzone, De Felice, Dias, Massei and Solis, Eds

MAS 2019 INTERNATIONAL PROGRAM COMMITTEE

NAAMANE AZIZ, *LABORATOIRE DES SCIENCES DE L'INFORMATION ET DES SYSTÈMES, FRANCE*
EDGAR ALONSO LOPEZ-ROJAS, *NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY, NORWAY*
ELEONORA BOTTANI, *UNIVERSITY OF PARMA, ITALY*
YOUSSEF BOUANAN, *LAB GRETHA UNIVERSITÉ BORDEAUX*, FRANCE
AGOSTINO G. BRUZZONE, *UNIVERSITY OF GENOA, ITALY*
ALESSANDRO CHIURCO, *UNIVERSITY OF CALABRIA, ITALY*
GERSON CUNHA, *LAMCE COPPE UFRJ, BRAZIL*
FABIO DE FELICE, *UNIVERSITY OF CASSINO, ITALY*
DAVID DEL RIO VILAS, *FERROVIAL SERVICES, SPAIN*
MARCO FRASCIO, *UNIVERSITY OF GENOA, ITALY*
CLAUDIA FRYDMAN, *LSIS, FRANCE*
YILIN HUANG, *DELFT UNIVERSITY OF TECHNOLOGY, NETHERLANDS*
JANOS SEBESTYEN JANOSY, *CENTRE FOR ENERGY RESEARCH HUNGARIAN ACADEMY OF SCIENCES, HUNGARY*
ANDREAS KÖRNER, *TECHNISCHE UNIVERSITÄT WIEN, AUSTRIA*
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*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
ANTONIO PADOVANO, *UNIVERSITY OF CALABRIA, ITALY*
HORACIO EMILIO PÉREZ SÁNCHEZ, *UNIVERSIDAD DE MURCIA, SPAIN*
ANTONELLA PETRILLO, *UNIVERSITY OF CASSINO, ITALY*
MARIA CELIA SANTOS LOPES, *LAMCE/COPPE/UFRJ – BRASIL*
PEER OLAF SIEBERS, *UNIVERSITY OF NOTTINGHAM, UK GEORGIOS CH*
ADRIANO SOLIS, *YORK UNIVERSITY, CANADA*
GIUSEPPE VIGNALI, *UNIVERSITY OF PARMA, ITALY*
GREGORY ZACHAREWICZ, *IMS UNIVERSITÉ BORDEAUX 1, FRANCE*

TRACKS AND WORKSHOP CHAIRS

SIMULATION FOR STRATEGIC ENGINEERING
CO-CHAIRS: AGOSTINO G.BRUZZONE, SIMULATION TEAM GENOA, UNIVERSITY OF GENOA (ITALY); FRANCESCO LONGO, MSC-LES (ITALY); IAN MAZAL, BRNO UNIVERSITY OF DEFENCE (CZECH REPUBLIC); MARINA MASSEI, LIOPHANT SIMULATION (ITALY); ANNA FRANCA SCIOMACHEN, DIEC (ITALY)

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;

SIMULATION INTEROPERABILITY IN ENTERPRISES
CHAIRS: GREGORY ZACHAREWICZ, MS UNIVERSITÉ BORDEAUX 1 (FRANCE) ; 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 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 G.BRUZZONE, UNIVERSITY OF GENOA, ITALY

HUMAN-CENTRED AND HUMAN-FOCUSED MODELLING AND SIMULATION (COMMON TRACK MAS-EMSS)
CHAIRS: AGOSTINO G. 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 2019!

We are pleased to announce the 18th International Conference on Modelling & Applied Simulation in Lisbon, Portugal.
We are delighted to see that year by year, The International Conference on Modelling & Applied Simulation is growing and is becoming even more rich in terms of ideas, topics covered and above all in terms of achievements. It makes us feel privileged to be Co-Chairs of this important event.
As a matter of facts, over the last 18 editions, The International Conference on Modelling & Applied Simulation has collected high quality as well as innovative contributions in research and applications resulting in good impacts within and outside its community. Our vision is to keep working on the conference growth to create even greater value.
This year, the scientific program will cover a broad spectrum of topics that span disciplines in the area of Industry 4.0, Mixed Reality, Computer Science, Ergonomics, Internet of Things, Automation, Machine Learning, Modeling &Simulation, Autonomous Systems, Virtual Reality, and many others.
A special effort has been paid to cover current and future research directions in order to make the conference an opportunity of inspiration and exchange able to foster scientific innovation as well as lasting scholarly collaborations across the globe.
Needless to say that setting up a conference according to the vision and the premises expressed above is quite a hard task. To this end, we would like to express our gratitude to the authors for their contributions which are the foundation of this conference, to the program Committee members for their tireless efforts to organize all sessions and tracks, to the Local Organization Committee for the work in support of many conference activities and most of all, we would like to thank all the participants for enriching the conference with their presence.
We hope that you will find the conference and your stay in Lisbon both valuable and enjoyable.

Marina Massei
University of Genoa,
Italy

Adriano Solis
York University,
Canada

ACKNOWLEDGEMENTS

The MAS 2019 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The MAS 2019 IPC would like to thank all the authors as well as the reviewers for their invaluable work.
A special thank goes to all the organizations, institutions and societies that have supported and technically sponsored the event.

I3M 2019 INTERNAL STAFF

AGOSTINO G. BRUZZONE, *DIME, UNIVERSITY OF GENOA, ITALY*
ALESSANDRO CHIURCO, *CAL-TEK SRL, ITALY*
JESSICA FRANGELLA, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
CATERINA FUSTO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
LUCIA GAZZANEO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
MARINA MASSEI, *DIME, UNIVERSITY OF GENOA, ITALY*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
ANTONIO PADOVANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
CATALDO RUSSO, *CAL-TEK SRL, ITALY*
KIRILL SINELSHCHIKOV, *SIMULATION TEAM, ITALY*
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

TOWARDS THE SEMI-AUTOMATIC ADAPTATION OF SIMULATED VIRTUAL LABORATORY EXPERIMENTS

Athanasios Sypsas[(a)], Dimitris Kalles[(b)]

[(a)]PhD candidate, Hellenic Open University, School of Science and Technology, Greece
[(b)]Associate Professor, Hellenic Open University, School of Science and Technology, Greece

[(a)]sipsas@gmail.com
[(b)]kalles@eap.gr

ABSTRACT
Since real-world problems are complex a system model is usually required in advance to be built for such a problem to be properly investigated. Virtual laboratories constitute a special category of simulations and are based on models of physical laboratories and the experimental processes carried out therein. Similar experiments can be adapted to suit various learners' needs if they can be transformed to satisfy the expected learning outcomes for each audience. We compare such experimental procedures using the Activity Diagrams which correspond to these experiments, in order to detect differences between them. These differences are, then, used for the required transformation of the experimental steps. The algorithm implemented uses a uses a BFS-like traversal to detect the differences between Activity Diagrams. The evaluation of the distance between the Activity Diagrams is carried out by the user and the possible needed transformation is decided to meet the learning outcomes in the educational environment selected by the user, educator or learner.

Keywords: simulation, virtual laboratory, activity diagrams

1. INTRODUCTION

Quite often, to investigate and understand how a real-life system works, a model of the system at an abstract level is constructed. A model contains mathematical or logical relationships. If relationships are simple, mathematical methods (algebra, calculus) may be used to acquire accurate information concerning the system's behavior via an analytic solution (Law and Kelton 2000). However, real-world problems are complex and analytic solution cannot provide the required information to thoroughly study a real-world system. These models are usually studied using simulation. There are quite a few alternative definitions for simulation. According to Robinson (2004), simulation is the imitation of a system, while according to Maria (1997), "simulation of a system is the operation of a model of the system" and, as stated in Banks et al. (2010), "simulation is the imitation of the operation of a real-world process or system over time". Generally, simulation is a mechanism to evaluate over time the behavior of a working, or under construction, system, under changing conditions of operation Simulation is a tool to understand artificial or natural systems and explain their performance (Ramat and Preux 2003), whereas a computer simulation is "a program that contains a model of a system or a process" (De Jong and Van Joolingen 1998). Furthermore, simulation is used to improve current systems or operations, offering a better understanding of a system and potentially helping identify suggested improvements (Robinson, Nance, Paul, Pidd, and Taylor 2004). A simulation model is a mathematical model used to evaluate the outputs of a system given certain inputs. It is used to analyze and study complex, which that cannot be studied using simple mathematical operations and methods. A simulation model consists of input variables, system entities, performance measures and functional relationships (Maria 1997). A state is an ordered tuple of values which completely describe the system's entities (nodes), and which are used to depict - via state diagrams - the transitions among system entities. The state transitions are made following defined sets of rules. Given a set of inputs and model characteristics, the model is run and the simulated behavior and performance are observed.

In the field of education, virtual laboratories constitute a special category of simulations and are based on models of physical laboratories and the experimental processes taking place therein (Rossiter 2016). Technology-enhanced simulation offers educators the ability to form an attractive and interactive learning environment so that learners become active participants in the educational process (Cook et al. 2013). There are virtual laboratories, which can be used by educators in order to create experiments for students in a variety of educational institutions, including high schools and universities. However, a simulator for a specific experimental procedure and specific instruments cannot, in principle, be used without changes in a similar experimental environment, since the availability of the instruments also affects the experimental procedure, as different instruments might require certain changes in

the experimental steps. Consequently, in order to map the simulated experiment to equipment with fewer capabilities, as might be the case with microscopes available in a high-school laboratory, and still achieve the educational goals of the experiment, it will be essential to adjust the experimental process. Due to the various experimental applications that are addressed to different audiences, models for virtual laboratories continuously evolve in order to meet the educational needs formed at the educational institutions.

In order to use similar experiments so as to be addressed to various learners audience, a formal description of simulation experiments is needed (Peng, Warnke, Haack and Uhrmacher 2016). Various Domain Specific Languages have been developed to describe different aspects of simulation experiment (Schützel, Peng, Uhrmacher and Perrone 2014) as each of these languages is specialized to particular domain problem. Description languages diagrams and visual modeling environments and tools (i.e. UML, Rational rose) are used in order to describe abstractly the procedures-steps of the experimental process and the data objects needed (Hucka et al. 2015). Then UML diagrams describing virtual experiments can be mapped to XML or any concrete implementation language, like C or Java, so that can be embedded in any virtual laboratory environment. Rules for simulation experiments apply, including precise description of the simulation steps and other procedures, so that can be reused in different simulation environments (Waltemath et al. 2011).

2. EXAMPLE

As a typical use case of virtual laboratory experiment transformation, we use the steps required for the microscoping procedure, aimed for university level students and secondary education students. Given a description of the experimental steps and the original microscope description, a transformation is needed so as to execute the experiment using another, similar but less sophisticated, microscope, to achieve a subset of the initial learning outcomes, targeted at secondary school level audience. Specifically, in a biology experiment in Hellenic Open University (HOU), which features the microscopic observation of plant cells, the experimental procedure is specified so as to achieve a high level of realism in the simulation, using an appropriate simulator (http://onlabs.eap.gr), which is depicted in Fig. 1, below

Figure 1: Onlabs virtual laboratory

The same biology experiment is used in secondary education schools. However, many schools are not equipped with the required laboratories including the appropriate tools, equipment and required instructional media (Ejiwale 2013). Moreover, in secondary education, only a small number of countries innovated in this domain and the absolute change in the access to science laboratories amounted to just 3 percentage points (Vincent-Lancrin, Urgel, Kar and Jacotin 2019). This small percentage reveals the limited investment in the development of new science laboratory facilities in secondary education and strengthens the argument for investing in the virtual laboratories as an alternative educational tool, when physical science labs are not easily accessible.

Firstly, the simulation educational environment is selected by the user, educator and/or learner. Then, we use as comparison parameters the learning outcomes (O) (to study cells, using the various parts and operation of a microscope), the description of the microscope (I) (including the different parts) and the experimental procedure (E) for the specific experiment. So, the proximity (δ) or "distance calculation" between the different experimental processes is the key element for experiment transformation/ mapping.

Thus, the comparison of the experimental procedures lies on the comparison between the diagrams used for experiments description, since semantic differencing presents differences as elements in the semantics of the one model but not the other (Maoz and Ringert 2018). Such elements are the traces of action execution in an Activity Diagram (AD).

UML 2 Activity Diagrams (ADs) are used for experimental procedure description in virtual laboratory environment. For example, the AD shown in Fig. 2 describes the initial steps of the experimental procedure of microscoping at the HOU level. The AD shown in Fig. 3 describes the corresponding steps of the experimental procedure at the secondary school level.

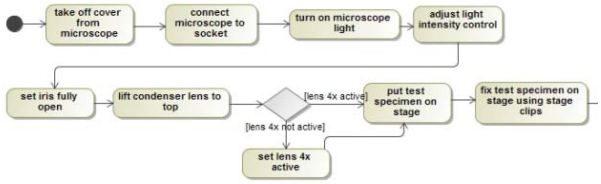


Figure 2: Steps of the Experimental Procedure in HOU

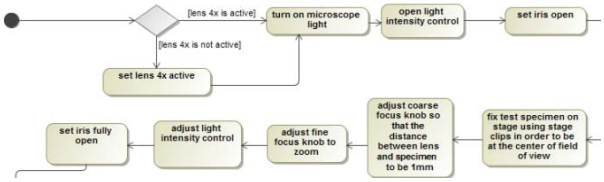


Figure 3: Steps of the Experimental Procedure in Secondary School

The above example is simple and thus easy to spot the differences when looking at the ADs. We use it in order to show the basic implementation of our idea.

3. BACKGROUND

An AD can be described as Maoz, Ringert, and Rumpe (2011) proposed, $AD = (A, V^{inp}, V^{loc}, AN, PN, T)$, where

- A is a set of action names.
- V^{inp} is a set of input variables over finite domains.
- V^{loc} is a set of local variables over finite domains.
- AN is a set of action nodes $an_{1,\dots,}$ with name $acname(an) = ac \in A$
- PN is a set of pseudo nodes, like initial nodes PN^{Init}, final nodes PN^{Fin}, decision nodes PN^{dec}, etc.
- T is a set of transitions of the form $t = (n_{src}, n_{trg}, guard)$ where $n_{src}, n_{trg} \in (AN \cup PN)$ and $guard$ is a Boolean expression.

The traces (tr) inside the AD are the sequences of AD states (ad), as described above:
$tr = s_0, s_1, \dots, s_k$ of ad with $s_{i+1} \in successors(s_i)$ and $s_0 \in ad.PN^{Init}$, where $successors(s_i)$ is the next AD state reachable from s_{i-1}.
Thus, in order to investigate the AD differences they are compared and diff traces are defined as following (Maoz et al. 2011):
Given ADs ad_1 and AD_2 a diff trace is a sequence of states $tr_1 = s_1^0, s_1^1, \dots, s_1^k$
so that

- $tr_1 \in traces(ad_1)$
- $\exists tr_2 = s_2^0, s_2^1, \dots, s_2^k$ *so that* $tr_2 \in traces(ad_2)$
  $\wedge \forall i, 0 \le i \le k, s_1^i \sim s_2^i,$
  $\wedge \exists s_2^{k+1}$ *so that* $s_2^{k+1} \not\sim s_1^{k+1}$
  $\wedge s_2^0, s_2^1, \dots, s_2^k \in traces(ad_2)$

tr_2 is called a corresponding diff trace of tr_1.

Two states s_1 and s_2 are corresponding, denoted $s_1 \sim s_2$ iff

- $s_1.ac = s_2.ac;$
- $\forall v \in V_1^{inp} \cap V_2^{inp}, s_1.val(v) = s_2.val(v).$

Based on these definitions different algorithms have been developed to calculate the diff traces (Maoz, Ringert and Rumpe 2011).
The implementation presented in the next section takes under consideration both action nodes and pseudo nodes, such as initial, final, fork, join and decision.

4. AN ALGORITHM TO COMPUTE THE DISTANCE BETWEEN ACTIVITY DIAGRAMS

As stated, there are algorithms in order to compute the differences between ADs. We have chosen to use the algorithm named ADDiff developed by Maoz, Ringert and Rumpe (2011) with slight changes. The specific algorithm uses a queue for corresponding states-pairs that have been reached but whose successors have not yet been traversed. Also, the visited states are stored using a Boolean visited array, in order to avoid processing a state more than once, as the algorithm uses a BFS-like traversal. The pseudo-code for the algorithm is given in Proc. 1, below.

Procedure 1: Pseudo-code to Detect Differences

Procedure 1 differencesinad(*ad₁*, *ad₂*)
Def *queueofpairs* **as queue of** *Pair*
Def *visitedpairs* **as list of** *Pair*
Def *rejectedpairs* **as list of** *Pair*
Def adtraces **as list of lists of** *Pair*
For all *ini₁* **in** *ad₁.inistates* **do**
  CorrespondingFound←false
  For all *ini₂* in *ad₂.inistates* **do**
    If *corresponding*(*ini₁*, *ini₂*) = true **then**
      Add *Pair*(*0*, *ini₁*, *0*, *ini₂*) **to** *queueofpairs*
      Add *Pair*(*0*, *ini₁*, *0*, *ini₂*) **to** *visitedpairs*
      CorrespondingFound←true
    Endif
  Endfor
  If *CorrespondingFound* = false **then**
    Add *Pair*(*0*, *ini₁*, *0*, *0*) **to** *rejectedpairs*
  Endif
Endfor
visitedpairs←*traverse*(*ad₁*, *ad₂*)
rejectedpairs←*traverse*(*ad₁*, *ad₂*)
adtraces←trace(*visitedpairs*, *rejectedpairs*)
return *adtraces*

The main structure Pair contains two pairs of states: predecessor and current state in ad_1 (pre_1, cur_1) and in ad_2 (pre_2, cur_2). The algorithm checks if the initial states of ad_1 exist in ad_2. If the corresponding states exist the pair is stored in structures *queueofpairs* and *visitedpairs*. In case no corresponding state exists a pair only for ad_1 state is stored in structure *rejectedpairs*. Structure *adtraces* is used to store the diff traces which will be calculated.
Procedure *traverse* iterates on the queue until it is empty, searching for the successors of the states in ad_1

and then for the corresponding states of ad_2. Procedure *trace* builds the traces from the initial states to the rejected states.

We have not yet fully implemented the evaluation of the distance between the ADs but this is work in progress using the API of the MagicDraw tool for AD design.

The differences between the two ADs for the same experiment are presented using dotted frames, in Fig. 4 and Fig. 5 below.

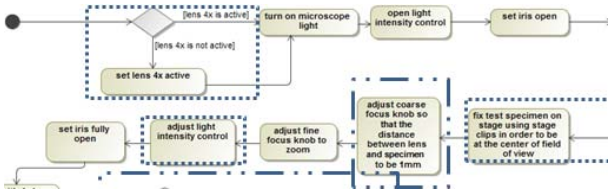


Figure 4: Experimental Procedure in Secondary School with Differences shown in a Dotted frame

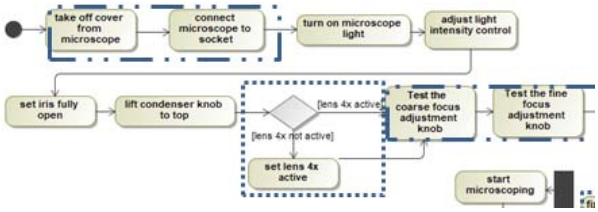


Figure 5: Experimental Procedure in HOU with Differences shown in a Dotted frame

Some of the steps are only in the one AD and some steps are in different order. For example, the decision node of whether lens 4x is active in the microscope is in different order. Thus, when algorithm is running the above differences are spotted, as they are when ADs are compared. However, the learning outcome-skill (manipulate microscope lenses) is achieved for the specific part of the experimental procedure. As a consequence, the AD is divided into subsections according to the learning outcome-skill to be achieved. Finally, these subsections are compared in order to reveal the differences which may lead to possible differences in learning outcomes-skills, as it is shown in Fig. 6 and Fig. 7.

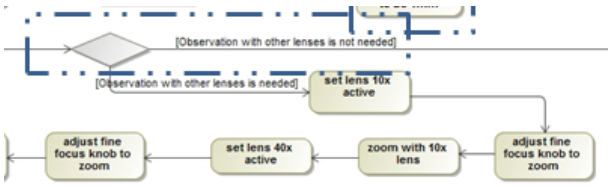


Figure 6: Steps of Experimental Procedure in Secondary Education level with Differences shown in a Dotted frame

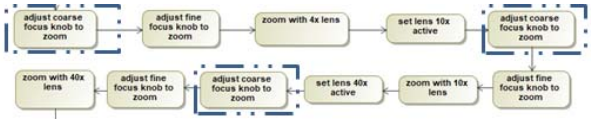


Figure 7: Steps of Experimental Procedure in HOU with Differences shown in a Dotted frame

The differences in the specific steps are spotted because, in the secondary education context, the use of lenses besides 4x ones cannot be supported, as the microscope does not have this type of lenses. So, it is decided that the differences affect the achievement of specific learning outcomes ("change lenses and use them"), and the educator may decide whether the experimental steps need to be transformed/changed to meet the learning outcomes (in that case, the decision node is deleted and the steps are mapped to the steps of the HOU experiment).

5. DISCUSSION

The idea of simulation experiments reuse is not new. Peng, Warnke, Haack and Uhrmacher (2016) have developed a mechanism to automatically generate and execute simulation experiments for extended models based on the reuse of the original experiment specifications. To facilitate the reuse of experiments, their robust and detailed description is needed. The aforementioned approach uses a declarative domain specific language SESSL to specify the experiments.

Various tools have also been developed in order to support the specification of experiments, as it is the first step for reusing simulation experiments (Hillston 1995; Smith, Llado and Puigjaner 2011). Our research focuses on detecting the differences between specified experiments that are depicted via ADs, in order to transform them to meet the learning outcomes in the educational environment selected by the user, educator or learner.

6. CONCLUSION AND FUTURE WORK

We presented a general framework for detecting the differences between experiments for virtual laboratories. These differences then are used for the needed experimental steps transformation in order to be executed by the selected simulator. We formalized the framework and presented the ADs that are primarily used to validate it.

We plan to build a repository containing ADs for experimental procedures for various educational settings, so that distance evaluation between the ADs will lead to simulator selection and experimental procedure transformation based on the learning outcomes set by the experiment designers.

ACKNOWLEDGMENTS

This research is co-financed by the European Union (European Social Fund - ESF) and Greek National Funds through the Operational Programs:

a) "The Human Resources Development, Education and Lifelong Learning Program (Digi-Lab, 70381) and

b) "The Financial Support for Development and Innovation Research Projects in RIS3 Priority Area - Information Technologies and Communications (XRLabs, 80014).

REFERENCES

Banks, J., Nelson, B. L., Carson, J. S. and Nicol, D. M., 2010. Discrete-Event System Simulation. PrenticeHall International Series in Industrial and

Systems Engineering, 640. https://doi.org/10.2307/1268124

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., … Hatala, R., 2013. Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. Medical Teacher, 35.

De Jong, T. and Van Joolingen, W. R., 1998. Scientific Discovery Learning with Computer Simulations of Conceptual Domains. Review of Educational Research, 68(2), 179–201. https://doi.org/10.3102/00346543068002179

Ejiwale, J. A., 2013. Barriers To Successful Implementation of STEM Education. Journal of Education and Learning (EduLearn), 7(2), 63–74. https://doi.org/10.11591/edulearn.v7i2.220

Hillston, J., 1995. A tool to enhance model exploitation. Performance Evaluation, 22(1), 59–74. https://doi.org/10.1016/0166-5316(93)E0038-7

Hucka, M., Bergmann, F. T., Hoops, S., Keating, S. M., Sahle, S., Schaff, J. C., … Wilkinson, D. J., 2015. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. Journal of Integrative Bioinformatics (JIB), 12(2), 382–549. https://doi.org/10.2390/biecoll-jib-2015-266

Law, A. M. and Kelton, W. D., 2000. Simulation Modeling and Analysis. New York: McGraw-Hill

Maoz, S. and Ringert, J. O., 2018. A framework for relating syntactic and semantic model differences. Software and Systems Modeling, 17(3), 753–777. https://doi.org/10.1007/s10270-016-0552-y

Maoz, S., Ringert, J. O. and Rumpe, B., 2011. A manifesto for semantic model differencing. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6627 LNCS, 194–203. https://doi.org/10.1007/978-3-642-21210-9_19

Maria, A., 1997. Introduction to modelling and simulation. Winter Simulation Conference, 7–13. https://doi.org/10.1109/WSC.1997.640371

Peng, D., Warnke, T., Haack, F. and Uhrmacher, A. M., 2016. Reusing simulation experiment specifications to support developing models by successive extension. Simulation Modelling Practice and Theory, 68, 33–53. https://doi.org/10.1016/j.simpat.2016.07.006

Ramat, E. and Preux, P., 2003. "Virtual laboratory environment" (VLE): A software environment oriented agent and object for modeling and simulation of complex systems. In Simulation Modelling Practice and Theory, 11, 45–55. https://doi.org/10.1016/S1569-190X(02)00094-1

Robinson, S., 2004. Simulation: the practice of model development and use. Journal of Simulation. https://doi.org/10.1057/palgrave.jos.4250031

Robinson, S., Nance, R. E., Paul, R. J., Pidd, M. and Taylor, S. J. E., 2004. Simulation model reuse: Definitions, benefits and obstacles. Simulation Modelling Practice and Theory, 12(7-8 SPEC. ISS.), 479–494.

Rossiter, J. A., 2016. Low production cost virtual modelling and control laboratories for chemical engineering students. IFAC-PapersOnLine, 49(6), 230-235.

Schützel, J., Peng, D., Uhrmacher, A. M. and Perrone, F., 2014. Perspectives on languages for specifying simulation experiments. In Proceedings of the Winter Simulation Conference, pp. 2836–2847. December 7-10, Savannah, (Georgia, USA).

Smith, C. U., Llado, C. M. and Puigjaner, R., 2011. Model Interchange Format Specifications for Experiments, Output and Results. The Computer Journal, 54(5), 674–690. https://doi.org/10.1093/comjnl/bxq065

Sypsas, A. and Kalles, D., 2018. Virtual laboratories in biology, biotechnology and chemistry education: a literature review. Proceedings of the 22nd Pan-Hellenic Conference on Informatics pp. 70-75. 30 November 30-December 1, Athens, (Greece).

Vincent-Lancrin, S., Urgel, J., Kar, S. and Jacotin, G., 2019. Measuring Innovation in Education 2019. OECD. Available from: https://doi.org/10.1787/9789264311671-en [accessed 15 March 2019]

Waltemath, D., Adams, R., Bergmann, F. T., Hucka, M., Kolpakov, F., Miller, A. K., … Le Novère, N,. 2011. Reproducible computational biology experiments with SED-ML - The Simulation Experiment Description Markup Language. BMC Systems Biology, 5(1), 198. https://doi.org/10.1186/1752-0509-5-198

APPRAISAL OF OFDM FOR MULTI-CARRIER, HIGH-SPEED DATA RATE WIRELESS COMMUNICATION NETWORKS

Engr. Muhammad Bello Abdullahi, Member NSE[1]

Department of Electrical and Electronic Engineering,
Abdu Gusau Polytechnic, Talata Mafara,
Zamfara State, Nigeria

engr.mbello@yahoo.co.uk

Abstract:
Orthogonal Frequency Division Multiplexing (OFDM) is used to achieve multi-carrier signals and high- Speed data rate in free space. OFDM-based systems operate in the hostile multipath radio environment, which allows efficient sharing of limited resources. This research work was designed, developed and simulated an OFDM System using the basic blocks of Simulink in MATLAB/Simulink software, to support multi-carrier, high-speed data rates. This was achieved in backing of collection and review of high-quality research papers, which reported the latest research developments in OFDM communications networks, and its applications in future wireless systems.

The research work significantly increases the speed of data rate signals, and many critical problems associated with the applications of OFDM technologies in future wireless systems are still looking for efficient solutions. This would overcome the global issues and challenges facing the limited bandwidth in wireless communication network.

Keywords: - *MATLAB/Simulink, OFDM, high-speed data rate, and multicarrier signals*

1. Introduction

In these days, the need of multicarrier network and high speed data rates on mobile communications networks becomes high due to many issues related to wireless transmission and broadband and multimedia applications. The multi-carrier high-speed data rates technologies are the answer to these issues. A summary of major research in significant stages are noted that used of multi-carrier high-speed data rates technologies such as OFDM systems to replace the used of single-carrier modulation techniques are necessary (Mondragon-Torres, Kommi and Battacharya (2011). Other contributions of multi-carrier modulation techniques include the limitations of bandwidth, and the speed of the data rates or power of transmission which need significant increase. However, OFDM transmit different sub-carriers; between sub-carriers is guard-band to avoid interference that is to eliminate or reduce inter-symbol interference (ISI) and inter-carrier interference (ICI). The sub-carriers are designed to be orthogonal, these allow sub-carriers to overlap and saved Bandwidth, that achieving higher data rate signals (Bodhe, Narkhede and Joshi (2012).

Furthermore, in single carrier network if you send single carrier frequency and when that carrier frequency become useless, due to either fading or other interference like ISC and ICI one station loses of service. But since OFDM takes one-word carrier and spreads it over several smaller carriers, so even if you have one fading in one specific frequency the station still has other useful channels, which can still convey the information. So instead of numerous adjacent symbols being destroyed, only some symbols are little bit distorted, spectral overlapping among sub-carriers is allowing improved spectral efficiency and used of steep band pass filter is eliminated (Premnath, Wasden and Farhang-Boroujeny (2013). In multicarrier transmission where subcarriers are orthogonal to each other in frequency domain, it also distributes data over many numbers of carriers that are spread out at precise frequencies. This spacing gives the orthogonality in this method that prevents the demodulators from detecting other frequencies rather than their frequency. This process is demonstrated in figure 1 below by the guard interval and cyclic prefix. Cyclic prefix it reduces or even eliminates the ISI. This technique supports simple frequency domain processing, equalization and channel estimation. It used to change multipath by making channel estimation simple.

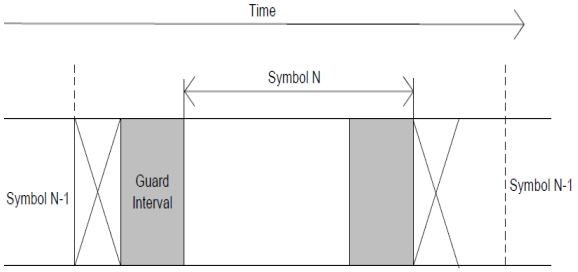


Figure: - 1. Guard Interval and cyclic Prefix (Gulzar., Nawaz and Thapa (2011)

In single carrier system frequency selective channels caused ISI at the receiver; equalization will apply noise critically in frequencies where channels reaction is poor as result single carrier performance is disturb because of higher attenuation in some bands since all used frequencies are haven equal important (Bodhe, Narkhede and Joshi (2012). But in multicarrier (OFDM) the spectrum used is divided into narrow sub-bands and then separate data is transmitted in each band using different carriers. Power and rate of transmission

in a different band depend on the responds of the channel in that band, and there is no ISI since in each narrow sub-band the channel responds are almost flat (Baltar and Nossek (2012). The first product profiles for mobile WiMAX have yet to be chosen as the standard is not yet approved, OFDM is the very key modern multi-carrier wireless technology forms the basis for 4G (fourth generation) wireless communication system in LT and WiMAX (LaSorte, Barnes and Refai, (2008).

In the OFDM the sub-carriers are designed to be orthogonal, these allows sub-carriers to overlap and saved Bandwidth, that make data rate higher (Baltar and Nossek (2012). In OFDM the carriers are all generated by single transmitter in a special way that allow them to be tight much closer together and spend a much wider Bandwidth that means we can reduce or eliminate Guard-Band. And sub-carriers can be packed tightly without interference with each other and fall offset at band edges. The different between non-overlapping multicarrier signals and overlapping ones as using overlapping methods can save much more bandwidth up to 50% than the non-overlapping one and used to transmit the same signal (Hariprasad and Sundari (2015). The figure: - 8 below shows OFDM system build using the basic block of Simulink using BPSK. The general block diagram of OFDM is illustrates in figure 2 below:

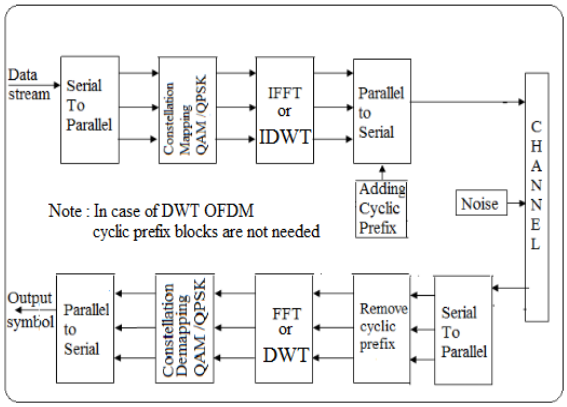


Figure: 2. General Block Diagram of OFDM System (Kanpur (2013)

A. The main aims and objectives of this research work are:

To design and develop OFDM systems that; show both OFDM transmitted and received signals (multi-carrier signals for information exchange between systems).

1. To study and understand the principle of OFDM in wireless communications network.
2. To study characteristics of wireless communications channels and understand the performance of a simulation for different scenarios
3. Finally, to design and develop a system which will produce multi-carrier signals, significantly increase the speed of data rate on the simulation results, which can be used and be expanding in wireless communication network.

2. Literature Review

This section provides an overview on the other research works that have been done on OFDM systems the researches that carried out indicate the importance of OFDM systems by many researched work. Many researchers give different design solution to the challenges associated with multiple carrier transmission using OFDM systems. Many studies have been getting desired results which can be used to improve the system performance and reduce the effect of ISC and ICI which still need further enhancement.

2.1 Global Wireless Communication systems

High data rate up to Gbps and precisely operate in 60-GHz frequency will be used in Broadband wiless communications. Furthermore, one can see this matter from another point of view, many people debating whether such higher-capacity systems is required, beleaving that all of the compression procedure deloped and the type of applications 10s of meggerbits per second. Since, high- capacitysystems become necessary there is a need to give perspective of what should be the ‘’underline research topic’’ in the area of the telecommunications. In views of what are the possible solution, in order to folow the demands of society in the future as far as communication is concerned, systems capacity is one of the major matters to be addressed due to the forseen increase in needs for new services, based on multimedia. personnel mobility will developed new challenges to the development of new personel and mobile communication systems. According to (Prasad (2004) ‘it is unsafe to limits on wireless data rates considering economic constraints since broadband is depend on data rate capacity’.

From this it can be concluded that: Even if at an unquestionabe point it may look “ academic’’ to developed a system for a high capacity much high than required, even though there are no need for such high capacity system appllications. It is suffecently good while to do it since almost surely in the fureture such applications will be in reality for the need of such capacity system or even beyoung that 5 Gigger bits per second, the story of fiber optics is shed light on that (Prasad (2004).

I. Wireless Technology Today

In this present day, basically, five wereless technologies have made an impact, namely, wireless global area networks (WGANs), wireles wide area networks (WWANs), wireless personal area networks (WPANs),

wireless local area networks (WLANs), and wireless broadband-personel area networks (WB-PANs), as demonastrates in figure 3.

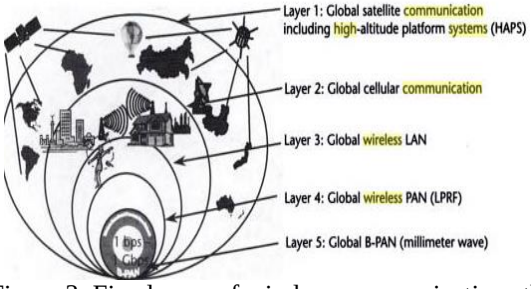


Figure 3: Five layers of wireless communications that provide [10]

II. Wireless Technology in the Future

To make sure that, all process associated with better performance of services of wireless technology, each wireless technology is moving toward future standardization. This standardization work is focusing on wireless technology for provision of any types of data. Here data consist of everything, audio, gamming, video, or any other application. A possible future situation is illustrates in figure 4, all relevant parties in wireless communication technologies should work together while providing all of the services to users anywhere any time. Clearly, it is believed that the OFDM systems are next target technologies of most modern wireless communication system (Kulkarni and Bhalchandra (2012).

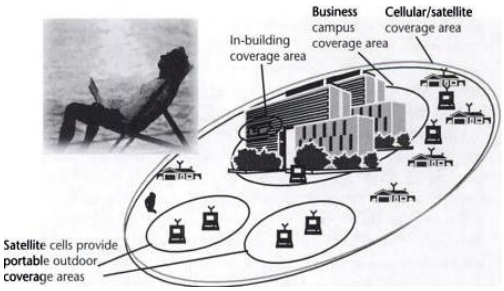


Figure 4: Future of Wireless Standardization [10]

2.2 The concept of OFDM

The idea of OFDM was intended in earlier 1960s, it was not able to be attained until the emergence of Fast Fourier transform (FFT) with the emergence of inverse Fast Fourier transform (IFFT) was made possible to generate OFDM system (Gomes, Al-Daher and Fernandes (2014). The idea of using Multi-Carrier technologies such as OFDM parallel-data transmission and frequency division multiplexing (FDM) was earliest started in the middle of 1960s (Kulkarni and Bhalchandra (2012).

OFDM involves FDM which divide the available Bandwidth into many subcarriers and allow multiple users to access the system simultaneously. In multi-carrier FDM the data of the user can be separated into multiple sub-stream and transmit them in parallel, to make the data rate high. We can see in figure 5 below how using overlapping methods can save much more bandwidth than the non-overlapping one, 50% of bandwidth used to transmit the same was saved (Bodhe, Narkhede and Joshi (2012).

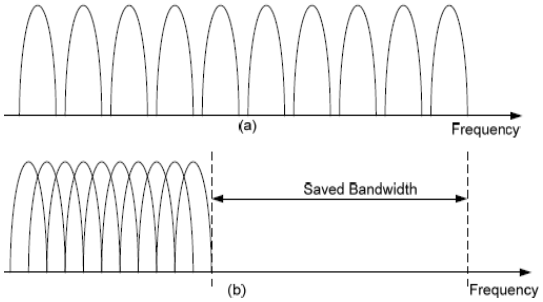


Figure: 5. Conventional multicarrier method (a) - FDM, and (b) - OFDM (Hariprasad and Sundari (2015)

3. Methodology

This stages shows the techniques approached adapted to carry out the research work, of designing and developing of OFDM systems, which would be simulated on MATLAB/Simulink software to produce OFDM input and output signals. These have been achieved by reviewing of high-quality research papers.

3.1 Implementation

After the systems are designed and simulated this section provides comprehensible explanation about the final stages of the search work, and simulation results obtained by the system. The final stages of this research work implementation were OFDM input and output signals from the simulation results as shown in figures below using different modulation schemes.

The first stage of the design is flow chart which illustrates in figure 6 below.

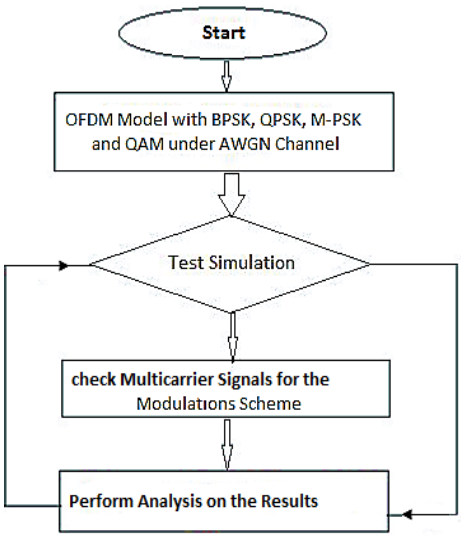


Figure 6: Flow Chart

3.2 The OFDM Model

The second stage of the design is the OFDM system as shown in figure 7 below. This OFDM model was designed using Simulink blocks in MATALAB/Simulink software, OFDM spectrum was achieved using Mux and Demux which were used in serial to parallel and parallel to serial conversion respectively, but better result were obtained using Buffer and Unbuffer to replace the used of Mux and Demux. BPSK, QPSK and QAM modulator and de modulator with up to 32 numbers of subcarriers used were used to carry out the OFDM model system designed differently, to simulate the FFT OFDM and IFFT OFDM model system. The modulators and demodulators are placed in transmitter and receiver side accordingly. The generated OFDM input signal of the FFT transmitter transmits without noise to the Multipath Rayleigh fading and AWGN channels. The IFFT receiver signals received with the channel noise as expected.

The transmitted signal is likely free from out-of-band harmonics, the out of band harmonics were more pronounced in received signal. Thus, the receiver was subject to certain degree of noise, errors are introduced in the demodulation and decoding processes. The performance of BPSK, QPSK and QAM simulations modulators and demodulators placed in transmitter and receiver side was observed in the frequency spectrum. Both signals were observed and analysed the effect of noise over AWGN and multipath Rayleigh fading channels by comparing the transmitted and received signals and compared the simulated data with theory. This paper has achieved better performance in terms of OFDM spectrum in FFT-IFFT based OFDM transmitter and receiver that for input signals were the same with output signals.

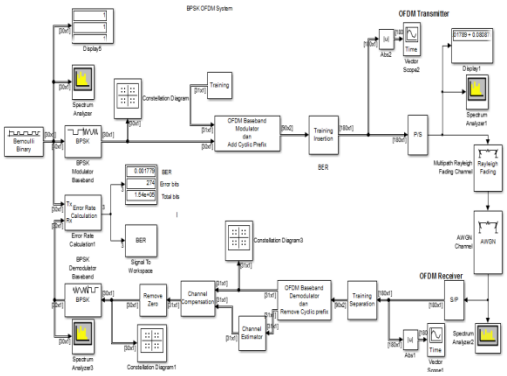


Figure 7: BPSK OFDM System

4. Simulation Results of OFDM input and output Signals

However, OFDM simulation results between the transmitter and receiver signals, transmitted single carrier signal and received multiple carrier signals in the frequency spectrum respectively. Both signals are observed and analyze the frequency spectrums conclude the results by comparing the transmitted and received signals and compared the simulated data with theory. The simulation results obtained has achieved better results.

4.1 BPSK, QPSK and QAM Simulation Results of OFDM spectrums

The OFDM receiver and transmitter signals are illustrates in figures 8 to 43 below thus, the transmitted OFDM signal transmitted without noise to the Multipath Rayleigh fading and AWGN channels. Likewise, the OFDM receiver signal received with channel noise as anticipated, channel induced noise disturbs the received signal, and OFDM output signal is the same with original input signal.

4.2 Analysis of Power Spectrum of BPSK, QPSK and QAM OFDM Signals

In each of the BPSK, QPSK and QAM four OFDM signals were observed, namely; input, transmitter, receiver and output signals respectively. However, by locking at the higher of the display in the power spectrum (dBm) which is the power of the frequency, it was observed that the single carrier signal was divided into narrow sub-bands in the transmitter and receiver signals. The sub-carriers were observed to be orthogonal with channel noise on the receiver signals. The spectrum signals confirm that QAM has less noise effect than QPSK and BPSK under the same transmission rate. The overall power spectra for the cases of OFDM systems with 32- QAM subcarriers has better performance in terms of noise enhance effect. The higher data rate signals were achieved since sub-carriers were designed to be orthogonal, overlap and saved Bandwidth as illustrates in figure 8 to 43 bellow.

4.2.1 The BPSK modulator

Baseband is a modulation scheme that modulates using the binary Phase Shift Keying technique, the output signal is baseband presentation of the modulated data. This modulation scheme accept vector and column input data, the input must be discrete time binary value signal. The output data type of this modulation scheme could be set to Inherit via back propagation, double, user-define, and single fixed-point.

4.2.4 QPSK Modulator Baseband

The modulator takes a scalar of column vector data. If input type parameters set to Bit, the input contains pairs of binary values; the modulator would takes column vectors with event lengths. And if the phase offset parameters set to $\pi/2$.

4.2.5 AQM Modulator Baseband

The Rectangular QAM modulator baseband is a modulation scheme that modulates the input signal using M-array quadrant amplitude modulation; it also has a constellation on a rectangular lattice. The output is baseband presentation of the modulated data.

4.3 BPSK Simulation Results of OFDM spectrums

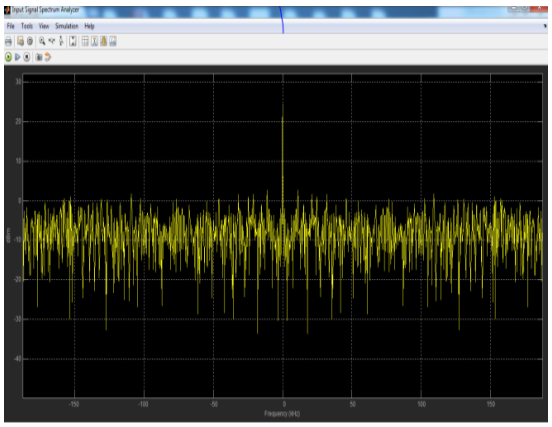


Figure 8: BPSK-OFDM input signal

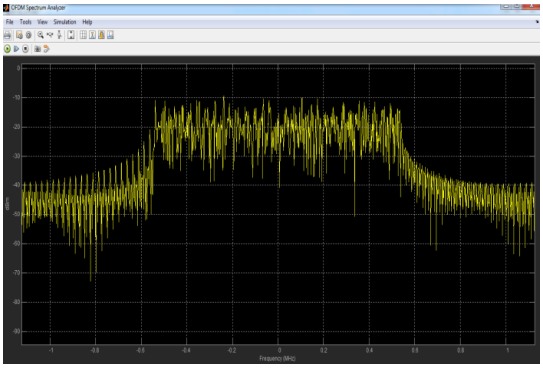


Figure 9: BPSK-OFDM transmitter signal

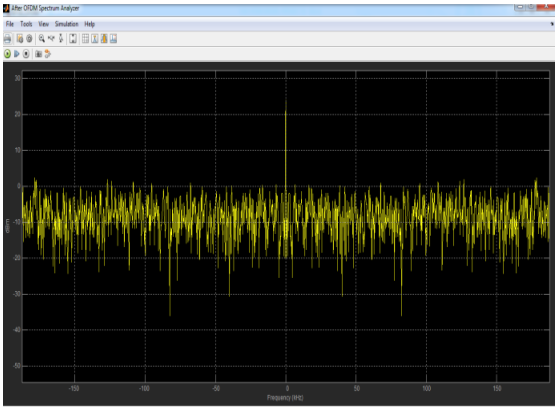


Figure 10: BPSK-OFDM Received signal

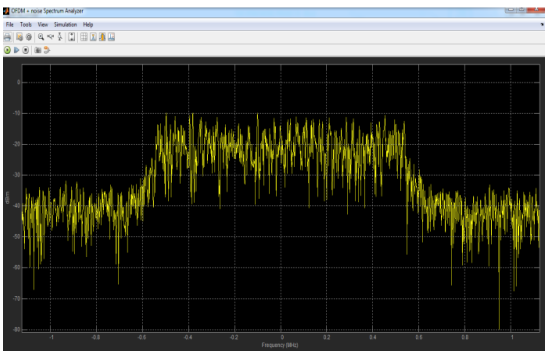


Figure 11: BPSK-OFDM Output signal

4.4 QPSK Simulation Results of OFDM spectrum

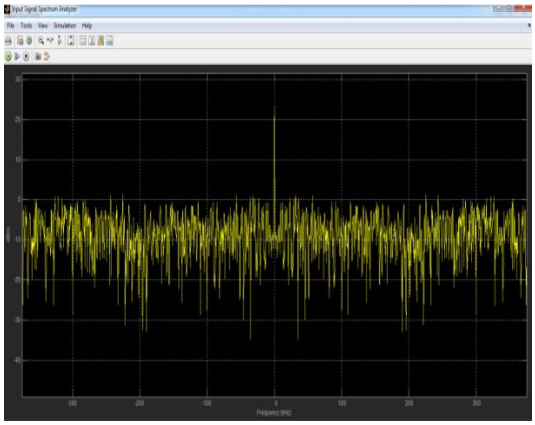


Figure 12: QPSK-OFDM input signal

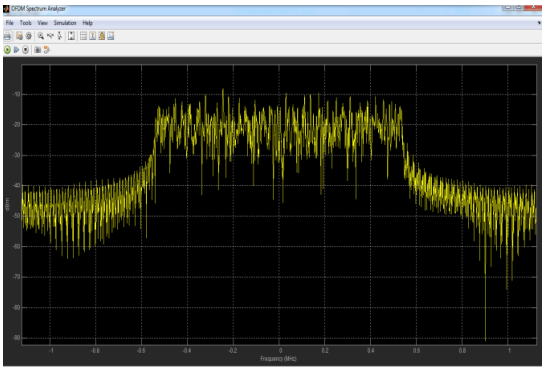


Figure 13: QPSK-OFDM transmitter signal

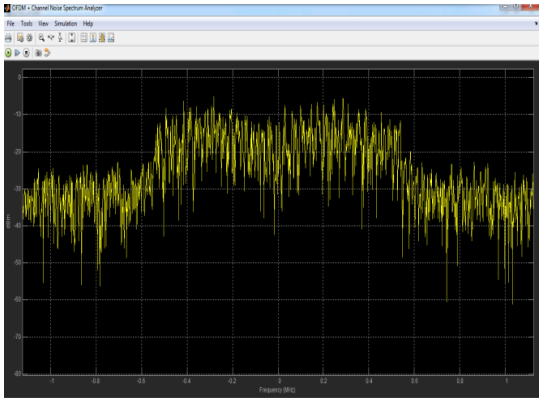


Figure 14: QPSK-OFDM Received signal

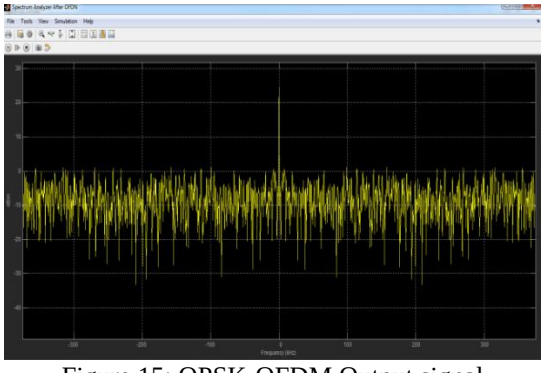


Figure 15: QPSK-OFDM Output signal

4.5 8-PSK Simulation Results

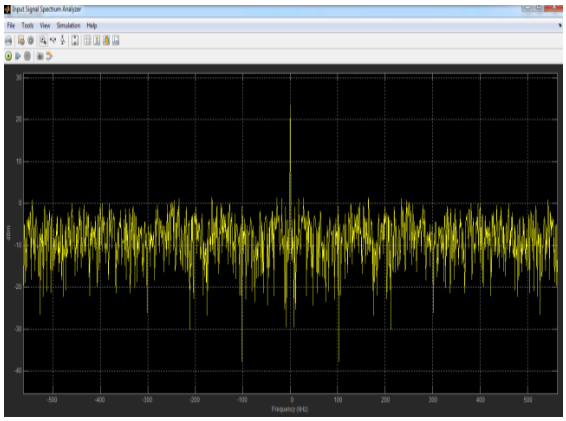


Figure 16: 8-PSK-OFDM input signal

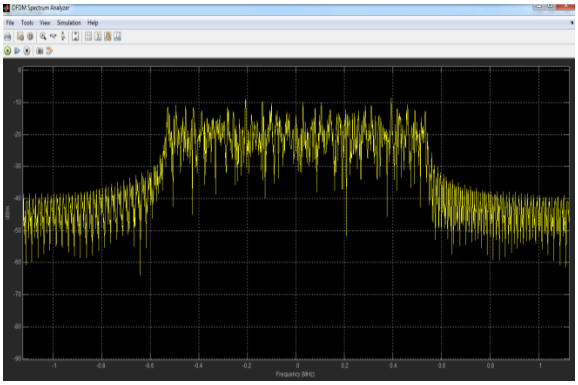


Figure 17: 8-PSK-OFDM transmitter signal

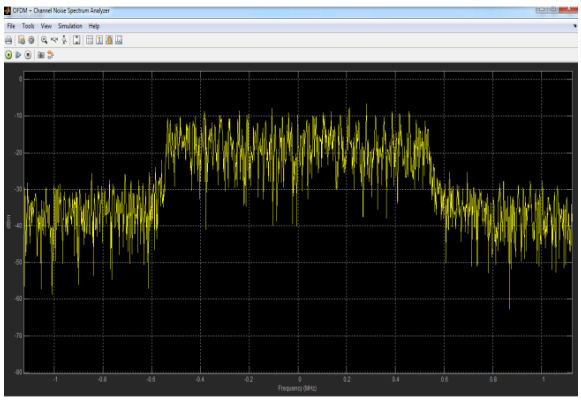


Figure 18: 8-PSK-OFDM receiver signal

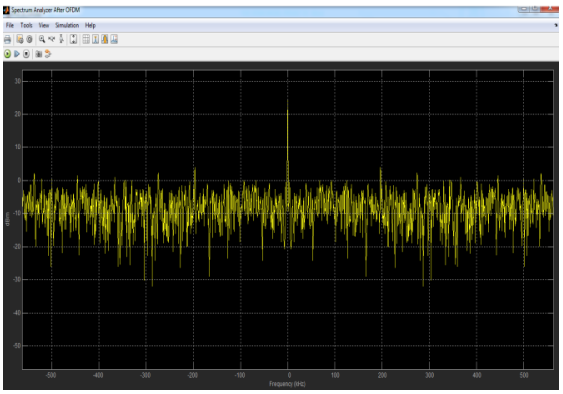


Figure 19: 8-PSK-OFDM Output signal

4.6 16-PSK Simulation Results of OFDM

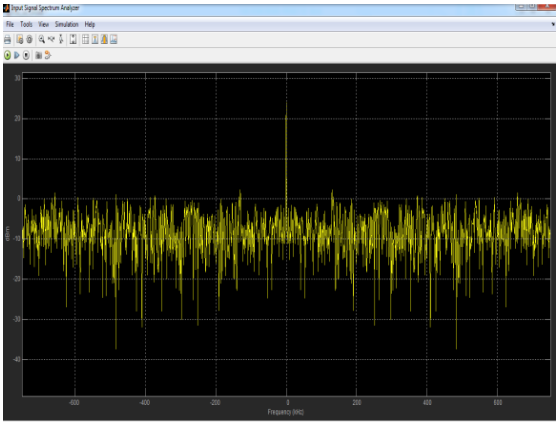


Figure 20: 16-PSK-OFDM input signal

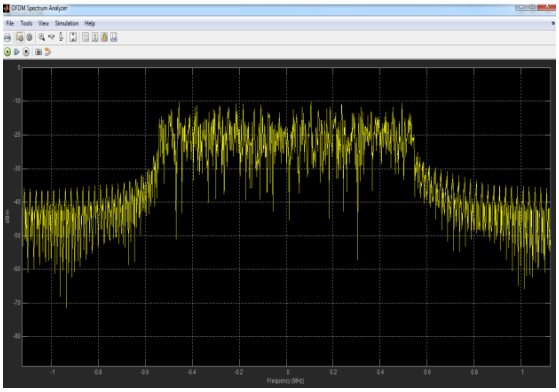


Figure 21: 16-PSK-OFDM transmitter signal

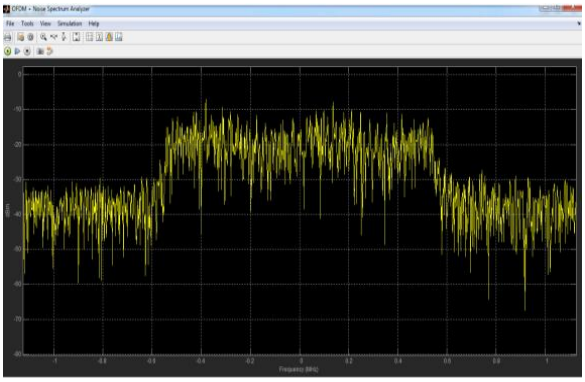


Figure 22: 16-PSK-OFDM Received signal

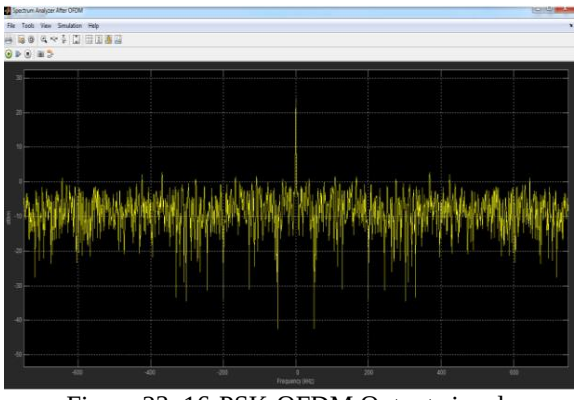


Figure 23: 16-PSK-OFDM Output signal

4.7 PSK Simulation Results of OFDM spectrum

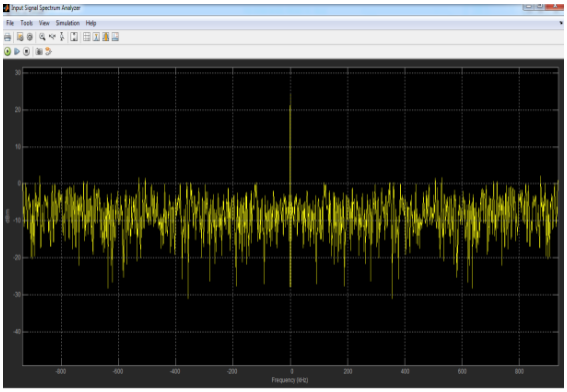


Figure 24: 32-PSK-OFDM input signal

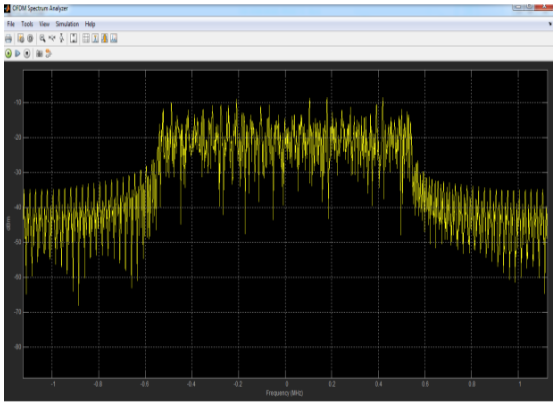


Figure 25: 32-PSK-OFDM transmitter signal

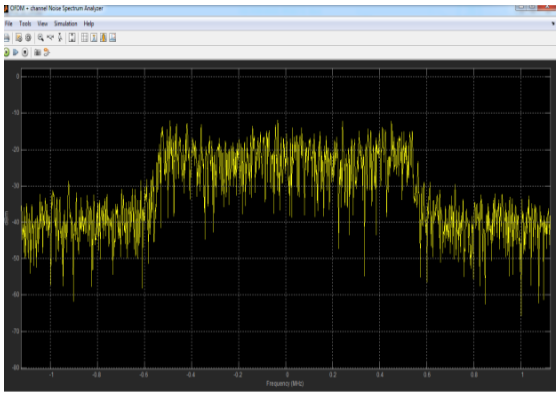


Figure 26: 32-PSK-OFDM Received signal

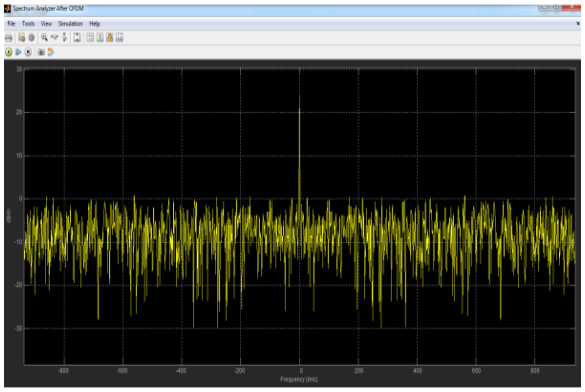


Figure 27: 32-PSK-OFDM Output signal

4.8 4 QAM Simulation Results of OFDM spectrum

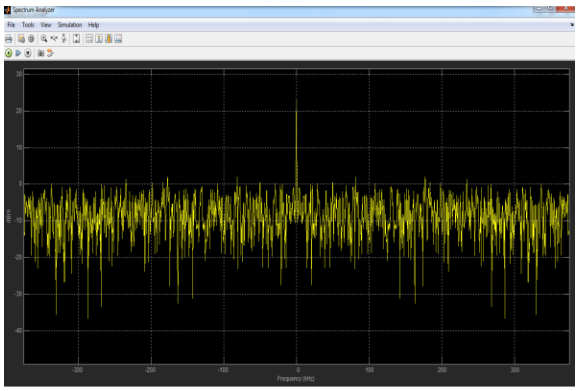


Figure 28: 4-QAM -OFDM input signal

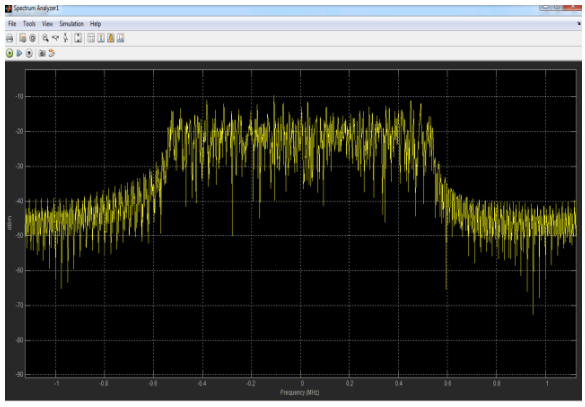


Figure 29: 4-QAM-OFDM transmitter signal

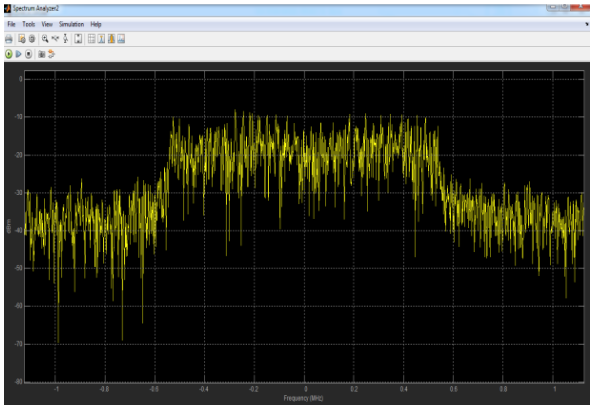


Figure 30: 4-QAM-OFDM Received signal

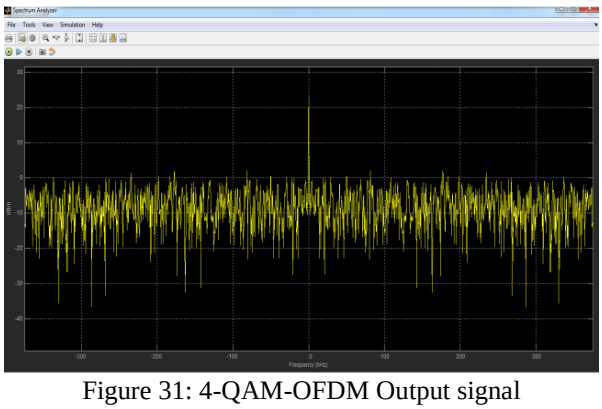


Figure 31: 4-QAM-OFDM Output signal

4.9 8-QAM Simulation Results of OFDM spectrum

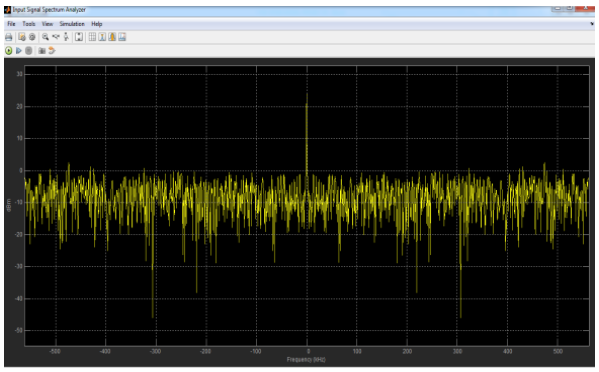


Figure 32: 8-QAM-OFDM input signal

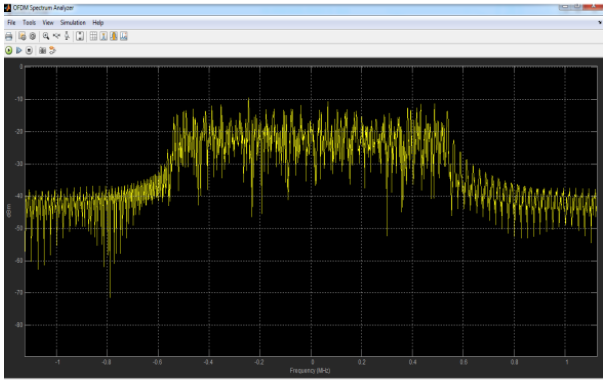


Figure 33: 8-QAM-OFDM transmitter signal

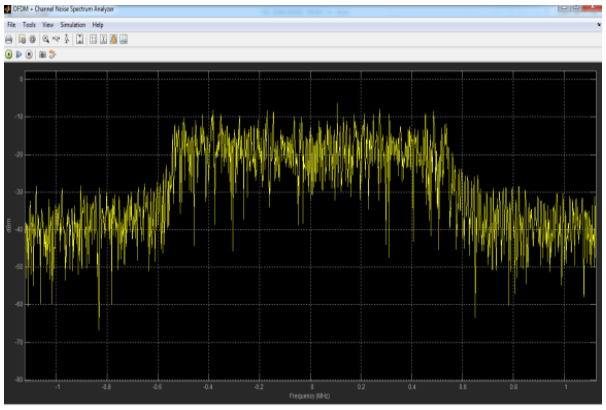


Figure 34: 8-QAM-OFDM Received signal

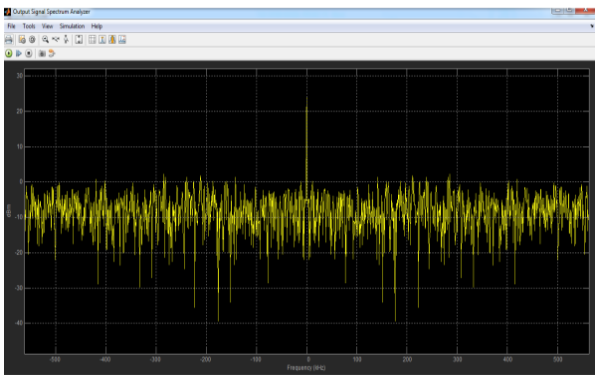


Figure 35: 8-QAM-OFDM Output signal

4.10 16QAM Simulation Results of OFDM spectrum

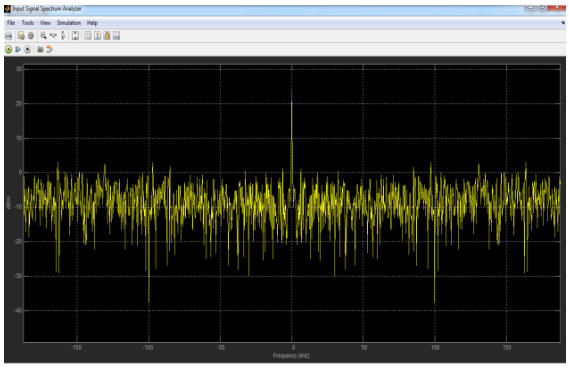


Figure 36: 16-QAM-OFDM input signal

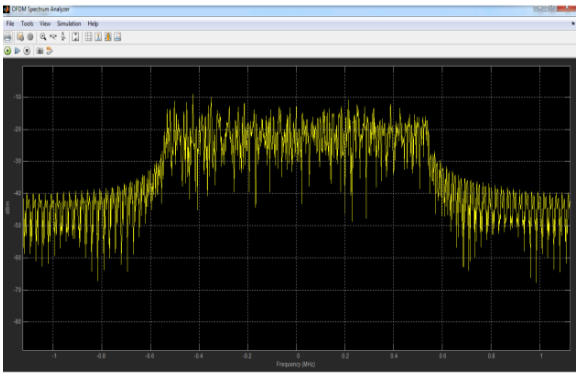


Figure 37: 16-QAM-OFDM transmitter signal

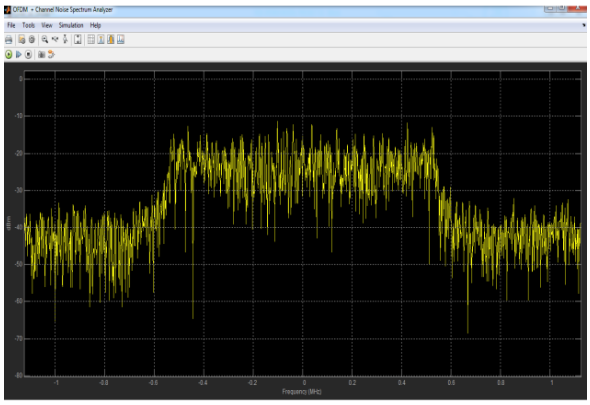


Figure 38: 16-QAM-OFDM Received signal

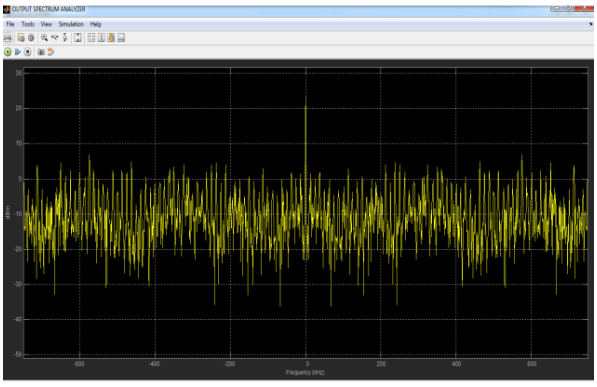


Figure 39: 16-QAM-OFDM Output signal

4.11 32-QAM Simulation Results of OFDM spectrum

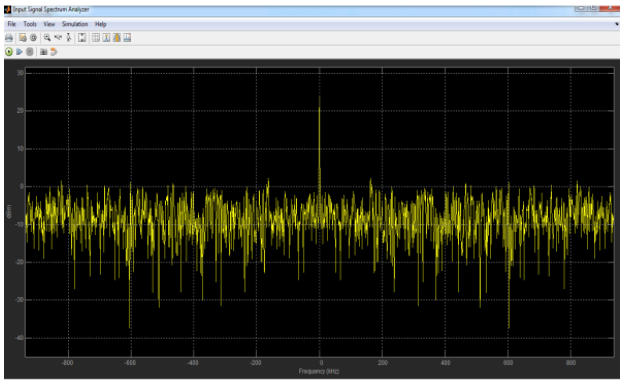


Figure 40: 32-QAM-OFDM input signal

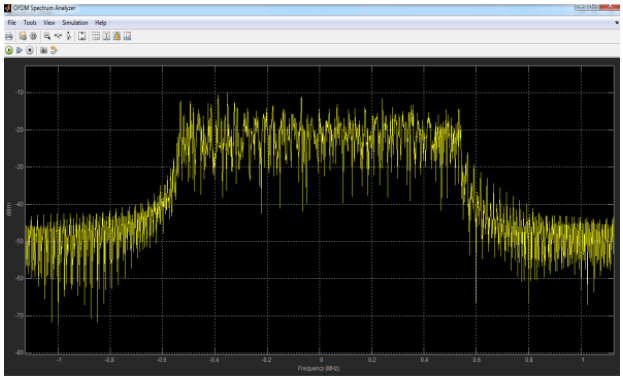


Figure 41: 32-QAM-OFDM transmitter

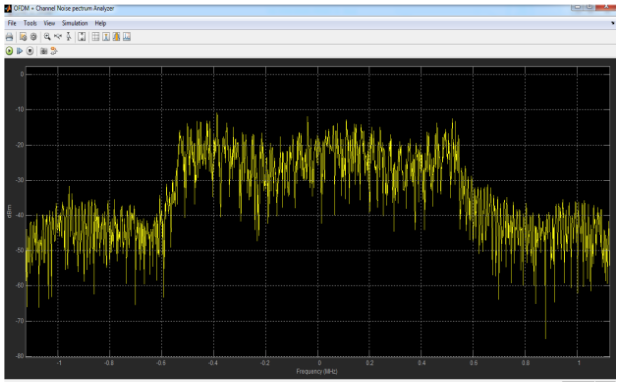


Figure 42: 32-QAM-OFDM Received signal

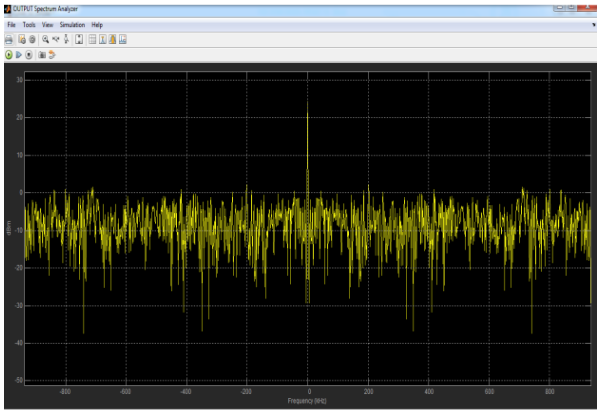


Figure 43: 32-QAM-OFDM Output signal

4 Conclusion

As it demonstrates above in the OFDM spectrums the aims of the research work were achieved, the paper shows the spectra of the sub-carriers accurately, noise effect on the OFDM transmitted signals and how OFDM return the channel noise sine output signals were the same with the input signals. The input (single) carrier signal was divided into narrow sub-bands and then separate data is transmitted in each band using different carriers. Power and rate of transmission in a different band depend on the responds of the channel in that band. The sub-carriers are designed to be orthogonal, these allows sub-carriers to overlap and saved Bandwidth, that achieving higher data rate signals.

A user close to the base station would normally be assigned a larger number of channels with high modulation scheme such as 32 QAM (quadrature amplitude modulation). The modulation scheme could gradually shift from 16 QAM to Quaternary Phase Shift Keying (QPSK) (four channels) and even binary phase shift keying (BPSK) (two channels) at longer ranges. The data amount drops as the channel capacity and modulation change, but the link maintains its strength.

5 Recommendation

Since the data amount drops as the channel capacity and modulation change at longer ranges, a high modulation scheme such as 64 QAM (quadrature amplitude modulation) and above should be used in the next OFDM system design.

REFERENCES

Baltar, L. G. and Nossek, J. A. (eds.) (2012) OFDM 2012, 17th International OFDM Workshop 2012 (InOWo'12); Proceedings of. 'Multicarrier Systems: A Comparison between Filter Bank Based and Cyclic Prefix Based OFDM'

Bodhe, Narkhede, Joshi, R. (2012) Design Of Simulink Model For Ofdm And Comparison Of Fft ofdm And Dwt-Ofdm [online] available from <https://www.idc-online.com/technical_references/pdfs/electronic_engineering/DESIGN%20OF%20SIMULINK%20MODEL%20FOR.pdf> [16 September 2018]

Gomes, R., Al-Daher, Z., Hammoudeh, A., Sobaihi, K., Caldeirinha, R. and Fernandes, T. (2014) 'Performance And Evaluation Of OFDM And SC - FDE Over An AWGN Propagation Channel Under RF Impairments Using Simulink At 60Ghz'. 2014 Loughborough Antennas and Propagation Conference (LAPC)

Gulzar, M., Nawaz, R. and Thapa, D. (2011) Implementation Of MIMO-OFDM System For Wimax [online] available from <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A421361&dswid=4817> [10 July 2018]

Hariprasad, N. and Sundari, G. (2015) 'Comparative Analysis Of The BER Performance Of DWT OFDM Over That Of FFT OFDM In Presence Of Phase Noise'. Robotics, Automation, Control and Embedded Systems (RACE), 2015 International Conference on [online]1-4.availablefrom <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7097288> [1 July 2018].

Kanpur, K. (2013) OFDM Simulator Using MATLAB [online] available from <http://www.ijetae.com/files/Volume3Issue9/IJETAE_0913_79.pdf> [10 July 2018].

LaSorte, N., Barnes, W. and Refai, H. (2008) 'The History Of Orthogonal Frequency Division Multiplexing'. IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference

Mondragon-Torres, A., Kommi, M. and Battacharya, T. (2011) 'Architecture Exploration, Development And Teaching Platform For Orthogonal Frequency Division Multiplexing (OFDM) Systems'.2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR).

Prasad, R. (2004) OFDM for Wireless Communications Systems.:Artech House.

Premnath, S., Wasden, D., Kasera, S. K., Patwari, N., and Farhang-Boroujeny, B. (2013) 'Beyond OFDM: Best-Effort Dynamic Spectrum Access using Filterbank Multicarrier'. Networking, IEEE/ACM Transactions on 21 (3), 869-882.

Surekha, T., Ananthapadmanabha, T. and Puttamadappa, C. (2011) 'ModelingAnd Performance Analysis Of QAM-OFDM System With AWGN Channel'. 2011 Third Pacific-Asia Conference on Circuits, Communications and System (PACCS)

AUTHORS BIOGRAPHY

My name is **Engr. Muhammad Bello Abdullahi MNSE;** I am indigent of Bungudu local government, Zamfara State, Nigeria and an academic staff of Abdu Gusau Polytechnic, Talata Mafara, Zamfara State, Nigeria. I am currently lecturing in the department of electrical and electronic engineering as senior lecturer, serving as head of the department. I have obtained Master of Science (MSc) in electrical and electronic engineering from Coventry University United Kingdom. I am registered engineer with the Council for the Regulation of Engineering in Nigeria (COREN) and member Nigerian Society of Engineers (NSE). I have published many papers in the polytechnic journal as well as international journals some of them include; Securing Nigeria's Sustainable Economic Growth through Online Marketing, Distribution Automation the Better Way forward in Nigeria's Power Supply, HVDC Power Network, Evaluation of BER Performance of FFT-OFDM for Wireless Communication Network and Free-Space Optical Communication, Design And Implementation of Helical Antenna For Transceiver Propagation. I have always had an interested in the field of electrical and electronic engineering particularly communication engineering. The cause is very personal; it could be my ability for working independently, team work and problem solving.

SIMULATION MODEL OF SUPPLY NETWORKS DEVELOPMENT

Petr Fiala[(a)], Martina Kuncová[(b)]

[(a),(b)]College of Polytechnics Jihlava

[(a)]pfiala@vse.cz, [(b)]kuncova@vspj.cz

ABSTRACT
The paper is dedicated to network development in the network economy. The current economy needs to look not only at networks with only dynamic flows and with a fixed structure, but as a dynamic system its structure evolves and changes. Structure and behaviour dynamics of network systems can be modelled as complex adaptive systems and use agent-oriented simulation to demonstrate origin, perturbation effects, and sensitivity with regard to initial conditions. Survival of firms is associated with the value of so-called fitness function. Firms whose fitness value falls below a certain threshold will be extinguished. In this way, it is possible to partially model network growth. A simulation model in SIMUL8 is proposed.

Keywords: network economy, supply networks, simulation, dynamics

1. INTRODUCTION

Dynamic models for the description and analysis of economic systems play a crucial role (Fiala, 2009, 2016). Growth dynamics and network development can be modelled by expanding the Utterback industrial growth model (Utterback, Suarez, 1993, Utterback 1994) and using the concept of behavioural patterns to form networks. It can be assumed that the pattern of supply networks will be similar to the bell curve proposed by Utterback. Simulation model divides a complex adaptive network into two basic parts:

- the supply network environment,
- networked firms.

Firms are represented by nodes in the network. The nodes operate on the basis of simple decision-making rules, with an effort to meet the environmental demand. As new industries emerge, supply networks begin to develop, new businesses can enter the market. Some firms are successful and build relationships with others, and there is growth. Some firms are unsuccessful and disappear, either because of local conditions or because they cannot become part of a viable supply network. In time, supply networks can grow into relatively stable structures, based on the interactive effects of local decision rules and environmental factors. CAESAR (Pathak, Dilts, Biswas, 2003; Pathak, Dilts, 2004) and its advanced Complex Adaptive Supply Network simulator (CAS-SIM) version have been designed to monitor supply network developments, including patterns of emergencies and extinctions.

The aim of the paper is to describe a possibility of the conceptual and simulation models creation based on the idea of the Utterback model. The simulation software SIMUL8 is used. Chapter 2 describes the main facts about the Utterback model. In chapter 3 the conceptual model of a network using arcs and nodes is created. Afterwards the architecture and algorithm that can be used in any simulation software to model the behaviour of the supply network are described (chapter 4 and 5). Finally, the SIMUL8 software is used to demonstrate the possibility to model the supply network similar to the Utterback ideas.

2. UTTERBACK MODEL

As new industries emerge, supply networks grow, creating new relationships between firms that cooperate to meet demand. Utterback industrial growth model (Utterback 1994) takes into account that at the beginning are low barriers to entry for firms in the industry and is not clearly defined market structure. At this stage, many firms want to enter the industry by becoming leaders. In the next stage, there is a desire to create a clearly defined market structure with a focus of firms on economies of scale and network externalities. Not all firms are successful, those unsuccessful are pushed out of the market and over time the number of firms is decreasing. The course of the individual phases of the industrial growth cycle creates a bell curve. The Utterback model is used as a premise that is verified by practice but can be modified in the simulation process.

Utterback (1994) also assumes that companies learn to play specialized roles over time. At the beginning of the development of the sector, all companies are trying to play a versatile role, then there remains less and the others either expire or specialize. The Utterback model takes into account the growth of the industry only with regard to the number of entering and exiting firms.

Another dimension that should be taken into account is the size of companies. In one possible scenario, with market growth over time, there is a different growth in firms. Some companies are significantly expanding capacity, while others weaken. An alternative scenario may be the fact that no companies become dominant and the market is relatively evenly distributed among participating firms. In a comprehensive adaptive supply network model, two aspects are considered:

- number of firms,
- size of firms.

3. CONCEPTUAL MODEL

The model (Figure 1) is based on a supply network as a system consisting of an environment (market) in which companies (nodes - N) create interactions (relations - R) based on simple rules of conduct to meet global demand. Stochastic environmental parameters, describing market conditions and demand, a node-based decision-making scheme, so-called fitness functions modelling the strength of companies, all influence the dynamics of structure and behaviour of the developing supply network.

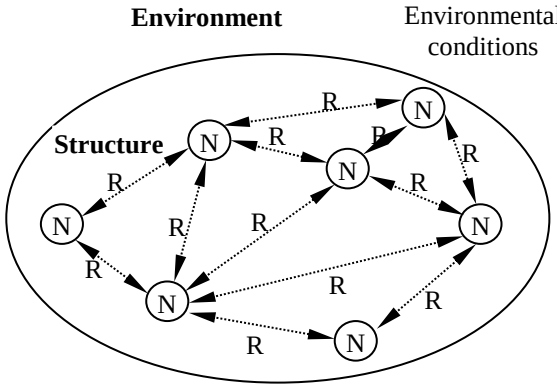


Figure 1: Conceptual model

Complex Adaptive System (CAS) is well suited for modelling systems with structure and behaviour dynamics. The CAS is characterized by three main components:

- environment,
- internal mechanisms,
- coevolution.

Firms (nodes) are emerging, exist and disappear in the environment. The environment is characterized by two main groups of conditions:

- operating conditions,
- setting up the market structure.

Operating conditions specify demand and time and cost information. Market structure settings specify the regulation, policies, and trading rules applied in the system. The environment also sets the fitness threshold, which is the minimum level of fitness required to survive the node in the environment.

Internal mechanisms concern nodes, rules, relations, and decisions in networks. The nodes represent the firms (agents) in the supply network. Nodes focus on achieving local goals, with each node having a set of possible strategies to achieve them. The rules allow the implementation of these strategies to achieve goals while respecting the constraints given by the environment and the node itself.

Nodes make two basic types of decisions:

- with whom to cooperate,
- strategic decisions.

Choosing a node with which to cooperate and create relationships is also partly affected by market rules. Strategic decisions include capacity determination, product prices, outsourcing levels, etc.

Coevolution describes changes in states and the formation of quasi-equilibrium in the system. The results of implementing node strategies in a specific environment generate coevolutionary network structures. Coevolution is the result of interaction between the environment in which the supply network exists and the internal mechanisms used by the nodes in the network. The coevolutionary process will result in different firms' growth relative to the values of built-in fitness functions.

4. SIMULATION MODEL ARCHITECTURE

To understand the growth and development of dynamics, it is necessary to monitor the time-dependent behaviour of the model. Simulation is a frequently used methodology for analysing the time-varying properties of a system. A multi-paradigmatic architecture can be used to create a simulator (Figure 2). In multi-paradigm architecture, some of the model's components fit the discrete-time modelling paradigm, such as Environment, while others fit the discrete-event modelling paradigm, such as individual node behaviour. Environment and Evaluator, which acts as an environment controller, are clustered models that interact with nodes in the system.

Evaluator acts as a coordinator of all nodes and communicates with them through the protocol, launches nodes and sends them information about demand and other messages. After a certain number of demand cycles, he evaluates the nodes by the value of the fitness function and the failed nodes disappear. Nodes are elementary models of discrete events owned and coordinated by Evaluator. Nodes respond to seven basic events:

- break,
- message
- flag,
- demand,
- bid request,
- time

- fitness update.

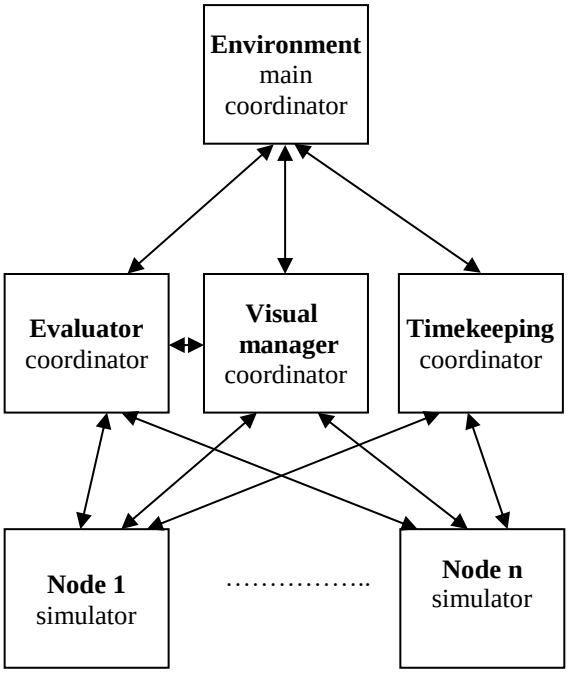


Figure 2: Multi-paradigmatic architecture

5. SIMULATION ALGORITHM

According to the previous conceptual model and the basic architecture of the simulation model connected with the supply network, it is possible to define an algorithm for a simulation model preparation. Figure 3 shows a sequence of events in a simulation algorithm cycle that leads to the development of a supply network. The simulation starts with the initialization of the environment and the setting of external system parameters such as the start of the simulation watch, the creation of a demand function, the activation of Evaluator part and the assignment of values of other conditions. After Environment is initialized, an initial number of network nodes is generated.

Environment triggers a new demand cycle and the Evaluator divides demand between all nodes based on a set market structure. Nodes interact on the basis of internal mechanisms to meet demand in a given period. Final production is delivered through cooperating nodes to end customers. After calculating the gains and losses of individual nodes, the values of their fitness functions are recalculated. Evaluator periodically checks the node condition in the population and takes nodes whose values have fallen below the thresholds. New nodes are introduced into Environment depending on demand and supply. The number of simulated demand cycles is determined during Environment initialization process. The course of time is given explicitly in the supply network by the fact that demand cycles run at regular intervals. Individual behaviour of nodes and their interaction are determined by discrete events.

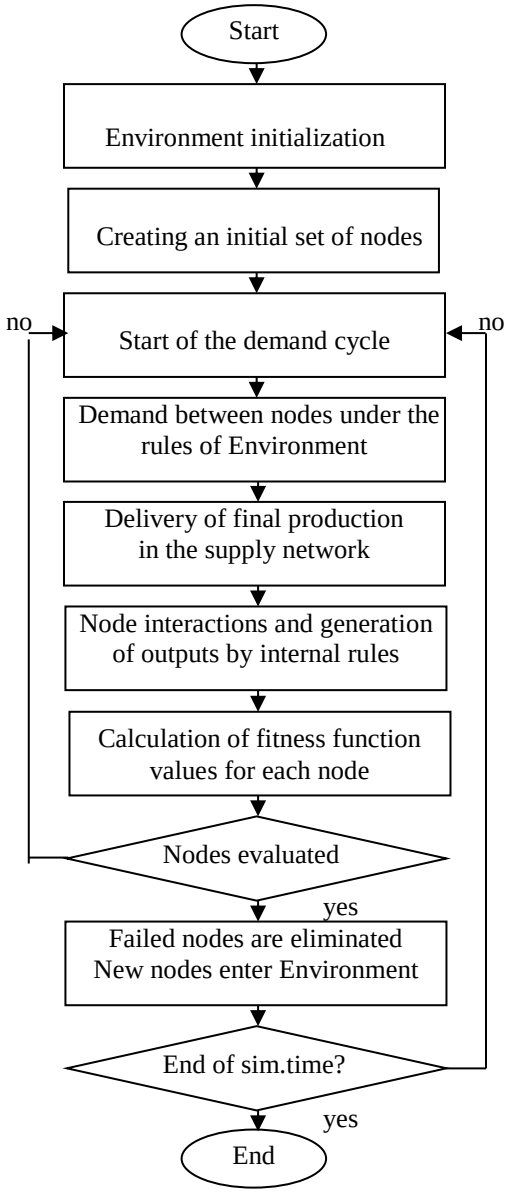


Figure 3: Simulation algorithm

6. SIMULATION MODEL IN SIMUL8

SIMUL8 is a software package designed for Discrete Event Simulation or Process simulation. It has been developed by the American firm SIMUL8 Corporation (www.simul8.com). The software has been used since 1994 and usually every year a new release has come into being. It allows a user to create a visual model of the analysed system by using pre-defined objects and by putting them directly on the screen of a computer. That is why the basic models are easily created. The SIMUL8 software is widely used in mainly for the process simulation mainly to model production/service processes (Greasley 2003; Ficova and Kuncová 2013; Omogbai, Salonitis 2016; Fousek, Kuncova and Fabry 2017; Kuncova, Skalova 2018).

The main 6 pre-defined objects are:

- Work Item – to define the main entity used in the model,
- Work Entry Point – for the entity inter-arrival times generation,
- Storage Bin – or a queue as a space where the entities might gather while waiting for a resource,
- Work Centre – for modelling any activity,
- Work Exit Pont – for the end of the modelled system, where an entity finishes its movement through the model,
- Resource - for modelling of limited capacities of the workers, material or means of production that are used during the activities.

Connectors are the last part that is due in all models to link the objects together, except of resources that are connected via the definition in a Work Centre (Concannon et al. 2007; Fousek, Kuncova, and Fabry 2017).

Although SIMUL8 itself has not the same possibilities as CAS-SIM, we tried to create a basic model of a market to simulate the Utterback model according to the conceptual model at Figure 3.

First, the environment and the set of nodes must be prepared. We model a fictive market driven by demand where the number of companies varies according to the market situation. The size of the companies is not directly modelled but the demand satisfaction is variable for each company and also for each set of companies. To be able to model more different companies in SIMUL8, the easier scheme is created where the 5 sets of companies are used as 5 nodes but behind each node max. 100 of companies may arise – it means max. number of companies on the market can be 500 (set of nodes representation).

As the market is fictive the demand is generated in fictive units and fictive inter-arrival times (exponential distribution was used) but the main idea was to simulate the whole life cycle of a new product. Four different periods were set for the demand increase, demand expansion, demand small decrease and demand big decrease.

After the demand generation the first set of companies (1-100) should try to satisfy it in a given time – the normal distribution for this activity is used with the average fifty times higher than the average inter-arrival time used for the demand generation, standard deviation is taken as 10% of the average activity time (in SIMUL8 named as "Average" distribution – Figure 4). The company is chosen randomly but the queuing time must not exceed one fifth of the processing time (Environment rules - Figure 5). When the demand is increasing and it is not possible by these 100 companies to handle it (the queuing time has reached the pre-set value and all companies are busy), new companies arise on the market – so the new set (1-100) is being used (node interactions representation) but as new-comers on the market they need higher time for the demand satisfaction. In the simulation model the demand (product) is sent to another queue or changes the queue after the long waiting time (as the customer looks for another, not busy, company). The whole model during the demand increase phase is can be seen on Figure 6.

This process continues if necessary based on the demand generated and based on the other companies' success in the demand fulfilment. This is taken as one of the possibilities of node interactions simulation when at the same time the new companies might need more time for the demand satisfaction (modelled via normal distribution with 10% higher average than for the previous set of companies).

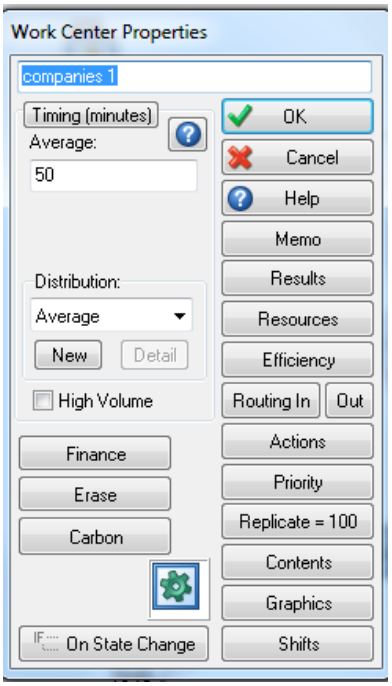


Figure 4: Definition of the work time for the first set of companies

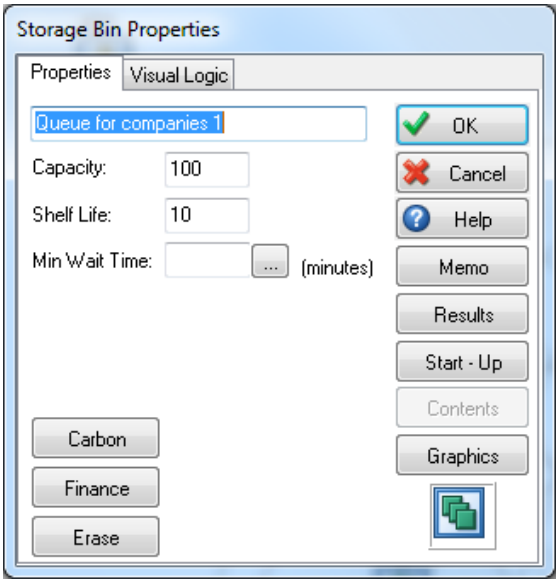


Figure 5: Queue settings for demand waiting for processing

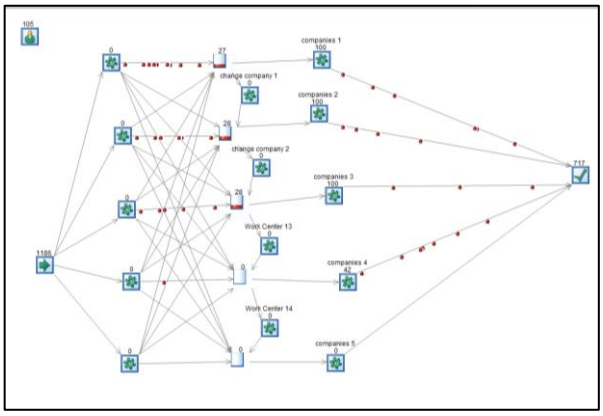


Figure 6: Simulation model in process – higher demand

Finally in each simulation time the number of active nodes (companies) is visible and in any time it is possible to see how productive each company is – so this might be takes as one example of the nodes evaluation (Figure 7).

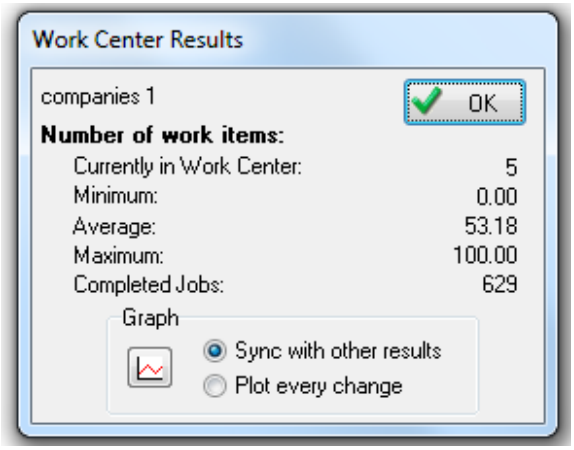


Figure 7: Simulation model results - number of companies (nodes) out of the first companies set needed for the demand satisfaction during the simulation time

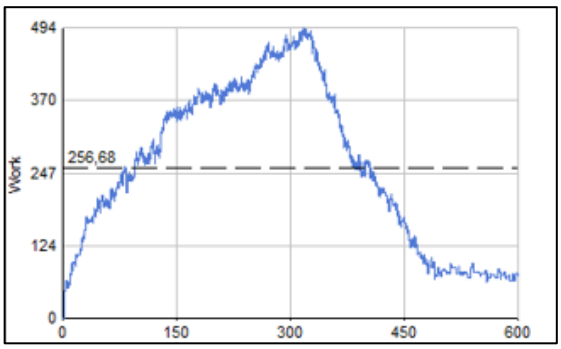


Figure 8: Simulation model results - changes in the number of companies on the market during the time

After the whole economic life cycle simulation (in out example about 600 time units) we might observe the typical curve of the number of companies on the market based on the time under the changing demand conditions (Figure 8).

7. CONCLUSION

The proposed simulation model of network economy development is based on the extension of the Utterback industrial growth model. The article proposes a conceptual model that captures the system being studied as a graphical network structure located in the market (market). Graphic structure nodes represent companies and edges represent interactions between companies. When tracking structure and behaviour dynamics, the problem is taken as a complex adaptive system. Company status is described by a fitness function value that is compared to the minimum level required for survival. The conceptual model is based on the simulation model architecture, which includes 4 coordinators: Environment, Evaluator, Visual Manager and Timekeeping, from which the Environment is the main and triggers the rest. A simulation algorithm is proposed which describes the development of the supply network. A simple model in the SIMUL8 software environment was created to demonstrate the given concept.

The proposed simulation model and algorithm are very flexible and allow modifications and generalizations when analysing a particular system. This tool seems appropriate for modelling the evolution of the network economy. The SIMUL8 itself might be used as the modelling environment but more other aspects should be added in real life situation which requires better and detailed specification of the market conditions using other probability distributions. Also programming some parts of the model via Visual logic in SIMUL8 might be useful.

ACKNOWLEDGMENTS

The paper was supported by the contribution of long term institutional support of research activities by the College of Polytechnics Jihlava and by the Internal Grant of the College of Polytechnics Jihlava No. 1170/4/199.

REFERENCES

Concannon, K., et al., 2007. Simulation Modeling with SIMUL8. Canada:Visual Thinking International.

Fiala P., 2009. Dynamic supply networks (in Czech). Praha: Professional Publishing.

Fiala P., 2016. Dynamic pricing and resource allocation in networks (in Czech). Praha: Professional Publishing.

Ficova, P. and Kuncova, M., 2013. Looking for the equilibrium of the shields production system via simulation model. Proceedings of the Conference Modeling and Applied Simulation (MAS), pp. 50-56. September 25-27, Athens (Greece).

Fousek, J., Kuncova, M. and Fabry, J., 2017. Discrete event simulation – production model in SIMUL8. Proceedings of the European Conference on Modelling and Simulation (ECMS), pp. 229–234, May 23-26, Budapest (Hungary).

Greasley, A., 2003. Simulation modelling for business. Innovative Business Textbooks. London: Ashgate.
Kuncova, M., Skalova, M., 2018. Discrete Event Simulation Applied to the Analysis of the Cash-desk Utilization in a Selected Shop of the Retail Chain. Proceedings of the Conference Modeling and Applied Simulation (MAS), pp. 76–82, September 17-19, Budapest (Hungary).
Omogbai, O. and Salonitis, K. 2016. Manufacturing system lean improvement design using discrete event simulation. Procedia CIRP – Conference on Manufacturing Systems 57, pp. 195 – 200.
Pathak S., Dilts D.M., 2004. CAS-SIM: Complex Adaptive Supply Network Simulator, A Scenario Analysis Tool For Analyzing Supply Networks. Proceedings of Production and Operations Management Society Conference, Cancun.
Pathak S., Dilts D.M., Biswas G., 2003. A Hybrid Simulator for Simulating Complex Adaptive Supply Chain Networks. Proceedings of Winter Simulation Conference, New Orleans.
Utterback J. (1994). Mastering the Dynamics of Innovation. Boston: HBS Press.
Utterback J., Suarez F., 1993. Innovation: Competition and Industry Structure. Research Policy 15, 285-305.

AUTHORS BIOGRAPHY

Petr Fiala: He holds the Full Professor degree from the University of Economics in Prague, Czech Republic, where he currently works as the Deputy Head of the Department of Econometrics. He is a member of other Czech universities. Professor Fiala is graduated from the University of Economics in Prague, in 1974, Charles University, Prague, in 1981, and Rochester Institute of Technology, USA, in 1996. His current research interests include modelling of production systems, supply chain management, project management, multi-criteria and group decision making. He is a member of professional societies for Operations Research and Multiple Criteria Decision Making and the Czech Republic representative in EURO (The Association of European Operational Research Societies). He has made research visits to many countries. He is the author of books, lecture notes, papers, and reports in quantitative models and methods in management and economics.

Martina Kuncova: She has got her degree at the University of Economics Prague, Czech Republic, at the branch of study Econometrics and Operational Research (1999). In 2009 she has finished her doctoral study at the University of West Bohemia in Pilsen, Czech Republic (Economics and Management). Since the year 2000 she has been working at the Department of Econometrics, University of Economics Prague, since 2007 also at the Department of Economic Studies of the College of Polytechnics Jihlava (Czech Republic). She is a member of the Czech Society of Operational Research, she participates in the solving of several grants. She is the co-author of four books and the author of many scientific papers and contributions at conferences. She is interested in the usage of the operational research, simulation methods and methods of multi-criteria decision in real life problems.

MODELLING OF SUPPLY NETWORK DESIGN

Renata Majovská[(a)], Petr Fiala[(b)]

[(a)] University of Finance and Administration, Prague
[(b)]University of Economics, Prague

[(a)] renata.majovska@mail.vsfs.cz, [(b)]pfiala@vse.cz

ABSTRACT
The paper is dedicated to proposed modelling approach for supply networks. The original structure of network systems can be modelled as complex adaptive systems and use agent-oriented simulation to demonstrate origin. The structure is clarified by expert opinion with use of DEMATEL method. The suitability of supply networks can be measured by multiple objectives, such as economic, environmental, social, and others. Traditional concepts of optimality focus on valuation of already given systems. We propose to use a methodology for optimal system design. As a methodology of optimal system design can be employed De Novo Multi-objective Linear Programming for reshaping feasible sets in linear systems.

Keywords: supply networks, simulation, dynamics, DEMATEL, De Novo optimisation

1. INTRODUCTION

Supply network management has generated a substantial amount of interest both by managers and researchers (see Tayur et al. 2000). The design of supply networks plays an important role in supply network management. The suitability of supply networks can be measured by multiple objectives, such as economic, environmental, social, technological, and others. Supply network structure and behaviour is changing dynamically (Majovská and Fiala 2015, 2016).

Modelling of supply network design is not a simple matter, it is necessary to combine random elements with expert evaluation and parameter adjustments in dynamic changes. The aim of the paper is to propose such a procedure. The overall idea of the proposed procedure lies in the interconnected steps from simulating the design of the supply network, through expert assessment of the essential parts and links between them, to optimizing the design parameters, including the necessary resources to create a supply network within a specified budget. The contribution of this paper is the proposed three-stage procedure for multi-criteria design of supply networks:

- Generation of potential elements of supply network.
- Expert evaluation and simplification of the network structure.
- Continuous reconfiguration and reshaping of systems boundaries.

In this paper, the individual stages are modified and adapted to the needs of the design of supply networks and connected in an integrated approach. The first stage is based on a dynamic generation of potential elements of supply networks. Structure and behaviour dynamics of network systems can be modelled as complex adaptive systems and use agent-oriented simulation to demonstrate origin, perturbation effects, and sensitivity with regard to initial conditions. The model is based on a supply network as a system consisting of an environment in which firms create interactions based on simple rules of conduct to meet global demand (Fiala and Kuncová 2019).

The second stage is devoted to expert evaluation and simplification of the supply network structure. The DEMATEL (Decision Making Trial and Evaluation Laboratory) method (Gabus and Fontela 1972) is used for this stage.

The third stage is oriented on a continuous reconfiguration and reshaping of systems boundaries. Multi-objective supply network design is formulated and solved by De Novo approach. Traditional concepts of optimality focus on valuation of already given systems. New concept of designing optimal systems was proposed (Zeleny 2010). Multi-objective linear programming (MOLP) is a model of optimizing a given system by multiple objectives. As a methodology of optimal system design can be employed De Novo programming for reshaping feasible sets in linear systems. The paper presents approaches for solving the Multi-objective De Novo linear programming (MODNLP) problem for design of multi-objective supply networks. The approach is based on reformulation of MOLP problem by given prices of resources and the given budget. Searching for a better portfolio of resources leads to a continuous reconfiguration and reshaping of systems boundaries.

Technological innovations bring improvements to the desired objectives and the better utilization of available resources. These changes can lead to beyond tradeoff-free solutions. The de Novo approach was adapted for supply network design.

2. GENERATING SUPPLY NETWORKS

Supply network is defined as a system of clusters with:

- suppliers,
- manufacturers,
- distributors,
- retailers,
- customers,

where

- material,
- financial
- information,
- decision

flows connect participants in both directions (see Fiala 2005). A supply network is a complex and dynamic supply and demand network of agents, activities, resources, technology and information involved in moving a product or service from supplier to customer. Most supply networks are composed of independent units with individual preferences. Each unit will attempt to optimize his own preference. Behaviour that is locally efficient can be inefficient from a global point of view. An information asymmetry is a source of inefficiency in supply networks. The so-called bullwhip effect, describing growing variation upstream in a supply network, is probably the most famous demonstration of inefficiency and system dynamics in supply networks. Information sharing is a very important issue for coordinating actions of units in networks.

The dynamic generation of supply networks consists in generating potential network members and inter-relationships. To understand the growth and development of dynamics, it is necessary to monitor the time-dependent behaviour of the model. Simulation is a frequently used methodology for analysing the time-varying properties of a system. Agent-oriented simulation takes place in a system consisting of an environment in which agents (nodes) create interactions (edges) based on simple rules of conduct to meet global demand. Stochastic environmental parameters, describing market conditions and demand, a node-based decision-making scheme, so-called fitness functions modelling the strength of companies, all influence the dynamics of structure and behaviour of the developing supply network (Fiala and Kuncová 2019).

As new industries emerge, supply networks grow, creating new relationships between firms that cooperate to meet demand. New start-ups are involved in supply networks. Some companies significantly expanding capacity, while others weaken. An alternative scenario may be the fact that no companies become dominant and the market is relatively evenly distributed among participating firms. In a comprehensive adaptive supply network model, two aspects are considered:

- number of firms,
- size of firms.

The next step is to assess the significance of elements and relationships between them in the supply network.

3. DEMATEL METHOD

Expert evaluation and simplification of the supply network structure is done using the method DEMATEL. The DEMATEL method can be summarized in the following steps (Gabus and Fontela 1972):

Step 1. *Find the initial direct relation matrix.*

Suppose we have *m* experts in this study and *n* elements to consider. Each expert is asked to indicate the degree to which he believes an element *i* affects an element *j*. These pairwise comparisons between any two elements are denoted by a_{ij} and are given an integer score ranging from 0, 1, 2, 3, and 4, representing:

0 no influence,
1 low influence,
2 medium influence,
3 high influence,
4 very high influence.

The scores by each expert will give us a (*n*, *n*) non-negative answer matrix $\mathbf{X}^k = [\, x_{ij}^k \,]$, with $1 \le k \le m$.

The diagonal elements of each answer matrix $\mathbf{X}^k$ are all set to zero. We can then compute the (*n*, *n*) average matrix **A** for all expert opinions by averaging the *m* experts' scores as follows:

$$a_{ij} = \frac{1}{m} \sum_{k=1}^{m} x_{ij}^k \qquad (1)$$

The average matrix $\mathbf{A}=[\, a_{ij} \,]$ is called the initial direct relation matrix. Matrix **A** shows the initial direct effects that an element exerts on and receives from other elements. Furthermore, we can map out the causal effect between each pair of elements in a system by drawing an influence map. DEMATEL can convert the structural relations among the elements of a system into an intelligible map of the system.

Step 2. *Calculate the normalized initial direct-relation matrix.*

The normalized initial direct-relation matrix **D** is obtained by normalizing the initial direct-relation matrix ***A*** in the following way:

Let

$$s = \max\left(\max_{1 \le i \le n} \sum_{j=1}^{n} a_{ij}, \max_{1 \le j \le n} \sum_{i=1}^{n} a_{ij} \right) \qquad (2)$$

Then

$$\mathbf{D} = \frac{1}{s} \mathbf{A} \qquad (3)$$

Since the sum of each row j of matrix **A** represents the total direct effects that element i gives to the other elements, $\max_{1\le i\le n}\sum_{j=1}^{n} a_{ij}$ represents the total direct effects of the element with the most direct effects on others. Likewise, since the sum of each column i of matrix ***A*** represents the total direct effects received by element i, $\max_{1\le j\le n}\sum_{i=1}^{n} a_{ij}$ represents the total direct effects received of the element that receives the most direct effects from others. The positive scalar s takes the lesser of the two as the upper bound, and the matrix **D** is obtained by dividing each element of **A** by the scalar s. Note that each element d_{ij} of matrix **D** is between zero and one.

Step 3. *Compute the total relation matrix.*
A continuous decrease of the indirect effects of problems along the powers of matrix **D** guarantees convergent solutions to the matrix inversion similar to an absorbing Markov chain matrix. Note that $\lim_{k\to\infty} \mathbf{D}^k = \mathbf{0}$ and $\lim_{k\to\infty}(\mathbf{I} + \mathbf{D} + \mathbf{D^2} + \cdots + \mathbf{D}^k) = (\mathbf{I} - \mathbf{D})^{-1}$, where $\mathbf{0}$ is the (n, n) null matrix and **I** is the (n, n) identity matrix. The total relation matrix **T** is an (n, n) matrix

$$\mathbf{T} = [t_{ij}] \quad i, j = 1, 2, \ldots, n,$$

and is defined as follow:

$$\mathbf{T} = \mathbf{D} + \mathbf{D}^2 + \ldots + \mathbf{D}^k = \mathbf{D}(\mathbf{I} + \mathbf{D} + \mathbf{D^2} + \cdots + \mathbf{D^{k-1}}) = \mathbf{D}(\mathbf{I} - \mathbf{D})^{-1}, \text{as } k \to \infty. \quad (4)$$

Vectors **r** and **c** are defined representing the sum of rows and sum of columns of the total relation matrix **T** as follows:

$$\mathbf{r} = (r_i) \quad (5)$$

where r_i be the sum of i-th row in matrix **T**. Then r_i shows the total effects, both direct and indirect, given by element i to the other elements, and

$$\mathbf{c} = (c_j) \quad (6)$$

where c_j denotes the sum of j-th column in matrix **T**. Then c_j shows the total effects, both direct and indirect, received by element j from the other elements.

Thus when $j = i$, the sum $(r_i + c_i)$ gives an index representing the total effects both given and received by element i. In other words, $(r_i + c_i)$ shows the degree of importance that element i plays in the system. In addition, the difference $(r_i - c_i)$ shows the net effect that element i contributes to the system. When $(r_i - c_i)$ is positive, element i is a net causer, and when $(r_i - c_i)$ is negative, element i is a net receiver (Tzeng et al. 2007).

Step 4. *Set a threshold value and obtain the impact-relations-map.*
In order to explain the structural relation among the elements while keeping the complexity of the system to a manageable level, it is necessary to set a threshold value p to filter out some negligible effects in matrix **T**. While each element of matrix **T** provides information on how one element affects another, the decision-maker must set a threshold value in order to reduce the complexity of the structural relation model implicit in matrix **T**. Only some elements, whose effect in matrix **T** is greater than the threshold value, should be chosen and shown in an impact-relations-map (Tzeng et al. 2007).

4. MULTI-OBJECTIVE SUPPLY NETWORKS

In the next part, we formulate a multi-objective supply network design problem. The mathematical program determines the ideal locations for each facility and allocates the activity at each facility such that the multiple objectives are considered and the constraints of meeting the customer demand and the facility capacity are satisfied. The presented model of a supply network consists of 4 layers with m suppliers, S_1, S_2, ... S_m, n potential producers, P_1, P_2, ... P_n, p potential distributors, D_1, D_2, ... D_p, and r customers, C_1, C_2, ... C_n. The following notation is used:
a_i = annual supply capacity of supplier i, b_j = annual potential capacity of producer j,
w_k = annual potential capacity of distributor k, d_l = annual demand - customer l,
f_j^P = fixed cost of potential producer j, f_k^D = fixed cost of potential distributor k,
c_{ij}^S = unit transportation cost from S_i to P_j, c_{jk}^P = unit transportation cost from P_j to D_k,
c_{kl}^D = unit transportation cost from D_k to C_l, e_{ij}^S = unit pollution from S_i to P_j,
e_{jk}^P = unit pollution from P_j to D_k, e_{kl}^D = unit environmental pollution from D_k to C_l,
x_{ij}^S = number of units transported from S_i to P_j, x_{jk}^P = number of units transported from P_j to D_k, x_{kl}^D = number of units transported from D_k to C_l,
y_j^P = bivalent variable for build-up of fixed capacity of producer j,
y_k^D = bivalent variable for build-up of fixed capacity of distributor k.
Using the above notations the problem can be formulated as follows:
The model has two objectives. The first one expresses minimizing of total costs. The second one expresses minimizing of total environmental pollution.

Min

$$z_1 = \sum_{j=1}^{n} f_j^P y_j^P + \sum_{k=1}^{p} f_k^D y_k^D + \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}^S x_{ij}^S + \sum_{j=1}^{n}\sum_{k=1}^{p} c_{jk}^P x_{jk}^P + \sum_{k=1}^{p}\sum_{l=1}^{r} c_{kl}^D x_{kl}^D$$

Min

$$z_2 = \sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}^S x_{ij}^S + \sum_{j=1}^{n}\sum_{k=1}^{p} e_{jk}^P x_{jk}^P + \sum_{k=1}^{p}\sum_{l=1}^{r} e_{kl}^D x_{kl}^D$$

Subject to the following constraints:

the amount sent from the supplier to producers cannot exceed the capacity

$$\sum_{j=1}^{n} x_{ij} \le a_i, \quad i = 1, 2, ..., m,$$

the amount produced by the producer cannot exceed the producer capacity

$$\sum_{k=1}^{p} x_{jk} \le b_j y_j, \quad j = 1, 2, ..., n,$$

the amount shipped from the distributor should not exceed the distributor capacity

$$\sum_{l=1}^{r} x_{kl} \le w_k y_k, \quad k = 1, 2, ..., p,$$

the amount shipped to the customer must equal the customer demand

$$\sum_{k=1}^{p} x_{kl} = d_l, \quad l = 1, 2, ..., r,$$

the amount shipped out of producers cannot exceed units received from suppliers

$$\sum_{i=1}^{m} x_{ij} - \sum_{k=1}^{p} x_{jk} \ge 0, \quad j = 1, 2, ..., n,$$

the amount shipped out of distributors cannot exceed quantity received from producers

$$\sum_{j=1}^{n} x_{jk} - \sum_{l=1}^{r} x_{kl} \ge 0, \quad k = 1, 2, ..., p,$$

binary and non-negativity constraints

$$y_j, y_k \in \{0,1\},$$

$$x_{ij}, x_{jk}, x_{kl} \ge 0, \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n, \quad k = 1, 2, ..., p, \quad l = 1, 2, ..., r.$$

The formulated model is a multi-objective linear programming problem (MOLP). The problem can be solved by some MOLP methods.

5. OPTIMIZING GIVEN SYSTEMS

Multi-objective linear programming (MOLP) is a model of optimizing a given system by multiple objectives. In MOLP problems it is usually impossible to optimize all objectives together in a given system. Trade-off means that one cannot increase the level of satisfaction for an objective without decreasing this for another objective. Multi-objective linear programming (MOLP) problem can be described as follows

$$\begin{aligned} &\text{“Max”} \quad z = Cx \\ &\text{s.t.} \quad Ax \le b, \; x \ge 0 \end{aligned} \tag{7}$$

where C is a (k, n) – matrix of objective coefficients, A is a (m, n) – matrix of structural coefficients, b is an m-vector of known resource restrictions, x is an n-vector of decision variables. In MOLP problems it is usually impossible to optimize all objectives in a given system. For multi-objective programming problems the concept of non-dominated solutions is used (see for example Steuer 1986). A compromise solution is selected from the set of non-dominated solutions. There are proposed many methods. Most of the methods are based on trade-offs. The next part is devoted to the trade-off free approach.

6. DESIGNING OPTIMAL SYSTEMS

Multi-objective De Novo linear programming (MODNLP) is a problem for designing an optimal system by reshaping the feasible set. By given prices of resources and the given budget the MOLP problem (7) is reformulated in the MODNLP problem (8)

$$\begin{aligned} &\text{“Max”} \quad z = Cx \\ &\text{s.t.} \quad Ax - b \le 0, \; pb \le B, \; x \ge 0 \end{aligned} \tag{8}$$

where b is an m-vector of unknown resource restrictions, p is an m-vector of resource prices, and B is the given total available budget.

From (8) follows $pAx \le pb \le B.$

Defining an n-vector of unit costs $v = pA$ we can rewrite the problem (8) as

$$\begin{aligned} &\text{“Max”} \quad z = Cx \\ &\text{s.t.} \quad vx \le B, \; x \ge 0 \end{aligned} \tag{9}$$

Solving single objective problems

$$\begin{aligned} &\text{Max} \quad z^i = c^i x, \quad i = 1,2,\ldots,k \\ &\text{s.t.} \quad vx \le B, \; x \ge 0 \end{aligned} \tag{10}$$

z^* is a k – vector of objective values for the ideal system with respect to B.

The problems (10) are continuous “knapsack” problems, the solutions are

$$x_j^i = \begin{cases} 0, & j \ne j_i \\ B/v_{j_i}, & j = j_i \end{cases},$$

where $j_i \in \left\{ j \in (1,\ldots,n) \middle| \max_j (c_j^i / v_j) \right\}$.

The meta-optimum problem can be formulated as follows

$$\begin{aligned} &\text{Min} \quad f = vx \\ &\text{s.t.} \quad Cx \ge z^*, \; x \ge 0 \end{aligned} \tag{11}$$

Solving the problem (11) provides solution: x*, B* = vx*, b* = Ax*.

The value B^* identifies the minimum budget to achieve z^* through solutions x^* and b^*.

The given budget level $B \le B^*$. The optimum–path ratio for achieving the best performance for a given budget B is defined as

$$r_1 = \frac{B}{B^*}$$

The optimum-path ratio provides an effective and fast tool for the efficient optimal redesign of large-scale linear systems. Optimal system design for the budget B:

$x = r_1 x^*$, $b = r_1 b^*$, $z = r_1 z^*$.

7. DE NOVO APPROACH FOR MULTI-OBJECTIVE SUPPLY NETWORKS

The De Novo approach can be useful in the design of the multi-objective supply network. Only a partial relaxation of constraints is adopted. Producer and distributor capacities are relaxed. Unit costs for capacity build-up are computed:

$p_j^P = \frac{f_j^P}{b_j}$ = cost of unit capacity of potential producer *j*,

$p_k^D = \frac{f_k^D}{w_k}$ = cost of unit capacity of potential distributor *k*.

Variables for build-up capacities are introduced:

u_j^P = variable for flexible capacity of producer *j*,

u_k^D = variable for flexible capacity of producer *k*.

The constraints for non-exceeding producer and distributor fixed capacities are replaced by the flexible capacity constraints and the budget constraint:

$$\sum_{k=1}^{p} x_{jk} - u_j^P \le 0, \quad j = 1, 2, ..., n,$$

$$\sum_{l=1}^{r} x_{kl} - u_k^D \le 0, \quad k = 1, 2, ..., p,$$

$$\sum_{j=1}^{n} p_j^P u_j^P + \sum_{k=1}^{p} p_k^D u_k^D \le B.$$

The multi-objective optimization can be then seen as a dynamic process. Technological innovations bring improvements to the desired objectives and the better utilization of available resources. The technological innovation matrix $T = (t_{ij})$ is introduced. The elements in the structural matrix A should be reduced by a technological progress. The problem (8) is reformulated in to the innovation MODNLP problem (12)

$$\text{“Max”} \quad = Cx$$
$$\text{s.t.} \quad TAx - b \le 0 \,, pb \le B, x \ge 0 \qquad (12)$$

De Novo approach provides a better solution in multiple objectives and also with lower budget because of flexible capacity constraints. The capacity of supply network members has been optimized with regard to flows in the supply network and to budget.

8. CASE STUDY

We tested the De Novo approach on a case study. A supply network is proposed with 3 potential suppliers, 3 potential manufacturers, 3 potential distributors, 3 customers. The network is evaluated according to 2 criteria, the first criterion is aimed at minimizing total costs and the second one at minimizing overall environmental pollution. Inputs for the model are as follows:

Capacities $a_i = 100$, $i = 1, 2, 3$; $b_j = 100$, $j = 1, 2, 3$; $w_k = 100$, $k = 1, 2, 3$; $d_l = 50$, $l = 1, 2, 3$.
Fixed costs $f_1^P = 110$, $f_2^P = 100$, $f_3^P = 120$, $f_1^D = 120$, $f_2^D = 110$, $f_3^D = 150$.
Unit transport costs and unit pollution are shown in the Table 1 and Table 2.

Table 1: Unit transport costs

c_{ij}^S	1	2	3	c_{jk}^P	1	2	3	c_{kl}^D	1	2	3
1	5	10	6	1	7	5	9	1	8	3	10
2	8	9	7	2	6	8	4	2	6	5	4
3	3	6	8	3	5	7	9	3	7	3	5

Table 2: Unit pollution

e_{ij}^S	1	2	3	e_{jk}^P	1	2	3	e_{kl}^D	1	2	3
1	4	3	8	1	8	7	9	1	8	6	2
2	8	9	2	2	6	8	4	2	8	9	8
3	7	6	8	3	4	7	9	3	5	3	5

This model was solved by different approaches. The first two approaches minimize each criterion separately. The compromise solution is calculated by the traditional STEM interactive approach for multi-criterion tasks and the De Novo approach was used. The following are non-zero values of the variables that express the number of units of product shipped between each supply network layer. These values are given for each problem-solving approach.

Min z_1: $x_{13}^S = 50$, $x_{31}^S = 100$, $x_{12}^P = 100$, $x_{31}^P = 50$, $x_{12}^D = 50$, $x_{21}^D = 50$, $x_{23}^D = 50$.
Min z_2: $x_{12}^S = 100$, $x_{23}^S = 50$, $x_{23}^P = 100$, $x_{31}^P = 50$, $x_{13}^D = 50$, $x_{31}^D = 50$, $x_{32}^D = 50$.
STEM: $x_{11}^S = 58.13$, $x_{23}^S = 91.87$, $x_{12}^P = 58.13$, $x_{31}^P = 91.87$, $x_{12}^D = 46.87$, $x_{13}^D = 45$, $x_{21}^D = 50$, $x_{22}^D = 3.12$, $x_{23}^D = 50$.
De Novo: $x_{23}^S = 62.86$, $x_{32}^S = 87.14$, $x_{21}^P = 10$, $x_{23}^P = 77.14$, $x_{31}^P = 62.86$, $x_{12}^D = 50$, $x_{13}^D = 22.86$, $x_{31}^D = 50$, $x_{33}^D = 27.14$.

The criteria values z_1a z_2 and the budget B are compared according to these solutions. De Novo solution is better in all values than the STEM solution. De Novo approach provides better solutions on both criteria and also with a lower budget due to flexible capacity constraints. The capacity of supply network members have been optimized for flows in the supply network and budget. The comparison of results are shown in Table 3.

Table 3: Comparison of solution results

	Min z_1	Min z_2	STEM	De Novo
z_1	2460	3490	3070	3000
z_2	3100	1800	2030	2000
B	460	490	460	365.71

9. CONCLUSIONS

The proposed methodology of supply network design modelling consists of three stages. In the first stage, the network system of potential members and their interconnections is simulated. The second stage is

devoted to expert evaluation of the proposed network system by the DEMATEL method. In the third stage, De Novo approach was applied for multi-objective supply network design problem and provides better solution than traditional approaches applied on fixed constraints. The design problem was formulated as MOLP problem. The economic and environmental objectives were used in the model but multiple objectives can be used in general. Technological innovations bring improvements to the desired objectives and the better utilization of available resources. These changes can lead to beyond tradeoff-free solutions. The combination of these three stages and their repetition over time provides an efficient and flexible tool for modeling supply network design. The approach can be enriched with other tools such as game theory (Fiala 2016).

ACKNOWLEDGMENTS

The paper is a result of institutional research project no. 7429/2018/08 "Analysis of ICT startups" supported by University of Finance and Administration, Prague and Grant No. IGA F4/66/2019, Faculty of Informatics and Statistics, University of Economics, Prague.

REFERENCES

Fiala, P., 2005. Information sharing in supply chains. Omega, 33 (4), 419-423.

Fiala, P., 2016. Profit allocation games in supply chains. Central European Journal of Operations Research, 24 (2), 267-281.

Fiala, P., Kuncová, M., 2019. Simulation model of supply networks development. Proceedings of International Conference on Modelling and Applied Simulation. September 18-20, Lisbon (Portugal).

Mathematical Methods in Economics, pp. 86–91. September 12-14, Jindřichův Hradec (Czech Republic).

Fiala, P., Majovská, R., 2018. Modeling of supply network coordination. Proceedings of Mathematical Methods in Economics, pp. 86–91. September 12-14, Jindřichův Hradec (Czech Republic).

Fiala, P., Majovská, R., 2018. Design of Supply Networks by De Novo Method (in Czech). Trendy v podnikání, 8 (3), 32–38.

Gabus, A., Fontela, E., 1972. World problems, an invitation to further thought within the framework of DEMATEL. Geneva: Battelle Geneva Research Centre.

Majovská, R., Fiala, P., 2015. Dynamic analysis of multi-criteria network systems. Proceedings of Mathematical Methods in Economics, pp. 484–490. September 9-11, Cheb (Czech Republic).

Majovská, R., Fiala, P., 2016. Modelling of dynamic network systems. Proceedings of the Quantitative Methods in Economics, pp. 236–240. May 25-27, Vrátna (Slovak Republic).

Steuer, R. E. 1986. Multiple Criteria Optimization: Theory, Computation and Application. New York: Wiley.

Tayur, S., Ganeshan, R., Magazine, M., 2000. Quantitative models for supply chain management. Boston: Kluwer.

Tzeng, G. H., Chiang, C. H., Li, C. W., 2007. Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. Expert Systems with Applications 32, 1028–1044.

Zelený, M., 2010. Multiobjective Optimization, Systems Design and De Novo Programming. In: C. Zopounidis, P. M. Pardalos, ed. Handbook of Multicriteria Analysis. Berlin: Springer, 243-262.

AUTHORS BIOGRAPHY

Renata Majovská: She earned her degree in Mathematics and Physics at the Technical University in Ostrava, Czech Republic. She gained her PhD. at the Constantine the Philosopher University in Nitra, Slovakia in 2008. She is the Assistant Professor at the University of Finance and Administration in Prague, Czech Republic. She is a member of the Czech Society of Operational Research. Dr. Majovská has participated in the solving of several grants and is a co-author of several books and author of many scientific and conference papers in modelling of production systems, supply chain management, multi-criteria and group decision making. She has prepared many electronic university textbooks.

Petr Fiala: He holds the Full Professor degree from the University of Economics in Prague, Czech Republic, where he currently works as the Deputy Head of the Department of Econometrics. He is a member of other Czech universities. Professor Fiala is graduated from the University of Economics in Prague, in 1974, Charles University, Prague, in 1981, and Rochester Institute of Technology, USA, in 1996. His current research interests include modelling of production systems, supply chain management, project management, multi-criteria and group decision making. He is a member of professional societies for Operations Research and Multiple Criteria Decision Making and the Czech Republic representative in EURO (The Association of European Operational Research Societies). He has made research visits to many countries. He is the author of books, lecture notes, papers, and reports in quantitative models and methods in management and economics.

OPTIMAL HYBRID DRIVE SYSTEM ARCHITECTURE EXPLORATION CONSIDERING PERFORMANCE INDEX OF 48V MILD HEV

Yonghyeok Ji[(a)], Taeho Park[(b), (c)], Hyeongcheol Lee[*(d)]

[(a),(b)]Department of Electrical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea
[(c)]Green Car Power System R&D Division, Korea Automotive Technology Institute, 303 Pungse-ro, Pungse-myeon, Dongnam-gu, Cheonan-si, Chungnam 31214, Korea
[(d)]Department of Electrical and Biomedical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea

[(a)]youg0839@hanyang.ac.kr, [(b), (c)]koreapow@hanyang.ac.kr, [(d)]hclee@hanyang.ac.kr
**Corresponding Author hclee@hanyang.ac.kr*

ABSTRACT
The 48V hybrid system has mostly adopted parallel hybrid system architecture. In the parallel hybrid system, various architecture can be derived depending on the location of the motor. In this paper, we explored a hybrid system architecture considering one or two motors and 48V electric supercharger and derived the optimal architecture by comparing the performance of each architecture. Performance of the hybrid system is mostly evaluated as fuel economy. However, since the hybrid system has increasingly been applied to various types of vehicles with different purpose of the operation, another performance index for evaluating a hybrid system is needed. Therefore, in this paper, we introduced an additional performance index to evaluate the hybrid electric drive system and used it to derive the optimal architecture of the hybrid electric drive system. We used *Dynamic programming (DP)* to evaluate each architecture and *DP* simulation was performed in the Matlab environment.

Keywords: 48V mild hybrid system, HEV architecture, *Dynamic programming*, Performance index

1. INTRODUCTION

As fuel economy and emission gas regulations are strengthened globally, environmentally friendly vehicles of various types (BEV, FCEV, (P)HEV, etc.) are being introduced to the market. However, due to the problem of the high-cost electric drive system or charging infrastructure, it is still difficult for vehicles to replace the demand for existing internal combustion engine vehicles. According to this trend, a mild hybrid system based on a 48V power supply has recently attracted attention as a new alternative and many kinds of research about 48V mild hybrid system are being conducted. (Malte Kuypers 2014; Mark Schudeleit and Christian Sieg 2015; Anthony Rick and Brain Sisk 2015; Andreas Baumgardt and Dieter Gerling 2015a,b; Anthony Rick and Brain Sisk 2015; Zifan Liu and Andrej Ivanco 2016; Junyong Park and Taeho Park 2017)

There are various hybrid drive system architectures according to the arrangement of the hybrid drive system component such as the engine, the motor, and the gearbox. 48V hybrid drive systems adopt parallel hybrid architecture mostly, and parallel hybrid architectures are generally divided into P0~P4 according to the location of the motor (Figure 1). And many studies have been done on the CO2 reduction effect and cost efficiency for each configuration. (Dr. Ing. Olivier COPPIN 2016; Ran Bao and Victor Avila 2017; Thomas Eckenfels and Florian Kolb 2018)

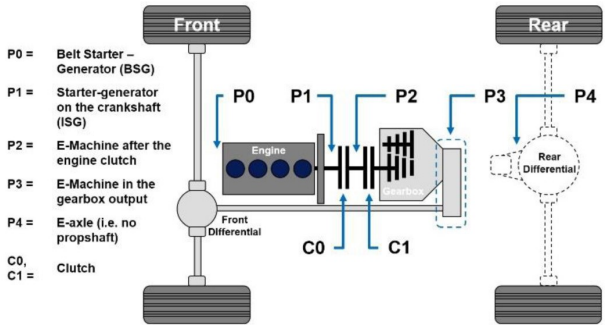


Figure 1: P0~P4 Architecture

However recently, there are the cases composing a 48V hybrid system including not only one motor but also two motor or a 48V electric supercharger, and there is still a lack of research on which architecture is optimal when these components are considered in 48V hybrid system configuration.

The performance of a hybrid system is evaluated with fuel economy mostly. But a hybrid system has been being adapted to various vehicle segments, and since a consumer's expectation is different by vehicle segments, it is needed that evaluating vehicle with performance

other than fuel economy. As related research, Kapadia (2017) compared P2 parallel hybrid architecture and input split hybrid architecture in terms of not only fuel economy but also drivability, launch, power, towing performance and packaging, etc.
Therefore, in this paper, we derived an optimal 48V hybrid system architecture according to fuel economy and performance index when one or two 48V motor and 48V electric supercharger are considered. To do this, we explored 48V hybrid system architecture manually when one or two 48V motors, 48V electric supercharger are considered, and we derived an optimal architecture through *Dynamic programming*. In this case, the cost function of *Dynamic programming* was modified to derive optimal architecture for fuel economy and performance index. Here, the performance index is the index which can evaluate hybrid system performance other than fuel economy, and in this paper, we defined an electric auxiliary load assist ability and drive power reserve as a performance index.

2. EXPLORATION TARGET HYBRID DRIVE ARCHITECTURE

When considering one motor, a 48V hybrid system architecture which can be derived is like Figure 1. In this case, since P0 and P1 architectures have lower fuel economy than P2, P3, and P4 generally, we select P2, P3, and P4 architecture as exploration target hybrid drive architecture having one motor.
And we select the exploration target hybrid drive architecture having two motors like Figure 2. In the figure, the P0D architecture is the architecture that can operate mechanical auxiliary load (ex. water pump, air-conditioner compressor) independently by separating the mechanical auxiliary load from the engine according to the situation. And the motor location is the same as P0 architecture. Since P0D has little higher fuel economy than P0 generally, which is not suggested in this paper, we select P0D architecture instead of P0 architecture for exploration.
Therefore, we select 10 exploration target 48V hybrid drive architecture like below according to motor location and the number of motors.

1. Using one motor: P2, P3, P4
2. Using two motor: P0D +P2, P0D+P3, P0D+P4
   P1 +P2, P1+P3, P1+P4
   P2+P4

But, above 10 architecture doesn't consider a 48V electric supercharger. Therefore finally, we select a total 20 exploration target 48V hybrid drive system architecture by adding 10 architecture having a 48V electric supercharger.
Table 1 below is the specification of the hybrid drive system main components. As we can see in the table, the motor power when using two motors is half of the power when using one motor.

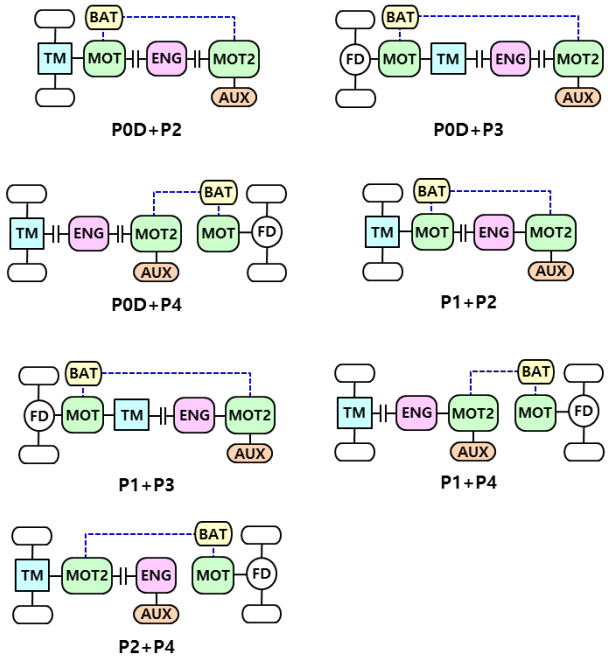


Figure 2: 48V Hybrid Drive System Architecture with Two Motor

Table 1: Specification of Vehicle Component

Component	Specification
Engine	1500 [cc] Gasoline Engine
Motor	10 [kW] PMSM (when using one motor) 5 [kW] PMSM (when using two motor)
Electric Supercharger	5 [kW] PMSM
Battery	48 [V] / 20 [Ah] Lithium-Ion

3. MODEL FOR DYNAMIC PROGRAMMING

We used *Dynamic Programming (DP)* to compare each exploration target hybrid drive system architecture. *DP* is a method of global optimization strategies that find optimal solutions to optimal control problems based on Bellman's principle of optimality. For the hybrid drive system, *DP* is mainly used to analyze the optimal fuel economy results by investigating all possible paths of the vehicle system in a given driving cycle in advance.
Therefore, the *DP* results for any hybrid system architecture can be said to be the optimal performance of the architecture. So, we carry out *DP* simulation about each exploration target hybrid drive architecture and derives optimal architecture by comparing *DP* results.

3.1. Dynamic programming with dpm.m function

In this paper, we use Matlab function dpm.m for *DP*. dpm.m is a *DP* algorithm introduced by Olle Sundström and Lino Guzzella (2009) to solve the general optimization problem. To solve the optimal control problem using the dpm.m function, the user must define the Main.m function and the Model function.m function. The Main.m function sets the range and grid of control

input, state for *DP* and executes dpm.m. Model function.m uses the given inputs to calculate the cost and the state of the next step and returns these values to dpm.m. In this paper, the model function is the vehicle model that calculates the fuel consumption and the SOC variation according to the power distribution ratio, the electric supercharger speed, and transmission gear stage.

3.2. Main.m function setting

As mentioned in Section 3.1, we need the Main.m function and the Model function.m to use dpm.m. This section introduces the settings for the Main.m function. First, we define the control input as below.

1. u_1 : Power distribution ratio for the first motor
2. u_2 : 48V electric supercharger speed
3. u_3 : Gear shift command
4. u_4 : Decoupler on/off command
5. u_5 : Power distribution ratio for the second motor

The decoupler on/off command u_4 is applied to the architecture considering P0D motor, the Power distribution ratio for the second motor u_5 is applied to the architecture considering two motors. The activated control inputs according to architecture is like Table 2.

Table 2: Control Input Setting according to Hybrid Drive System Architecture (O: Use, -: Not use)

Architecture	u_1	u_2	u_3	u_4	u_5
P2	O	O	O	-	-
P3	O	O	O	-	-
P4	O	O	O	-	-
P0D+P2	O	O	O	O	O
P0D+P3	O	O	O	O	O
P0D+P4	O	O	O	O	O
P1+P2	O	O	O	-	O
P1+P3	O	O	O	-	O
P1+P4	O	O	O	-	O
P2+P4	O	O	O	-	O

The mathematical equation of each control input is defined as follows.

First, u_1 and u_5 are defined like Table 3. In the equations below, $\gamma_{ply}, \gamma_{TM}$ is the gear ratio of engine pulley and transmission respectively, $T_{m,i}$ is the motor torque, $T_{tot,i}$ is the vehicle total desired torque (i=2 for P2, 3 for P3, 0D3 for P0D+P3 etc.). And $T_{tot\_f,j}$ is the value of vehicle total desired torque minus P4 motor torque in the architecture having P4 motor. (j=0D4, 14, 24). As we can see in the table, u_1 determines the power distribution of motor which is located relatively far from the engine, u_5 determines the power distribution of the motor which is located relatively close to the engine.

Table 3: Equation of Control Input u_1, u_5

Architecture	u_1	u_5
P2	$T_{m,2}/T_{tot,2}$	-
P3	$T_{m,3}/T_{tot,3}$	-
P4	$T_{m,4}/T_{tot,4}$	-
P0D+P2	$T_{m,2}/T_{tot,0D2}$	$\gamma_{ply}T_{m,0D}/T_{tot,0D2}$
P0D+P3	$T_{m,3}/T_{tot,0D3}$	$\gamma_{TM}\gamma_{ply}T_{m,0D}/T_{tot,0D3}$
P0D+P4	$T_{m,4}/T_{tot,0D4}$	$\gamma_{TM}\gamma_{ply}T_{m,0D}/T_{tot\_f,0D4}$
P1+P2	$T_{m,2}/T_{tot,12}$	$T_{m,1}/T_{tot,12}$
P1+P3	$T_{m,3}/T_{tot,13}$	$\gamma_{TM}T_{m,1}/T_{tot,13}$
P1+P4	$T_{m,4}/T_{tot,14}$	$\gamma_{TM}T_{m,1}/T_{tot\_f,14}$
P2+P4	$T_{m,4}/T_{tot,24}$	$\gamma_{TM}T_{m,2}/T_{tot\_f,24}$

u_3, u_4 are defined like Equation (1), (2) respectively. In Equation (2), Decoupler open means the case that mechanical auxiliary load is separated from an engine, and Decoupler close means the case that mechanical auxiliary load is connected to an engine.

$$u_3 = \begin{cases} 1 & (Upshift) \\ 0 & (No\ shift) \\ -1 & (Downshift) \end{cases} \quad (1)$$

$$u_4 = \begin{cases} 1 & (Decoupler\ open) \\ 0 & (Decoupler\ close) \end{cases} \quad (2)$$

The range and grid of each control input are like Table 4.

Table 4: Range and Grid of Control Input

Control Input	Range	Grid points	Unit
u_1	-1 ~ 1	21	-
u_2	0 ~ 143,000	21	RPM
u_3	-1 ~ 1	3	-
u_4	0 ~ 1	2	-
u_5	-1 ~ 1	21	-

We define states as 48V battery SOC, an engine on/off state, and gear stage of transmission. The range and grid of each state are like Table 5.

Table 5: Range and Grid of State

Control Input	Range	Grid points
SOC	0.3 ~ 0.9	21
Engine On/Off	0 ~ 1	2
Gear Stage	1 ~ 6	6

3.3. Model function.m setting

This section introduces the Model function.m for *DP*. As mentioned above, the model function in this paper is a vehicle model, and the model used in this study is basically the same as the model which was used in the study last year (Ji Y. and Park J. 2018). Therefore, in this paper, we only describe the values which are defined differently from the last year's study depending on the exploration target hybrid system architecture. The values that are defined differently by architecture are vehicle total desired torque, motor torque, cost function.

3.3.1. Total desired torque determination

The vehicle model used in last year's study is P0 architecture. So, the vehicle total desired torque is calculated at the engine crankshaft. In this paper, we define the calculation location of vehicle total desired torque according to architecture, for the convenience of calculation, like Figure 3.

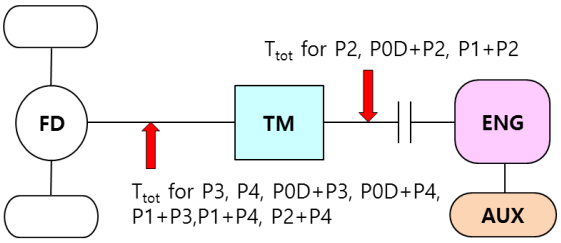


Figure 3: Total Desired Torque Calculation Locations according to Hybrid Drive System Architecture

The vehicle total desired torque according to architecture is calculated as Equation (3) ~ (9). We considered the efficiency of the transmission in a real simulation, but not present in this paper to simplify equation representation. In the equation below, T_{e0} is drive resistance torque of the engine, T_{wp} is load torque of the mechanical auxiliary load, $T_{m0,i}$ (i=2 for P2, 3 for P3, 0D3 for P0D+P3, etc.) is the drive resistance torque of the motor, and γ_{fd} is the gear ratio of the final drive gear. Here, T_{e0} and T_{wp} are calculated at the transmission input shaft. Therefore, for the architecture which calculates vehicle total desired torque at the transmission output shaft, we should calculate these values considering the transmission gear ratio. And since $T_{m0,i}$ is calculated at motor shaft, we should calculate the drive resistance torque of the motor considering the transmission gear ratio or motor location when calculating vehicle total desired torque. In the equation below, T_{e0_diff}, $T_{m0,i\_diff,}$, $T_{wp\_diff,}$ are drive resistance torque of the engine and motor considering the transmission gear ratio, mechanical auxiliary load torque considering the transmission gear ratio respectively.

1. P2

$$T_{tot,2} = \begin{cases} T_{e0} + T_{wp} + T_{m0,2} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & (u_1 \neq 1) \\ T_{m0,2} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & (u_1 = 1) \end{cases} \quad (3)$$

2. P3, P4 ($k = 3,4$)

$$T_{tot,k} = \begin{cases} T_{e0\_diff} + T_{wp\_diff} + T_{m0,k} + \dfrac{T_v}{\gamma_{fd}} & (u_1 \neq 1) \\ T_{m0,k} + \dfrac{T_v}{\gamma_{fd}} & (u_1 = 1) \end{cases} \quad (4)$$

3. P0D+P2

$$T_{tot,0D2} = \begin{cases} T_{e0} + T_{wp} + T_{m0,2} + T_{m0,0D}\gamma_{ply} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & \{(u_1 \neq 1)\,\&\,(u_4 \neq 1)\} \\ T_{e0} + T_{m0,1} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & \{(u_1 \neq 1)\,\&\,(u_4 = 1)\} \\ T_{m0,1} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & \{(u_1 = 1)\} \end{cases} \quad (5)$$

4. P0D+P3, P0D+P4 ($k = 3,4$)

$$T_{tot,0Dk} = \begin{cases} T_{e0\_diff} + T_{wp\_diff} + T_{m0,k} + T_{m0,0D\_diff} + \dfrac{T_v}{\gamma_{fd}} & \{(u_1 \neq 1)\,\&\,(u_4 \neq 1)\} \\ T_{e0\_diff} + T_{m0,k} + \dfrac{T_v}{\gamma_{fd}} & \{(u_1 \neq 1)\,\&\,(u_4 = 1)\} \\ T_{m0,k} + \dfrac{T_v}{\gamma_{fd}} & \{(u_1 = 1)\} \end{cases}$$

(6)

5. P1+P2

$$T_{tot,12} = \begin{cases} T_{e0} + T_{wp} + T_{m0,1} + T_{m0,2} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & (u_1 \neq 1) \\ T_{m0,1} + \dfrac{T_v}{\gamma_{TM}\gamma_{fd}} & (u_1 = 1) \end{cases} \quad (7)$$

6. P1+P3, P1+P4 ($k = 3,4$)

$$T_{tot,1k} = \begin{cases} T_{e0\_diff} + T_{wp\_diff} + T_{m0,k} + T_{m0,1\_diff} + \dfrac{T_v}{\gamma_{fd}} & (u_1 \neq 1) \\ T_{m0,k} + \dfrac{T_v}{\gamma_{fd}} & (u_1 = 1) \end{cases} \quad (8)$$

7. P2+P4

$$T_{tot,24} = \begin{cases} T_{e0\_diff} + T_{wp\_diff} + T_{m0,4} + T_{m0,2\_diff} + \dfrac{T_v}{\gamma_{fd}} & (u_1 + u_5 \neq 1) \\ T_{m0,4} + T_{m0,2} + \dfrac{T_v}{\gamma_{fd}} & (u_1 + u_5 = 1) \end{cases} \quad (9)$$

3.3.2. Motor torque determination by u_1 (First motor)

Table 6 shows the motor of which torque is determined by control input u_1, according to the architecture, and we let say this motor as the first motor from now on. The torque of the motor is calculated as Equation (10). In the Equation (10), $T_{m,i}$ is the torque of the motor of which torque is determined by u_1 for each architecture, and $T_{tot,k}$ is the vehicle total desired torque for each architecture.

Table 6: The Motor of which Torque is determined by Control Input u_1

Architecture		Architecture	
P2	P2	P0D+P4	P4
P3	P3	P1+P2	P2
P4	P4	P1+P3	P3
P0D+P2	P2	P1+P4	P4
P0D+P3	P3	P2+P4	P4

$$T_{m,i} = u_1 T_{tot,k} \quad (i = 2,3,4, k = 2,3,4,0D2,0D3,.....) \quad (10)$$

3.3.3. Motor torque determination by u_5 (Second motor)

Table 7 shows the motor of which torque is determined by control input u_5, according to the architecture, and we let say this motor as the second motor from now on.

Table 7: The Motor of which Torque is determined by Control Input u_5

Architecture		Architecture	
P0D+P2	P0D	P1+P2	P1
P0D+P3		P1+P3	
P0D+P4		P1+P4	
		P2+P4	

If the case of $u_1 = 1$, since the vehicle is propelled by the first motor only, the second motor doesn't generate any torque. Therefore, the second motor torque is determined differently by u_1. And in the case of P0D motor, the torque of the motor is differed by decoupler state. With consideration of these characteristics, the motor torque is determined like Equation (11) ~ (16), according to the architecture. In Equation (13), C_{eff} is the efficiency of delivered driving power, and η_{trans} has the value between 0 and 1. This is the parameter for reflecting power distribution characteristic when motors are located at both front and rear wheel. For example, when the motors of the front and rear wheel all propel the vehicle or all brake the vehicle, each drive power is delivered to ground directly. But when the motor of front wheel propels the vehicle and the motor of rear wheel generate power using front wheel power, the front wheel power is delivered rear wheel motor through the ground. So, in that case, we should consider the efficiency of the delivered drive power. For reflecting these characteristics, we use C_{eff} when architecture has a P4 motor.

1. P0D+P2

$$T_{m,0D} = \begin{cases} u_5 \dfrac{T_{tot,0D2}}{\gamma_{ply}} & \{(u_4 \neq 1) \,\&\, (u_1 \neq 1)\} \\ 0 & \{(u_4 \neq 1) \,\&\, (u_1 = 1)\} \\ T_{m0m0D} + \dfrac{T_{wp}}{\gamma_{ply}} & \{(u_4 = 1)\} \end{cases} \quad (11)$$

2. P0D+P3

$$T_{m,0D} = \begin{cases} u_5 \dfrac{T_{tot,0D3}}{\gamma_{TM}\gamma_{ply}} & \{(u_4 \neq 1)\ \&\ (u_1 \neq 1)\} \\ 0 & \{(u_4 \neq 1)\ \&\ (u_1 = 1)\} \\ T_{m0,0D} + \dfrac{T_{wp}}{\gamma_{ply}} & \{(u_4 = 1)\} \end{cases} \quad (12)$$

3. P0D+P4

$$T_{tot\_f,0D4} = T_{tot,0D4} - C_{eff} T_{m,4}$$

$$C_{eff} = \begin{cases} 1/\eta_{trans} & \{(T_{tot,0D4} \geq 0)\ \&\ (u_1 T_{m,4} \leq 0)\} \\ 1 & (otherwise) \end{cases}$$

$$T_{m,0D} = \begin{cases} u_5 \dfrac{T_{tot\_f,0D4}}{\gamma_{TM}\gamma_{ply}} & \{(u_4 \neq 1)\ \&\ (u_1 \neq 1)\} \\ 0 & \{(u_4 \neq 1)\ \&\ (u_1 = 1)\} \\ T_{m0,0D} + \dfrac{T_{wp}}{\gamma_{ply}} & \{(u_4 = 1)\} \end{cases} \quad (13)$$

4. P1+P2

$$T_{m,1} = \begin{cases} u_5 T_{tot,12} & (u_1 \neq 1) \\ 0 & (u_1 = 1) \end{cases} \quad (14)$$

5. P1+P3

$$T_{m,1} = \begin{cases} u_5 \dfrac{T_{tot,13}}{\gamma_{TM}} & (u_1 \neq 1) \\ 0 & (u_1 = 1) \end{cases} \quad (15)$$

6. P1+P4, P2+P4

$$T_{m,1} = \begin{cases} u_5 \dfrac{T_{tot\_f,14}}{\gamma_{TM}} & (u_1 \neq 1) \\ 0 & (u_1 = 1) \end{cases} \quad (16)$$

3.3.4. Cost

The dpm.m function sum the all cost per step to determine the optimal path. In this paper, the cost is defined by performance index differently.

The cost for fuel economy and additional performance index, electric auxiliary load assist ability is fuel consumption per step time like Equation (17). Therefore, The *DP* for fuel economy and electric auxiliary assist ability will find the optimal path which minimizes fuel consumption.

The cost for drive power reserve is like Equation (18). In the equation, Δm_{fuel} is fuel consumption per step time, k_{pwr} is the coefficient for drive power reserve, $P_{e,max}, P_e$ are the engine maximum power and current power respectively, ω_e , $\omega_{e\text{-}SC}$ are the speed of the engine and 48V electric supercharger respectively, H_{LHV} is a low-heating value of fuel, T_s is step time. And we can see that the second term of Equation (18) has the same unit with fuel consumption. By this term, the engine drive power reserve (The engine maximum power – current power) became larger, the cost became smaller. So, The *DP* for drive power reserve will find the optimal path which maximizes drive power reserve and minimizes fuel consumption.

$$J = \Delta m_{fuel} \quad (17)$$

$$J = \Delta m_{fuel} - k_{pwr} \frac{P_{e,max}(\omega_e, \omega_{e-SC}) - P_e}{H_{LHV}} T_s \quad (18)$$

4. DYNAMIC PROGRAMMING RESULT

4.1. Test scenario

In order to compare performance for each architecture, we carry out *DP* simulation for FTP-75 which is one of the driving cycles that used for evaluating fuel economy of a hybrid electric vehicle generally. The speed profile reference of a vehicle is like Figure 4.

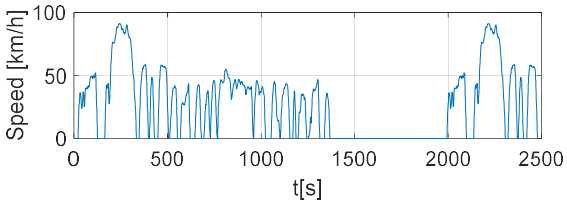


Figure 4: The Speed Profile of FTP-75 Cycle

4.2. Performance index

In this paper, we evaluate hybrid drive system architecture through fuel economy, electric auxiliary load assist ability, and drive power reserve. The index for evaluating each performance is calculated like below respectively.

4.2.1. Fuel economy

We calculate fuel economy like Equation (19) that calculate fuel economy by the total fuel consumption versus total travel distance of the vehicle. In this case, the cost of Model function.m is Equation (17). In the Equation (19), FE is fuel economy of the vehicle,

$Dist_{tot}$ is total travel distance of the vehicle when driving FTP-75 cycle, F_{base} is total fuel consumption when we use cost as Equation (17).

$$FE = \frac{Dist_{tot}}{\sum \Delta m_{fuel}}, \quad F_{base} = \sum \Delta m_{fuel} \tag{19}$$

4.2.2. Electric auxiliary load assist ability performance index

For the case of electric auxiliary load assist ability, the cost of the Model function.m is Equation (17) as the case of fuel economy also. We evaluate the electric auxiliary load assist ability by fuel consumption difference between when an additional electric auxiliary load is applied and when is not applied. If the difference is not large, we can say that the architecture has high electric auxiliary load assist ability because it has a low reduction of fuel economy by an additional electric auxiliary load.

In order to compare performance according to an additional electric auxiliary load size, we apply two kinds of additional loads, 300[W] and 1000[W]. The electric auxiliary load assist ability is defined as Equation (20). In Equation (20), $F_{FE,base}$ is the fuel consumption when an additional electric auxiliary load is not applied, $F_{FE,accelec}$ is the fuel consumption when an additional electric auxiliary load, $P_{elec,add}$, is applied by 300[W] or 1000[W].

$$\begin{aligned} F_{elec,acc} &= F_{FE,accelec} - F_{FE,base} \\ F_{FE,base} &= \sum \Delta m_{fuel} \,(@\, P_{elec,add} = 0) \\ F_{FE,accelec} &= \sum \Delta m_{fuel} (@\, P_{elec,add} = 300 \,or\, 1000) \end{aligned} \tag{20}$$

4.2.3. Drive power reserve performance index

We evaluate the performance of the drive power reserve as the total summation of drive power reserve per step time when we set the cost of Model function.m as Equation (18). In other words, the drive power reserve performance $F_{pwr,rsv}$ is defined as Equation (21).

$$F_{pwr,rsv} = \sum \frac{P_{e,max}(\omega_e, \omega_{e-SC}) - P_e}{H_{LHV}} T_s \tag{21}$$

4.2.4. Comprehensive performance index

We defined the performance of electric auxiliary load assist ability and drive power reserve as Equation (20), (21). We defined each performance index as the concept of fuel consumption, and the reason for this is to solve the scaling problem when calculating the comprehensive performance index. The comprehensive performance index (F_{total}) that integrates all performance index which is described earlier is Equation (22). In Equation (22), α_0, α_1, and α_2 are the weight of each performance index respectively, and a user can adjust the weight when deriving optimal hybrid system architecture, according to what performance is more important. For example, If a user wants the optimal architecture which has higher electric auxiliary load assist ability than fuel economy and the drive power reserve, the user can set the weight as $\alpha_0 = 0.1$, $\alpha_1 = 1$, $\alpha_2 = 0.1$. Each weight has a positive value.

$$F_{total} = \alpha_0 F_{base} + \alpha_1 F_{elec,acc} - \alpha_2 F_{pwr,rsv} \tag{22}$$

4.3. Comparison of fuel economy

The fuel economy results for each architecture obtained by *DP* simulation (FTP-75) is like Figure 5. In this case, we set the cost of Model function.m as Equation (17). We can see P2 architecture have the best fuel economy. The reason for this may be that the motor of P2 architecture can operate at a higher efficiency operating point than other architecture by optimizing the gear stage. And for all architectures, we can see that they have higher fuel economy when a 48V electric supercharger is applied. Therefore, we can conclude that the optimal architecture for fuel economy is P2 with e-SC (48V electric supercharger) architecture.

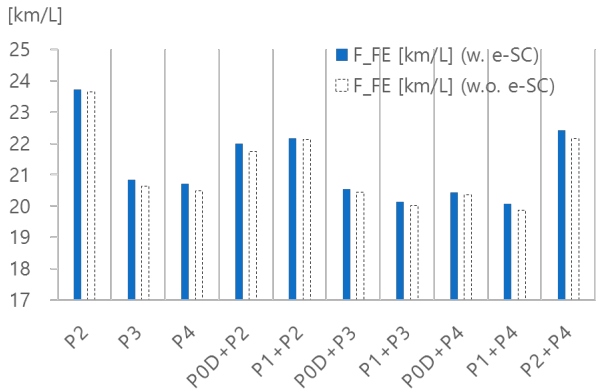


Figure 5: Comparison of FE

4.4. Comparison of electric accessory load assist ability

The electric auxiliary load assist ability for each architecture obtained by *DP* simulation (FTP-75) is like Figure 6, 7. In this case, we set the cost of Model function.m as Equation (17) and apply additional electric auxiliary load. Figure 6 and 7 are the case when an additional electric auxiliary load is 300[W] and 1000[W] respectively. The electric auxiliary load assist ability performance index has a low value when the electric auxiliary load assist ability is high. For both cases that additional electric auxiliary loads are 300[W] and 1000[W] respectively, we can see that the P0D+P2 with e-SC architecture has the best electric auxiliary load assist ability. Therefore, we can conclude that the optimal architecture for the electric auxiliary load assist ability is P0D+P2 with e-SC architecture.

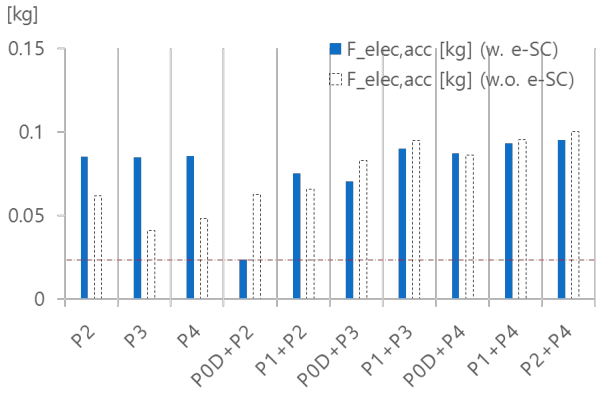


Figure 6: Comparison of $F_{elec,acc}$ ($P_{elec,add} = 300$)

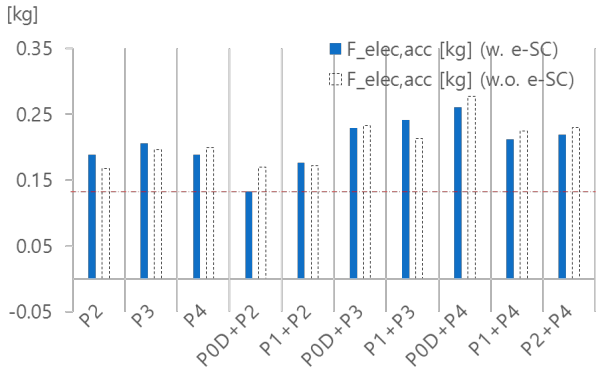


Figure 7: Comparison of $F_{elec,acc}$ ($P_{elec,add} = 1000$)

4.5. Comparison of drive power reserve

The drive power reserve for each architecture obtained by *DP* simulation (FTP-75) is like Figure 8. In this case, we set the cost of Model function.m as Equation (18). The drive power reserve performance index has a high value when the drive power reserve is high. So, we can see that the P0D+P2 without e-SC architecture has the best drive power reserve. Therefore, we can conclude that the optimal architecture for the drive power reserve is P0D+P2 without e-SC architecture.

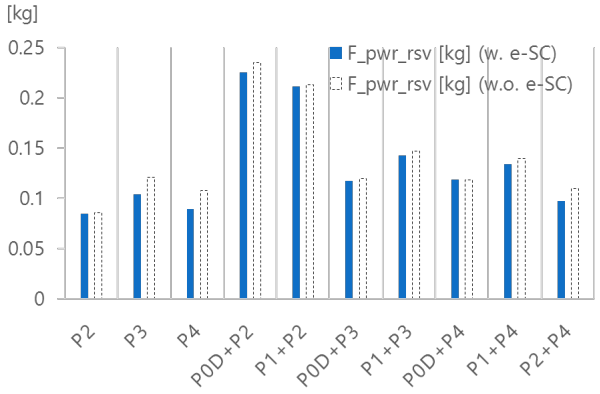


Figure 8: Comparison of $F_{pwr,rsv}$

4.6. Comparison of Comprehensive Performance Index

The comprehensive performance index is differed by weight value, α_0, α_1, α_2. In this paper, we show one examples by setting α_0, α_1, α_2 to 1 all. In other words, the optimal architecture derived from these weight values will be the architecture that has proper fuel economy, electric auxiliary load assist ability and drive power reserve. We calculate the comprehensive performance index for each architecture using performance indices which are calculated in previous sections, and the results are like Figure 9, 10. For both cases of additional electric auxiliary load are 300[W] and 1000[W] respectively, we can see that the P0D+P2 without e-SC architecture has the lowest comprehensive performance index. Therefore, we can conclude that the optimal architecture for the comprehensive performance index is P0D+P2 without e-SC architecture.

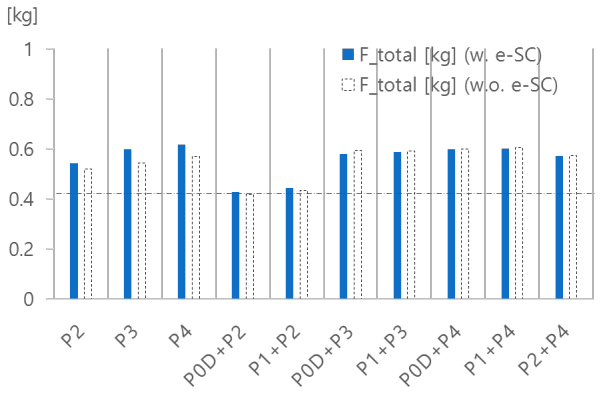


Figure 9: Comparison of F_{total} ($P_{elec,add} = 300$)

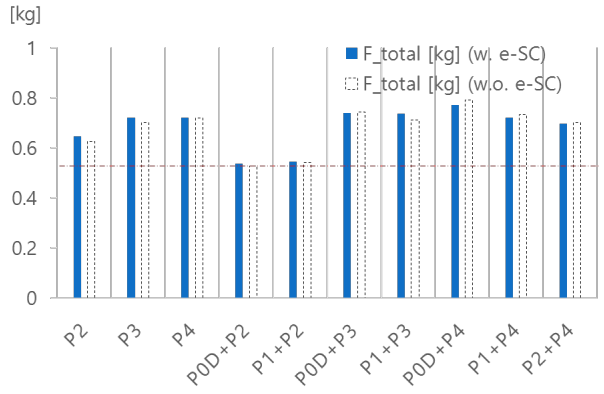


Figure 10: Comparison of F_{total} ($P_{elec,add} = 1000$)

4.7. Ranking by Each Performance Index

Table 8 and 9 are the ranking of the performance indices which are calculated in previous sections. The ranking is low when performance is high. In other words, FE, $F_{pwr,rsv}$ are ordered in which the performance indices are high, and $P_{elec,add}$ are ordered in which the performance indices are low.

In the case of fuel economy, we can see that the architectures having the P2 motor tend to have a high fuel economy. And In the case of the drive power reserve, we can see that the architecture having the motor connected directly to the engine crankshaft (P0D, P1) tend to have a high drive power reserve. In the case of the electric auxiliary load assist ability, we can't find a clear trend when an additional electric auxiliary load is 300[W]. But when an additional electric auxiliary load is 1000[W], we can see that the architecture having both P2 motor and the motor connected directly to the engine crankshaft tend to have a high electric auxiliary load assist ability.

Therefore, taking above analysis together, we can conclude that we should select the P2 architecture when fuel economy is the most important, and select the architecture having crankshaft direct connection motor and P2 motor when electric auxiliary load assist ability or drive power reserve is most important.

Table 8: Ranking by FE and $F_{pwr,rsv}$

Rank	FE		$F_{pwr,rsv}$	
	Architecture	48V e-SC	Architecture	48V e-SC
1	**P2**	O	**P0D**+**P2**	-
2	**P2**	-	**P0D**+**P2**	O
3	**P2**+P4	O	**P1**+**P2**	-
4	P1+**P2**	O	**P1**+**P2**	O
5	**P2**+P4	-	**P1**+P3	-
6	P1+**P2**	-	**P1**+P3	O
7	P0D+**P2**	O	**P1**+P4	-
8	P0D+**P2**	-	**P1**+P4	O
9	P3	O	P3	-
10	P4	O	**P0D**+P3	-
11	P3	-	**P0D**+P4	-
12	P0D+P3	O	**P0D**+P4	O
13	P4	-	**P0D**+P3	O
14	P0D+P4	O	**P2**+P4	-
15	P0D+P3	-	P4	-
16	P0D+P4	-	P3	O
17	P1+P3	O	**P2**+P4	O
18	P1+P4	O	P4	O
19	P1+P3	-	**P2**	-
20	P1+P4	-	**P2**	O

Table 9: Ranking by $P_{elec,add}$

Rank	$F_{elec,acc}$ ($P_{elec,add} = 300$)		$F_{elec,acc}$ ($P_{elec,add} = 1000$)	
	Architecture	48V e-SC	Architecture	48V e-SC
1	P0D+P2	O	**P0D**+**P2**	O
2	P3	-	**P2**	-
3	P4	-	**P0D**+**P2**	-
4	P2	-	**P1**+**P2**	-
5	P0D+P2	-	**P1**+**P2**	O
6	P1+P2	-	**P2**	O
7	P0D+P3	O	P4	O
8	P1+P2	O	P3	-
9	P0D+P3	-	P4	-
10	P3	O	P3	O
11	P2	O	**P1**+P4	O
12	P4	O	**P1**+P3	-
13	P0D+P4	-	**P2**+P4	O
14	P0D+P4	O	**P1**+P4	-
15	P1+P3	O	**P0D**+P3	O
16	P1+P4	O	**P2**+P4	-
17	P1+P3	-	**P0D**+P3	-
18	P2+P4	O	**P1**+P3	O
19	P1+P4	-	**P0D**+P4	O
20	P2+P4	-	**P0D**+P4	-

5. CONCLUSION

In this paper, we selected exploration target 48V hybrid drive architecture by exploring possible 48V hybrid drive system manually and compared the performance of each architecture through *DP* simulation. When comparing the performance of the hybrid system, we used not only fuel economy but also electric auxiliary load assist ability and drive power reserve. Here, we developed performance indices for electric auxiliary load assist ability and drive power reserve in order to compare performance. And the performance indices have a unit of fuel consumption in order to solve the problem scaling when calculating comprehensive performance index. As results, we found that P2 is the optimal architecture for fuel economy, and the architecture having crankshaft direct connection motor (P0D, P1) and P2 motor is the optimal architecture for electric auxiliary load assist ability or drive power reserve.

We considered an engine only when calculating drive power reserve performance index. But a hybrid electric vehicle has not only an engine but also a motor for driving. Therefore, in future work, if we consider a drive power reserve of an engine and a motor together, the drive power reserve performance is expected to be evaluated more accurately.

ACKNOWLEDGMENTS

This work was supported by the Industrial Strategic Technology Development Program (10076437, Development of hybrid drive topology exploration and control technology for fuel economy optimization of 48V mild HEVs) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea).

REFERENCES

Andreas Baumgardt, Dieter Gerling, 2015a. 48V Recuperation Storage Based on Supercaps for Automotive Application, EVS28 International Electric Vehicle Symposium and Exhibition, May 3-6, KINTEX (Korea).

Andreas Baumgardt, Dieter Gerling, 2015b. 48V Recuperation Storage Including Stabilizing 12V Tap for HEVs, 2015 IEEE Transportation Electrification Conference and Expo (ITEC), June 14-17, Dearborn (Michigan, USA).

Anthony Rick, Brain Sisk, 2015. A Simulation Based Analysis of 12V and 48V Microhybrid Systems Across Vehicle Segments and Drive Cycle, SAE 2015 World Congress & Exhibition, April 21-23. Detroit (Michigan, USA).

Dr. Ing. Olivier COPPIN, 2016. Valeo Presentation – From 12 plus 12V to 48V. Available from: https://48v-vehicles.iqpc.de/downloads/valeo-presentation-from-12-plus-12v-to-48v [accessed 5 March 2019]

Ji Y., Park J, Lee H., 2018. A study on operating characteristics of a 48V electric supercharger and P0 configuration motor for fuel economy improvement. MAS 2018, September 17, 39-47.

Junyong Park, Taeho Park, Hyeongcheol Lee, 2017. An optimization of power distribution considering supercharging characteristic for electric supercharger applied 48V mild hybrid vehicle, KSAE 2017 Annual Autumn Conference & Exhibition, pp.1328-1333. November 15-18, Yeosu (Jeollanam-do, South Korea).

Kapadia J., Kok D., Jennings M., Kuang M., Masterson B., Isaacs R., Dona A., Wagner C., and Gee T., 2017. Powersplit or Parallel - Selecting the Right Hybrid Architecture. SAE International Journal of Alternative Powertrains, 6.

Malte Kuypers, 2014. Application of 48 Volt for Mild Hybrid Vehicles and High Power Load, SAE 2014 World Congress & Exhibition, April 1, Detroit (Michigan, USA).

Mark Schudeleit, Christian Sieg, Ferit Kücükay, 2015. The Potential of 48V HEV in Real Driving, International Scholarly and Scientific Research & Innovation Vol: 9 No: 10, 1719-1728.

Olle Sundstrom, Lino Guzzella, 2009. A Generic Dynamic Programming Matlab Function, 18th IEEE International Conference on Control Applications Part of 2009 IEEE Multi-conference on Systems and Control Saint Petersburg, July 8-10, St. Petersburg (Russia).

Ran Bao, Victor Avila, James Baxter, 2017. Effect of 48V Mild Hybrid System Layout on Powertrain System Efficiency and Its Potential of Fuel Economy Improvement, WCX™ 17: SAE World Congress Experience, April 4-6, Detroit (Michigan, USA).

Thomas Eckenfels, Florian Kolb, Steffen Lehmann, Waldemar Neugebauer, Manel Calero, 2018. 48V Hybridization, A Smart Upgrade for the Powertrain. Available from: http://schaeffler-events.com/symposium/lecture/h3/index.html [accessed 6 July 2018]

Zifan Liu, Andrej Ivanco, and Zoran S. Filipi, 2016. Impact of Real-World Driving and Driver Aggressiveness on Fuel Consumption of 48V Mild Hybrid Vehicle, SAE International, April 5, 249-258.

FUEL-OPTIMAL PATH FINDING ALGORITHM USING TRAFFIC INFORMATION AT URBAN INTERSECTION

Jooin Lee[(a)], Hyeongcheol Lee[*(b)]

[(a)] Department of Electrical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea
[(b)]Department of Electrical and Biomedical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea

[(a)]galaxyeagle@hanyang.ac.kr, [(b)]hclee@hanyang.ac.kr
**Corresponding Author, E-mail: hclee@hanyang.ac.kr*

ABSTRACT
Intelligent Transportation System (ITS) is actively studied as the sensor and communication technology in the vehicle develops. The Intelligent Transportation System collects, processes, and provides information on the location, speed, and acceleration of the vehicles in the intersection. This paper proposes a fuel optimal route decision algorithm. The algorithm estimates traffic condition using information of vehicles acquired from several ITS intersections and determines the route that minimizes fuel consumption by reflecting the estimated traffic condition. Simplified fuel consumption models and road information (speed limit, average speed, etc.) are used to estimate the amount of fuel consumed when passing through the road. Dynamic Programming (DP) is used to determine the route that fuel consumption can be minimized. This algorithm has been verified in an intersection traffic model that reflects the actual traffic environment (Korea Daegu Technopolis) and the corresponding traffic model is modeled using AIMSUN.

Keywords: Fuel Consumption, Traffic Information, Dynamic Programming(DP), Intelligent Transportation Systems(ITS), Advanced Interactive Microscopic Simulator for Urban and Non-Urban Networks (AIMSUN)

1. INTRODUCTION

As the consumption of fossil fuels increases worldwide, environmental pollution becomes a more serious problem. The engine of an automobile burns fuel and exhaust many kinds of harmful gas. Typical automobile exhaust gas contains hydrocarbons (HC), nitrogen oxides (NOx), carbon monoxide (CO), carbon dioxide (CO2), and particle mass (PM). Automobile exhaust pollution has a detrimental effect on the human body and cause environmental changes such as global warming. Automobile emissions regulations are being strengthened to solve environmental problems caused by engine exhaust gas. Various technologies have been developed to improve fuel efficiency and satisfy environmental regulations. Engine control technologies such as 'lean-burn' and 'engine downsizing' have greatly increased engine efficiency, and component processing technology has significantly reduced power transmission losses. Eco-friendly vehicles using two or more distinct types of power (Fuel Cell Electric Vehicle, (Plug-in) Hybrid Electric Vehicle, etc.) are being actively developed and driving on the road.

However, not only the automotive manufacturing / control technology but also the actual driving environment experienced by the driver is highly related to fuel economy and exhaust pollution. (Min Zhou 2016) The actual driving environment includes the driving behavior of the driver, the traffic environment (queue length, speed limit, etc.), and the road environment (weather, traffic light, road curvature, etc.). Road information and traffic information are required to control the vehicle in accordance with the actual driving environment. The ITS has various measuring equipments (camera, radar, etc.) and communication systems (WAVE, 5G, etc.), so it can provide information such as traffic average speed and traffic signal schedule to the vehicles. There are many studies on how to reflect this information in the vehicle speed control algorithm and routing algorithm. (Hesham Rakha 2011; Matthew Barth 2011; Hao Yang 2016; Raj kishore Kamalanathsharma 2014; M.A.S. Kamal 2010; Matej 2016; Jie Sun 2015; Xian Huang 2018) These algorithms are very helpful in preventing traffic accidents and improving fuel economy.

In this paper, we propose an algorithm to determine the route that can improve fuel economy by using road / traffic information delivered from ITS. First, the traffic flow model is used to estimate the state of traffic such as the average speed and the average travel time for each road. Second, simplified fuel consumption model is used to estimate instantaneous fuel consumption at vehicle speed and acceleration. The estimated instantaneous fuel consumption and road information are used to determine the amount of fuel consumed to pass the road. Instantaneous fuel consumption refers to the amount of fuel consumed in a steady state, ignoring

the transient state of the engine. It is assumed that the transient state of the engine is negligible because it is very small compared to the normal operating state. Third, the algorithm defines a cost function to minimize the amount of fuel consumed to pass the road and derives the optimal path using Dynamic Programming (DP). The DP algorithm is an algorithm for finding a globally optimal solution when the status of the entire system is known. The algorithm was programmed using Python and validated at the intersection AIMSUN traffic model reflecting the actual traffic environment (Korea Daegu Technopolis). AIMSUN is a software that can microscope the traffic environment and verify vehicle behavior in traffic flow.

2. TRAFFIC STATE ESTIMATION MODEL

The traffic state estimation model estimates the 'Travel Time' and 'Travel Fuel Consumption'. The traffic state estimation model uses average speed and traffic density based on Green-shield linear traffic model. The Green-shield linear traffic model assumes that the traffic average speed and traffic density have a linear relationship. The relationship can be seen in Figure 1.

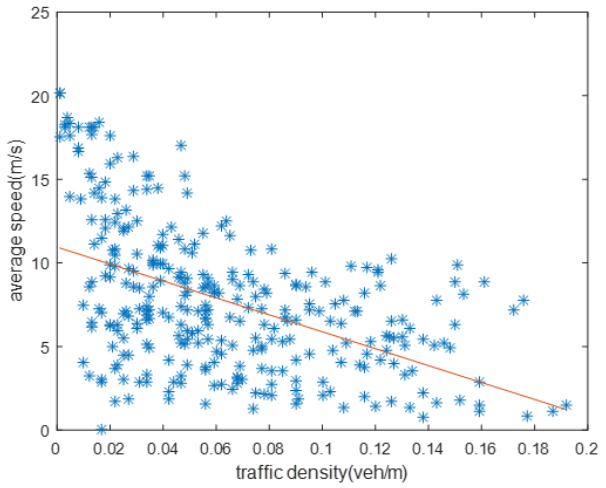


Figure 1: The relationship between space mean velocity and traffic density

2.1. Travel Time

The 'Travel Time' means the time required to pass through each road. Travel time(T_{Travel}) can be estimated using 'space mean speed(V_{space_mean})', 'speed limit($V_{\lim it}$)' and 'road length(L_{road})'. The equations for the travel time are (1).

$$T_{Travel} = \max\left(\frac{L_{road}}{V_{space\_mean}}, V_{\lim it}\right) \quad (1)$$

Space mean speed is closely related to fuel consumption. The relationship can be seen in Figure 2. It can be seen that the fuel consumption according to the average speed increases in a specific section. This means that the engine efficiency in not good in the corresponding speed section. Even if fuel consumption is minimized, excessive travel time is not a reasonable solution. Therefore, travel time can be applied as a constraint on the cost function.

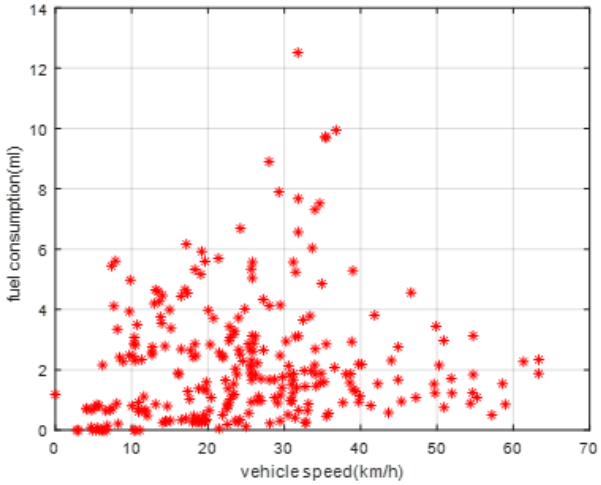


Figure 2: The relationship between space mean velocity and fuel consumption

2.2. Travel Fuel Consumption

The 'Travel Fuel Consumption' means the fuel consumption required to pass through each road. Travel fuel consumption(FC_{travel}) can be estimated using 'instantaneous fuel consumption rate(FR_{inst})' and 'travel time(T_{travel})'. The equations for the travel time are (2).

$$FC_{travel} = FR_{inst} \times T_{travel} \quad (2)$$

The instantaneous fuel consumption rate is the fuel consumption rate, which is determined only by the current state, assuming a steady state, and is determined by the longitudinal vehicle dynamics and the simplified fuel consumption rate model. The instantaneous fuel consumption rate is calculated by the simplified fuel consumption rate model. (Engin Ozatay 2013; Hesham Rakha 2011) Actual engines have nonlinear fuel consumption characteristics. It appears to be an overly complex formula and requires a long computation time. Simplified fuel consumption rate model cannot calculate accurate fuel consumption, but they can represent fuel consumption trends and have simple formulas and short computing times. As a result, a simplified fuel consumption rate model can be used to find the trends to minimize fuel consumption. The equations for the constructed model are (3).

$$FR_{inst} = \frac{m}{eH_L\eta_{drive}} \times a_{input}(t) \times v(t) + \frac{P_{Loss} \times V_d}{4\pi \times eH_L}\left(\frac{\gamma(t)}{R_{wh}}\right) \times v(t) + \dot{m}_{idle} \quad (3)$$

The fuel consumption model is related to longitudinal vehicle dynamics. The longitudinal vehicle dynamics can be seen in Figure 3 and the equation is (4) - (6). Table 1 shows the vehicle parameter.

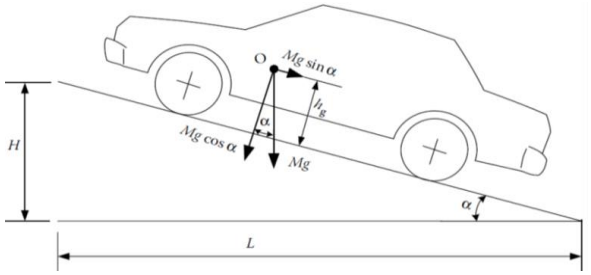


Figure 3: Longitudinal Vehicle Dynamics

$$M\frac{dv(t)}{dt} = F_{tract}(t) - F_{aero}(t) - F_{grade}(t) \quad (4)$$

$$F_{aero} = \frac{1}{2}\cdot\rho\cdot A_f\cdot C_d\cdot v^2(t) \quad (5)$$

$$F_{grade} = M\cdot g\cdot\sin(\alpha) \quad (6)$$

Table 1: Vehicle Dynamics Parameters

Vehicle Mass - M $[kg]$	1202
Frontal Area - A_f $[m^2]$	2.7
Air Drag Coefficient - C_d [–]	0.3
Air Density - ρ $[kg/m^3]$	1.206

A longitudinal vehicle dynamics model and a simplified fuel consumption model were verified using Carsim. Carsim is software that is primarily used to simulate the dynamic behavior of a vehicle. The verification result is the same as Figure 4 and Figure 5. In the graph, the blue line represents the nonlinear model and the red line represents the simplified model. The tendency of the simplified fuel consumption rate model is the same as the tendency of the non-linear fuel consumption rate model.

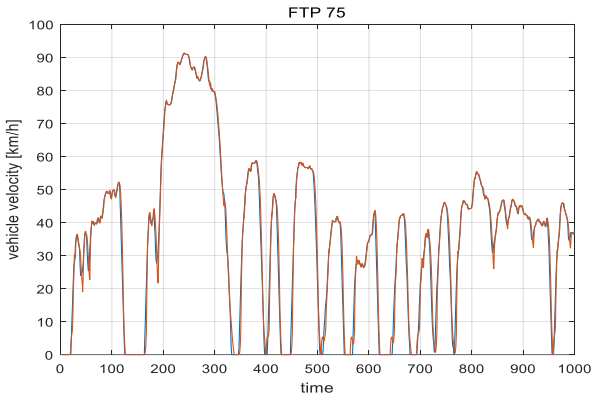


Figure 4: Vehicle velocity estimation result for FTP-75 city cycle

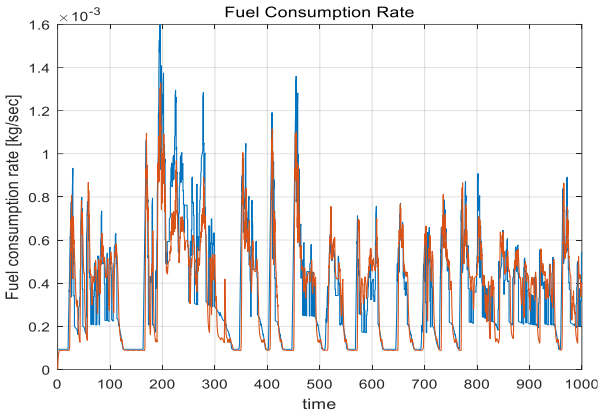


Figure 5: Instantaneous fuel consumption estimation result for FTP-75 city cycle

3. FUEL-OPTIMAL PATH FINDING ALGORITHM

The Fuel optimal routing algorithm uses dynamic programming to derive a path that minimizes fuel consumption.

3.1. Dynamic Programming

Dynamic programming, designed by Richard Bellman, is an analysis technique used to find global optimization solutions for complex systems. Dynamic programming divides a complex optimization problem into several simple sub-problems and derives the optimal solution of the complex problem by solving the sub-problems. In order to apply dynamic programming, a value function is designed for the purpose.

3.2. Cost Function

This algorithm aims at minimizing fuel consumption and reaching the destination. Depending on the purpose, the cost function($Cost\_function$) is designed to minimize fuel consumption(FC_{travel}). That is, the path that can minimize the total fuel consumption from the departure time(T_0) to the arrival time(T_f) at the destination is determined. The formula for the cost function is (7).

$$Cost\_function = \int_{t_o}^{t_f} FC_{travel}\, dt \quad (7)$$

The weighted graph of the road is determined by the cost function and is shown in Figure 6. The weight graph consists of nodes and sections, nodes represent intersections, and sections represent roads. Weights are assigned to each section. Each node is assigned an ID corresponding to an intersection.

Unlike other path finding algorithms that minimize travel time or travel distance, the route is determined by comparing the fuel consumption of the entire route.

Figure 6: Weighted graph for actual intersections

4. SIMULATION & RESULT

4.1. Simulation Environment

The Fuel - optimal path finding algorithm is implemented using Python. The algorithm was verified by an intersection traffic model reflecting the actual urban traffic environment. The intersection traffic model was designed using AIMSUN. AIMSUN is a microscopic traffic simulation software, which can simulate traffic flow through the movement of individual vehicles. AIMSUN worked with Python 2.7.13 to control the vehicles in the intersection traffic model.

The modeling target is 19 intersections located in Technopolis, Daegu, Korea. In order to simulate actual traffic environment, road characteristics (traffic signal schedule, road shape) and traffic characteristics were reflected in urban intersection traffic model. The urban intersection traffic model is the same as Figure 7. The traffic signal schedule configuration is shown in Figure 8. Figure 9 shows the traffic simulation screen using AIMSUN.

To verify this algorithm, we compare the simulation result with the shortest path finding algorithm to minimize travel distance. The performance of the algorithm was verified by comparing fuel consumption.

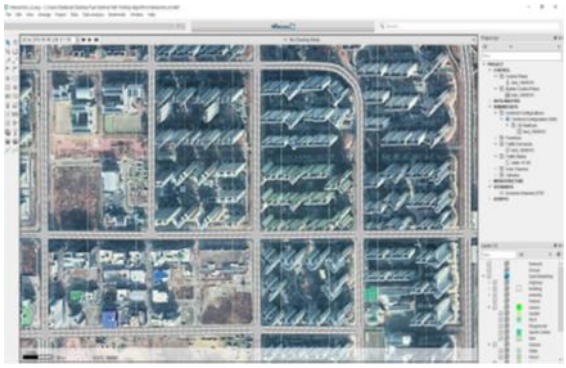


Figure 7: AIMSUN urban intersection traffic model

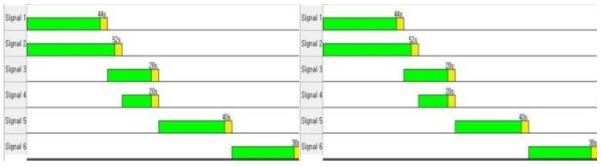


Figure 8: AIMSUN traffic model - traffic signal schedule

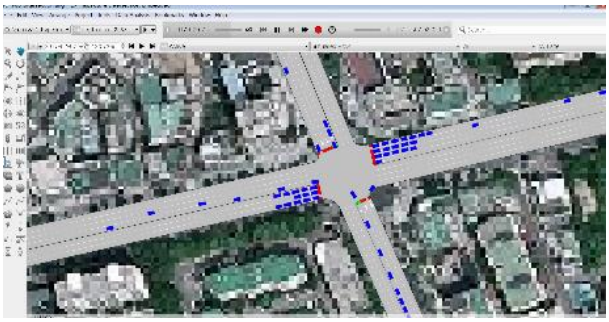


Figure 9: AIMSUN Traffic simulation screen

4.2. Simulation Result

The simulation results are shown in Figure 10. 200 vehicles of the total 4000 vehicles were controlled. The verification vehicle starts from node 24 and arrives at node 3. We simulated the saturation traffic situation of commuting time.

The red path represents the shortest path and the yellow represents the Fuel-optimal path finding algorithm. Table 2 shows the simulation result. The shortest path finding algorithm consumes 74 liters of fuel with an average travel time of 2148 seconds and the fuel-optimal routing algorithm consumes 68 liters of fuel with an average delay of 2250 seconds. As a result, it was confirmed that the fuel-optimal path finding algorithm improved the fuel efficiency of the controlled vehicle by approximately 9.04%. Although the fuel-optimal path finding algorithm has a longer travel time, it can be seen that the fuel consumption is lower.

Figure 10: Simulation results - path comparison

Table 2: Simulation results

Shortest path finding algorithm	
Fuel Consumption	74.86 [L]
Average Travel Time	2148.0 [sec]
Fuel-optimal path finding algorithm	
Fuel Consumption	68.09 [L]
Average Travel Time	2250.4 [sec]

5. CONCLUSION

In this paper, we propose an algorithm to find the path that minimizes fuel consumption. The algorithm uses traffic information and vehicle information. The average travel time is estimated using traffic information. The instantaneous fuel rate is calculated using simplified fuel consumption model and vehicle status. The instantaneous fuel rate and average travel time are used to calculate travel fuel consumption, which is the amount of fuel consumed in passing through the road. The cost function is designed using the travel fuel consumption and the weight of each road is given. Dynamic Programming is used to find a path that minimizes the corresponding cost function over the entire travel period. To verify the algorithm, we modeled intersection traffic model for 19 actual traffic environment of 19 actual intersections of Daegu Technopolis, Seoul, Korea. Micro traffic simulation software AIMSUN was used as a traffic modeling environment. A total of 4,000 cars were used to construct traffic, of which 200 vehicles controlled the route. We compared the fuel consumption of the vehicles with the shortest path finding algorithm and the fuel optimal path finding algorithm respectively. It was confirmed that the fuel-optimal path finding algorithm uses 9% less fuel and 100 seconds longer travel time. It can be seen from this result that reducing the travel time or travel distance is not the only wat to reduce fuel consumption.

ACKNOWLEDGMENTS

This work is supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No. R7117-16-0132, The development of future technology for mixed traffic intersection integrated environmental information regarding the autonomous at the crossroads of the city)

REFERENCES

Min Zhou, Hui Jin, Wenshuo Wang, 2016. A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing. Transportation Research Part D: Transport and Environment, New York.

Xianan Huang, Huei Peng, 2018. Eco-Routing based on a Data Driven Fuel Consumption Model. Cornell University, New York.

Hesham Rakha., and Raj kishore Kamalanathsharma., 2011. Eco-driving at signalized intersections using V2I communication. Intelligent Transportation Systems , Washington.

Matthew Barth., Sindhura Mandava., Kanok Boriboonsomsin., and Haitao Xia., 2011. Dynamic ECO-driving for arterial corridors. Integrated and Sustainable Transportation System (FISTS), Vienna..

Tien-Wen Sung., Liang-Cheng Shiu., Fu-Tian Lin., and Chu-Sing Yang., 2015. A Speed Control Scheme of Eco-Driving at Road Intersections. Third International Conference on Robot, Vision and Signal Processing (RVSP), Kaohsiung.

Hao Yang., Hesham Rakha., and Mani Venkat Ala., 2016. Eco-Cooperative Adaptive Cruise Control at Signalized Intersections Considering Queue Effects. IEEE Transactions on Intelligent Transportation Systems, China.

Raj kishore Kamalanathsharma., and Hesham Rakha., 2014. Leveraging connected vehicle technology and telematics to enhance vehicle fuel efficiency in the vicinity of signalized intersections. Journal of Intelligent Transportation Systems, London.

Md. Abdus Samad Kamal., Masakazu Mukai., Junichi Murata., and Taketoshi Kawabe., 2013. Model predictive control of vehicles on urban roads for improved fuel economy. IEEE Control Systems Society, India.

M.A.S. Kamal., M. Mukai., J. Murata., and T. Kawabe., 2010. On board eco-driving system for varying road-traffic environments using model predictive control. IEEE International Conference on Control Applications , Yokohama, Japan.

M.A.S. Kamal., M. Mukai., J. Murata., and T. Kawabe., 2011. Ecological driving based on preceding vehicle prediction using MPC. IFAC Proceedings Volumes 44.1 , Milano, Italy.

Matěj Kubička, Jan Klusáček, Jan Klusáček, Arben Cela, Hugues Mounier, Laurent Thibault, S. I. Niculescu, 2016. Performance of current eco-routing methods. IEEE Intelligent Vehicles Symposium, France.

Matěj Kubička, Jan Klusáček, Jan Klusáček, Arben Cela, Hugues Mounier, Laurent Thibault, S. I. Niculescu, 2016. Performance of current eco-routing methods. IEEE Intelligent Vehicles Symposium, France.

JieSun, Henry X.Liu, 2015. Stochastic Eco-routing in a Signalized Traffic Network. Transportation research procedia, USA.

JieSun, Henry X.Liu, 2015. Stochastic Eco-routing in a Signalized Traffic Network. Transportation research procedia, USA.

Engin Ozatay., Umit Ozguner., Dimitar Filev., and John Michelini., 2013. Analytical and numerical

solutions for energy minimization of road vehicles with the existence of multiple traffic lights. IEEE 52nd Annual Conference on Decision and Control (CDC), Florence , Italy.

Hiraoka Toshihiro., Taketoshi Kunimatsu., Osamu Nishihara., and Hiromitsu Kumamoto., 2005. Modeling of driver following behavior based on minimum-jerk theory. 12th World Congress ITS, San Francisco.

Hernsoo Hahn., Hanseok Ko., Osamu Nishihara., and Sukhan Lee., 2008. Multisensor Fusion and Integration for Intelligent Systems. the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems.

Luc Binette., Michael A. Dopita., and Ian R. Tuohy., 1985. Radiative shock-wave theory. II - High-velocity shocks and thermal instabilities. Astrophysical Journal.USA.

Tom V. Mathew., and K V Krishna Rao., 2007. Traffic stream models. National Programme on Technology Enhanced Learning (NPTEL). India.

Junbo Jing., 2014. Vehicle Fuel Consumption Optimization using Model Predictive Control based on V2V communication, The Ohio State University, Ohio.

MACHINE LEARNING TECHNIQUES APPLIED TO INDUSTRIAL ENGINEERING: A MULTI CRITERIA APPROACH

Fabio De Felice [(a)], Marta Travaglioni [(b)], Giuseppina Piscitelli [(b)], Raffaele Cioffi [(b)], Antonella Petrillo [(b)]

[(a)] Department of Civil and Mechanical Engineering - University of Cassino and Southern Lazio, Italy
[(a)] Department of Engineering - University of Naples "Parthenope", Italy

[(a)]defelice@unicas.it, [(b)]antonella.petrillo@uniparthenope.it, [(b)]marta.travaglioni@uniparthenope.it, [(b)]giuseppina.piscitelli@uniparthenope.it

ABSTRACT
With the Industry 4.0 (I4.0) beginning, the world is witnessing an important technological development. The success of I4.0 is linked to the implementation of enabling technologies, including Machine Learning, which focuses on the machines' ability to receive a series of data and learn on their own. The present research aims to systematically analyze the existing literature on the subject in various aspects, including publication year, authors, scientific sector, country, institution and keywords. Understanding and analyzing the existing literature on Machine Learning applied to predictive maintenance is preparatory to recommend policy on the subject.

Keywords: machine learning, predictive maintenance, literature review, industry 4.0

1. INTRODUCTION

With Industry 4.0, companies can exploit Artificial Intelligence applied as predictive analysis, to monitor the production process and have a more in-depth view of their activity in real time. The machines and components connected to the Internet of Things system send information any time, then they are historicized and analyzed with algorithms (Petrillo et al., 2019). The latter work on Machine Learning logics to predict system failures in advance and estimate the residual life of the machines. This means there is no longer the dilemma between waiting for the machine downtime or replacing the components ahead of time.
The challenge of proper scheduling grows with the complexity of machines. In this context, predictive can help to detect the anomalies and failure patterns and provide early warnings (Mobley 2002). It is based on the possibility of recognizing a progression anomaly through the notice and interpretation of weak signals that define a future failure. Therefore, it is not just talking about data generated by the interactions between machines and operators or by M2M (Machine to Machine) connection systems, but to process them with Analytics solutions.
The existence of a system that collects all the data coming from multiple sources allows to identify recursive patterns, that define the probability with which an event can happen in the future and to learn automatically from the feedback of the activities carried out. Thus, Machine Learning (ML) models are enabled to activate Predictive Maintenance (PdM).
With I4.0 and the consequent increase in data availability, the cases in which the ML is implemented for the PdM increased.
Given considerable importance of the subject, the present research will carry out a bibliographic survey that highlights the most important texts with respect to a research area, including year of publication, authors, scientific sector, country, institution and keyword. It not only offers a basis for future comparisons but prompts a number of new questions for investigations as well.
The rest of the paper is organized as follows: Section 2 presents a brief review of machine learning techniques; in Section 3 proposed bibliometric analysis method is explained; in Section 4 the sample building process is explained; in Section 5 results of the bibliometric analysis is developed; in Section 6 results of analysis are discussed. Finally, in Section 7 are summarized the main contribution of the research.

2. BRIEF REVIEW OF MACHINE LEARNING TECHNIQUES

One of the largest areas currently included in the label of artificial intelligence is Machine Learning. This term includes a large number of different techniques and approaches, but all of these share a clear and unambiguous property: solving problems through algorithms that have the ability to learn from what is entered as input. In these terms, the concept of knowledge changes. Knowledge is no longer the human one and transferred to the machine, but it is learned by the machine itself in an autonomous way from the activities that it carries out gradually. Consequently, the role of man also changes: he will only have to define, through programming, how the machine will have to learn. In some cases, the operator provides examples and information from which the machine must learn, so it can develop a "knowledge" that it will transfer or allow to automate activities.
The different approaches have been uniquely classified by scientific community, depending on their learning model. Each model works, in principle, on the basis of

two distinct approaches: supervised learning and unsupervised learning (Samuel A.L. 1959).
In the case of supervised learning, the computer is given concrete examples to be used in carrying out the required task, while in the case of unsupervised learning the software works without any kind of assistance.
Within these two categories there are subsets that allow us to classify Machine Learning even more in detail.
The first is *Supervised Learning*, the computer is given examples in the form of possible inputs and the respective desired outputs, to extract a general rule that associates the input with the correct output. It is usually used in applications where historical data can predict future events.
The second subset is *Unsupervised Learning*, where the system is provided only with data sets without indication of the result sought. The computer receives inputs in no way labeled, from which it must find a structure. This means the system is not given the "right" answer and therefore the algorithm must find out what is shown to it.
The third subset is *Reinforcement Learning*. The system (which can be a computer, software or algorithm) must be able to interact with a dynamic environment, from which to draw input data and reach a goal. If the goal is achieved, the system will be rewarded with a performance evaluation. The purpose is to learn an optimal policy that selects the subsequent actions to each state in order to maximize the rewards accumulated over time.
The fourth and final subset is the *Semi-Supervised Learning*. This, unlike the other subsets, is a hybrid model: incomplete data sets are supplied to the computer (some of the inputs are also output) as in Supervised Learning, and others lack them, as in Unsupervised Learning.

3. BIBLIOMETRIC ANALYSIS METHODS

The interest in using bibliometric techniques has increased thanks to the availability of large scientific databases. It involves a series of techniques that are used to quantify the process of written communication and identify patterns. The methodological used approach mixes bibliometric, content analysis and social network techniques. In this state-of-the-art study the research was initiated through the consultation of electronic databases, Scopus (SCP) and Web of Science (WoS), on April 9, 2019 and dividing it into the following phases.
In Phase 1, bibliometric data was collected from the SCP and WoS databases. This phase required three steps for the construction of the sample to be analyzed:

1. Step #1: identification;
2. Step #2: screening;
3. Step #3: inclusion.

Once Phase 1 is completed, the next phase is Phase 2, which is the analysis of the results.
The approaches used for the bibliometric analysis were:

- the use of indicators for the parameters studied;
- the SNA (Social Network Analysis) for the authors analysis.

The indicators chosen to perform the analysis are Total Papers (TP), which is the total number of publications and Total Citation (TC), which is the total number of citations. The SNA finds application in various social sciences, lately employed in the study of various phenomena such as international trade, information dissemination, the study of institutions and the functioning of organizations. The analysis of the use of the term SNA in the scientific literature has undergone an exponential growth in utilization of this mode of computable representation for complex and interdependent phenomena. For the purpose of this study, UCINET, NetDraw software was used, expressly designed for the creation and graphic processing of networks. It was used to represent the connection between publication and citation in authors' network, and Excel for data input. To analyze keywords, NVivo 12 software was used. At the end of the second phase, a third and final one follows, where the results will be discussed, and conclusions will be drawn. All analyzed documents are listed in Appendix (Table A).

4. BIBLIOMETRIC ANALYSIS: PHASE 1 - RESEARCH

In this paragraph, the results of the bibliometric research are analyzed, starting from Phase 1 in which the process of construction of the sample is described.

4.1. Step#1: Identification

First phase is dedicated to documents collection and the sample's construction.
It is divided into three steps; every step is useful for selecting the documents of interest and excluding ones that are not interesting for the purpose.
The first step is named "Identification".
In this step, Scopus (SCP) and Web of Science (WoS) databases were taken into consideration. In order to maintain the consistency of the results, the same keywords were used:

- Machine Learning;
- Predictive Maintenance.

In addition, a time period of 20 years was chosen, from 1999 to 2019, as shown in Table 1.

Table 1: Keywords And Time Period For Research

Keywords and Time Period		
Keywords	Machine Learning (ML)	Predictive Maintenance (PM)
Time Period	From 1999	To 2019

The results extracted by Scopus are numerically superior to Web of Science: 320 for the first database and just 173 for the second one (Table 2).

Table 2: Total Results Of Research On SCP And Wos

Data Extraction			
Source of research	SCP	WoS	Total
Results	320	173	493

Figure 2 shows the aggregated result for Scopus and Web of Science query; it points out that the results obtained are scarce in the period before I4.0. On the contrary, there is an increase in publications since the beginning of the fourth industrial revolution. Overall, this is a fairly poor operating result, just 493 documents, most of them were concentrated in year 2018, with 172 results (Figure 1).

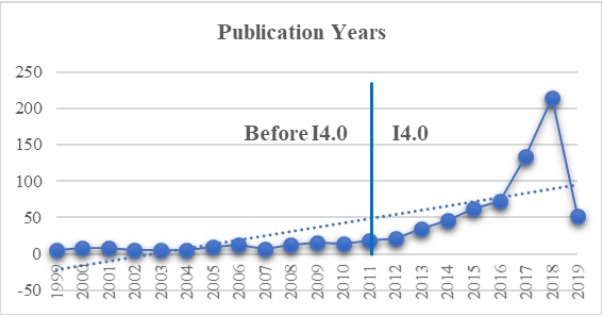


Figure 1: Research Growth

Wanting to analyze the numbers before and after I4.0 (Table 3), just before the fourth revolution, 23 documents were recorded in SCP and 19 in WoS. While after 2011 there are 297 documents in Scopus and 154 in WoS, as shown in Table 3.

Table 3: SCP And Wos Research Areas

Results			
-	Before I4.0	I4.0	Total
SCP	23	297	320
WoS	19	154	173
Total	42	451	-

With particular attention to year 2018, 122 and 50 documents are recorded on Scopus and Web of Science respectively.
The reason is not to be sought in the scarce interest, although Machine Learning is dated already to the last century, Predictive Maintenance has encountered an increase in development only with I4.0.
It follows that the analysis of the literature concerning the period 1999 - 2010 can be considered negligible, since it is an overall result equal to 42 documents out of 493.
On the contrary, it is useful to carry out an analysis starting from 2011 until the present. In this period the identified literature has a result equal to 451 documents, distributed unevenly along the time axis, with greater concentration in year 2018 (Figure 2).

4.2. Step#2: Screening

Starting from the result obtained in the first step, in the screening step, documents characterized by open access were chosen to be analyzed, excluding those with restrictions, from 2011 to 2019. With this step, the number of documents dropped dramatically. Thus, 38 documents are obtained from Scopus, while 29 documents are obtained from the Web of Science (Table 4).

Table 4: Results Of OA In Time Period 2011-2019

Results OA			
	OA	Time Horizon	
SCP	38	From 2011	To 2019
WoS	29		
Total	67		

This is a decrease of 87% and 83% respectively.

4.3. Step#3: Inclusion

The last step is dedicated to finalizing the final sample. Trying to give an overview of the topics and areas of the interface, filters have been applied in relation to the thematic areas to which they belong, shown in Table 5.

Table 5: Research Filters And Results Of Filters Application

Filters	
SCP	WoS
Engineering	Engineering Industrial
Computer Science	Engineering Manufacturing
Material Science	Computer Science Information System
Energy	Computer Science Interdisciplinary Applications
Business, Management and Accounting	Engineering Electrical Electronic
Chemical Engineering	Engineering Mechanical
Decision Science	Engineering Multidisciplinary
-	Telecommunication
-	Automation Control System
-	Computer Science AI
-	Engineering Civil
Results	
29 documents	18 documents

The result is another decrease in the documents number of our interest, obtaining 29 documents on Scopus and 18 on Web of Science (Table 5).
Finally, they were analyzed to exclude redundancies or document overlaps: at the end of phase 1, the sample to be analyzed consists of 37 total documents (Table A).

5. BIBLIOMETRIC ANALYSIS: PHASE 2 - ANALYSIS

This section presents and discusses the findings of this review.

First, an overview of the selected studies are presented. Second, the review findings according to the research criteria, one by one in the separate subsections, are reported.

5.1. Document Types

Before proceeding to the document's analysis, it is appropriate to establish their type. In fact, the 37 documents composing the sample are divided into two types: 17 are articles and 20 are conference papers, 46% and 54% respectively.

5.2. Publication by Years

The analysis of the years of publication shows that the research was not particularly intense in the years following the introduction of I4.0. In fact, it seems the research is peaking only recently, with documents concentration in year 2018, as shown in Table 6.

Table 6: Top Publications By Years

Results	
Year	TP
2011	1
2012	0
2013	0
2014	2
2015	3
2016	1
2017	6
2018	22
2019	2

In 2019 there is a small number, since the research was carried out at the end of the first four months. However, tracing a line representing the trend (Figure 2) it is plausible to think that there is a large number of documents.

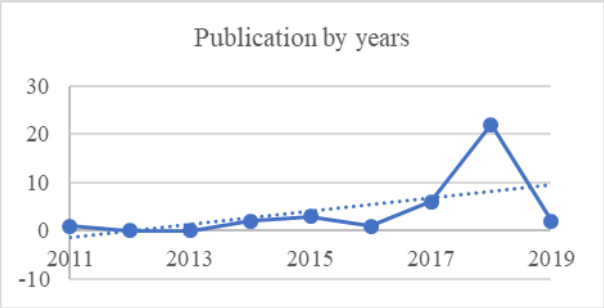


Figure 2: Publication By Years

5.3. Citation by Years

Compared to the years of publication, it is interesting to know the number of citations.

Table 7 shows the citations number for each year of publication and the graph relating to the table is shown in Figure 3.

Table 7: Top Citation By Years

Results	
Year	TC
2011	27
2012	0
2013	0
2014	9
2015	79
2016	3
2017	61
2018	10
2019	0

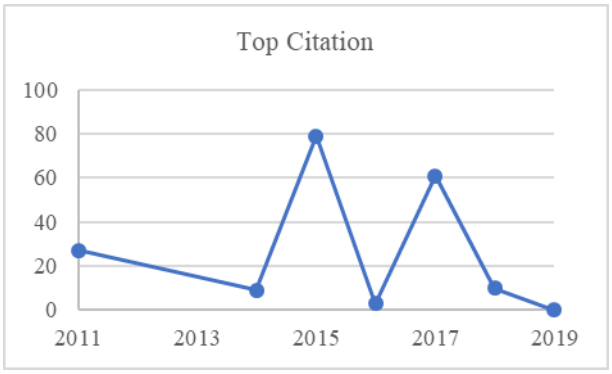


Figure 3: Top Citation By Years

The first aspect that emerges is the citations number in years 2015 and 2017, respectively with 79 and 61 citations. This can be interpreted as the source of the publications produced in year 2018, which is the most productive one (Figure 4).

One of the documents of 2015 is produced in Italy (Table A, ID4). Specifically, this includes 43 citations out of 79 totals. The document proposes a Predictive Maintenance methodology, that allows dynamical decision rules to be adopted for maintenance management and can be used with high-dimensional and censored data problems. Furthermore, the effectiveness of the methodology is demonstrated using a simulated example and a benchmark semiconductor manufacturing maintenance problem (Susto and Schirru and Pampuri and McLoone and Beghi 2015). Indeed, this is the document that counts the greatest number of citations of the whole sample (Table A).

5.4. Country Analysis

Referring to the previous paragraph, it is useful to understand which countries are more productive.

From the analysis, it is clear (Figure 5) that the countries most interested in the topic are China, USA and Italy, respectively with 7, 6 and 4 documents.

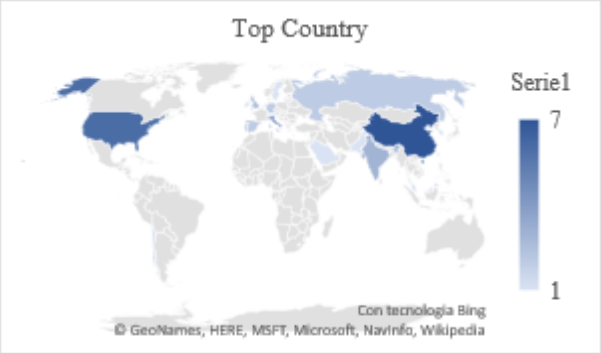


Figure 4: Top Country Analysis

Looking at Europe, it can be determined that it is the most productive, with an overall result of 19 documents, which means a 51% contribution.

5.5. Research Areas Analysis

According with the thematic areas used as a filter reported in Table 5 (paragraph 4.3) and coherently with the topic under examination, the analysis of the research areas revealed that the higher interest areas are "Engineering" and "Computer Science" (Table 8) with 24 and 16 documents respectively.

Table 8: Top Publications By Research Areas

Research	
Research Areas	TP
Engineering	24
Computer Science	16
Mathematics	7
Materials Science	5
Energy	4
Physics and Astronomy	4
Biochemistry, Genetics and Molecular Biology	2
Agricultural and Biological Sciences	1
Business, Management and Accounting	1
Decision Sciences	1
Economics, Econometrics and Finance	1
Social Sciences	1

The other areas also show interest, but with a much smaller number, as shown in Table 8.

5.6. Top Source Journal Analysis

In this section top 10 source or journals which are publishing most frequently are extracted. The total source journal detected from the document is 23 but, considering the top 10, given the source frequency distribution, Figure 6 shows that only the first 7 sources have more than one paper published with a percentage contribute of 57% on the total.

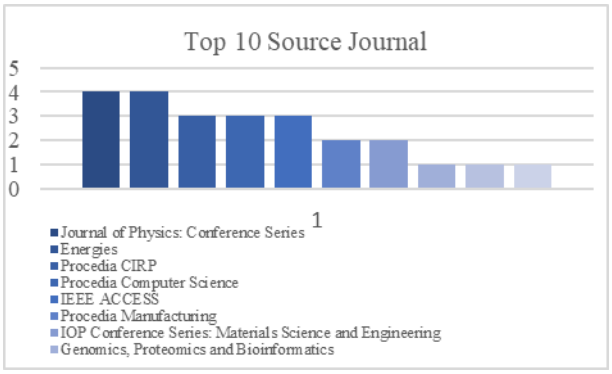


Figure 5: Top 10 Source Journal Analysis

Wanting to analyze the sources separately, the results obtained in the two databases are not the same. Referring to Table A, in Scopus the top source journal is "Journal of Physics: Conference Series" with 4 publications and a contribute of 11% of the total. In Web of Science the top source journal is "ENERGIES", with 3 publications and a contribute of 8%.
Aggregating the data collected from the two databases, the ranking moves to that obtained by Scopus, moving "ENERGIES" to second place (Figure 5). After, 3 source journals have 3 publications and 2 others have 2. The rest of the publications have a one-to-one relationship with the corresponding source journal: for this reason only some of them have been shown in Figure 6.
The low level of concentration of the sources suggests that there is a great deal of interest on these topics by several scientific journals.

5.7. Most Collaborative Authors

From the analysis carried out, it can be observed that all authors are authors or co-authors of a single article, except for Gao Robert X. (Table A, ID 8 and ID 26), who have published two documents.
The most useful analysis in this case is represented by a relationship between authors and citation. Only the top 20 authors are shown in Table 9, which represents a comparison between the number of publications (TP) with the number of citations (TC) associated with each author.

Table 9: Top 20 Authors Analysis

Results		
Top 20 Authors	TP	TC
Beghi, A.	1	43
Martins, T.	1	43
Oneto, L.	1	43
Pampuri, S.	1	43
Schirru, A.	1	43
He, Z.	1	27
Sun, C.	1	27
Zhang, Z.	1	27
Gao, R.X.	2	23
Jennings, C.	1	23
Kumara, S.	1	23
McLoone, S.	1	23

Miao, Q.	1	23
Susto, G.A.	1	23
Terpenny, J.	1	23
Triphathi, A.K.	1	23
Tsui, K.-L.,	1	23
Wang, C.	1	23
Wang, D.	1	23
Wu, D.	1	23

However, there are other researchers who are authors of two documents, but the number of citations associated with them is lower or even zero. For this reason they are not among the top 20 authors and therefore do not appear in Table 9. For completeness, the authors of two documents and their associated citations number are shown in Table 10.

Table 10: Authors Of 2 Documents

Results		
Authors	TP	TC
Gao, R.X.	2	23
Sahli, H.	2	14
Liu, Ling	2	1
Praça, I.	2	0

Starting from Table 9, it can represent the SNA, through which the link between publications number and citations number of each author is shown.

The graph shown in Figure 7, nodes and leaves can be identified. The nodes are represented by the blue square and correspond to publications number (TP) and citations number (TC). Red circles are the leaves and represent the authors' name. Each leaf is connected to both nodes with a different weight.

The weight of the connection between the authors and TP node is not showed, but the line that has a weight equal to 2.0 is highlighted. This line, colored with blue (Figure 7), is the connection between the author Gao R.X. and the number of documents produced by him. All other connections, consistent with Table 9, is equal to 1.0.

The weight that links the authors and the citations number, on the contrary, is reported in Figure 6. The highest weights among all the values shown in Table 9 are highlighted by the green lines and range from 43.0 to 27.0. Weight that is 23.0 is not highlighted.

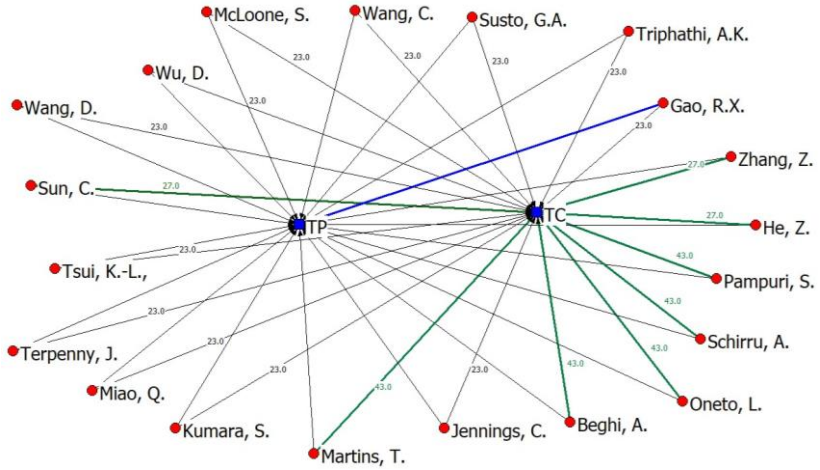


Figure 6: Publications And Citations' Connection For Each Authors

5.8. Top Most Keywords Analysis

Almost all the keywords are common in WoS and Scopus extracted documents and use one or more keywords. To analyze the most popular keywords, a software to create word clouds: NVivo 12, produced by QSR International Pty Ltd ® was used.

The word cloud in Figure 7 highlights the keywords most used by the authors to index their publications.

As shown in Figure 8, the largest words are "machine", "learning", "maintenance", "predictive", "system" and other words. These are the words that appear several times in the documents. For example, the word "learning" appears 14 times and "machine" appears 13 times. Top 20 keywords frequency (named F in Table 11) that appeared in the documents analyzed are shown in Table11.

Figure 7: Keywords Cloud Contribute By NVivo 12

Table 11: Top 20 Keywords Frequency

Results of Top 20 Keywords Frequency (F)			
Keywords	F	Keywords	F
Learning	14	Analytics	4
Machine	13	System	4
Data	11	Health	4
Maintenance	9	Prediction	3
Predictive	9	Industry	3
Monitoring	5	mining	3
Condition	5	fault	3
Analysis	5	identification	2
Prognostics	4	classification	2
Management	4	manufacturing	2

As the words frequency decreases, the size of words in the cloud also decreases. Therefore, the words that are very small in the cloud are those that appear only once.

6. BIBLIOMETRIC ANALYSIS: PHASE 3 - DISCUSSION

Interesting information emerged from the analysis carried out.

The first observation concerns the number of documents. The low number of documents available from 2011 to nowadays shows how the joint topic is still very young and unripe, but the sudden growth that has affected the last few years is indicative of an interest that has been made more alive. However, the technologies are not at all young (just think that the ML dates back to the late 1950s) but have not undergone such a rapid technological process as in recent years, with I4.0. In this regard, it is recalled that throughout history AI technology has undergone the so-called "AI winter" (this term first pronounced in 1984 by Roger Schank and Marvin Minsky and appeared as the topic of a public debate at the annual meeting of American Association of Artificial Intelligence).

Additional information that can be extracted from the analysis concerns the research area. Although Artificial Intelligence (and therefore ML) is a branch of Computer Science, the research area with the largest number of publications is Engineering. The reason is linked to the very high interest of Machine Learning for Predictive Maintenance interested in engineering applications.

Furthermore, it can be said that the countries most interested in scientific research are Europe, China and USA, which are the most industrialized countries.

With technological growth, the geographical spread of academic interest but also of applications is not excluded.

It is important to underline that this document was produced using only two databases. These are WoS and Scopus, in which only documents with open access were included.

7. CONCLUSION

This document focused on the study of the current state-of-the-art in Machine Learning for Predictive Maintenance topics. To date, with the tools available to scientific community, the literature available on any subject is very wide. As a result, complete coverage of all documents published in relation to a specific topic can be very difficult. Therefore, a selection of the most relevant literature was made and a document was produced, that provides a review of the applications in various scientific fields, using ML techniques. For the selection of documents, objective and clear investigation methods were used, independent of the researchers' experience. Among the objectives of the document there is not only the wish to provide a complete picture on research literature, but also a starting point for integrating knowledge through research in this field, suggesting future research paths. There are therefore many other documents with limited access and other indexing databases, such as Google Scholar, which could be integrated into the research in the near future.

ACKNOWLEDGMENTS

This work is the result of the Italian project "*Linee Guida per I4.0 – Campania*" - funded by Regione Campania within POR FSE 2014-2020 Asse IV "Capacità istituzionale e amministrativa" objectives 18 (RA) 11.3 and 21 (RA) 11.6

REFERENCES

Petrillo, A., De Felice, F., Zomparelli, F., 2019. Performance measurement for world class manufacturing: a model for Italian automotive industry. Total Quality Management & Business Excellence, Pages 908-935, Volume 30, 2019 - Issue 7-8.

Fernandes M., Canito A., Bolón-Canedo V., Conceição L., Praça I., Marreiros G., 2019. Data analysis and feature selection for predictive maintenance: a case-study in the metallurgic industry. International Journal of Information Management, 46, 252-262.

Lee C.Y., Huang T.S., Liu M.K., Lan C.Y., 2019. Data science for vibration heteroscedasticity and predictive maintenance of rotary bearings. Energies, 12 (5), art. no. 801.

Serio L., Antonello F., Baraldi P., Castellano A., Gentile U., Zio E., 2018. A smart framework for the availability and reliability assessment and management of accelerators technical facilities. Journal of Physics: Conference Series, 1067 (7), art. no. 072029.

Kleftogiannis D., Ashoor H., Bajic V.B., 2018. TELS: a novel computational framework for identifying motif signatures of transcribed enhancers. Genomics, Proteomics and Bioinformatics, 16 (5), 332-341.

Bedi J., Toshniwal D., 2018. Empirical mode decomposition based deep learning for electricity demand forecasting. IEEE Access, 6, 49144-49156.

Shafi U., Safi A., Shahid A.R., Ziauddin S., Saleem M.Q., 2018. Vehicle remote health monitoring and

prognostic maintenance system. Journal of Advanced Transportation, 2018, art. no. 8061514.

Pandit R., Infield D., 2018. Gaussian process operational curves for wind turbine condition monitoring. Energies, 11 (7), art. no. 1631.

Ullah I., Yang F., Khan R., Liu L., Yang H., Gao B., Sun K., 2017. Predictive maintenance of power substation equipment by infrared thermography using a machine-learning approach. Energies, 10 (12), art. no. 1987.

Wang D., Tsui K.-L., Miao Q., 2017. Prognostics and health management: a review of vibration based bearing and gear health indicators. IEEE Access, 6, 665-676.

Wilson D.L., Coyle J.R., Thomas E.A., 2017. Ensemble machine learning and forecasting can achieve 99% uptime for rural handpumps. PLoS ONE, 12 (11), art. no. e0188808.

Wu D., Jennings C., Terpenny J., Gao R.X., Kumara S., 2017. A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 139 (7), art. no. 071018.

Zhou F., Gao Y., Wen C., 2017. A novel multimode fault classification method based on deep learning. Journal Of Control Science And Engineering, 2017, art. no. 3583610.

Wang C., Liu X., Liu H., Chen Z., 2016. A fault diagnostic method for position sensor of switched reluctance wind generator, Journal Of Electrical Engineering And Technology, 11 (1), 29-37.

Cartella F., Lemeire J., Dimiccoli L., Sahli H., 2015. Hidden semi-markov models for predictive maintenance. Mathematical Problems in Engineering, 2015, art. no. 278120.

Tulabandhula T., Rudin C., 2014. On combining machine learning with decision making. Machine Learning, 97 (1-2), 33-64.

Susto, G.A., Schirru A., Pampuri S., McLoone S., Beghi A., 2015. Machine learning for predictive maintenance: a multiple classifier approach, IEEE Transactions on Industrial Informatics, 11 (3), 812-820.

Leahy K., Gallagher C., O'Donovan P., Bruton K., O'Sullivan D.T.J., 2018. A robust prescriptive framework and performance metric for diagnosing and predicting wind turbine faults based on scada and alarms data with case study, Energies 11 (7), art. no. 1738.

Peres R.S., Rocha A.D., Leitao P., Barata J., 2018. IDARTS - Towards intelligent data analysis and real-time supervision for industry 4.0, Computers in Industry, 101, 138-146.

Wu Z.Y., Luo H., Yang Y.N., Lv P., Zhu X.N., Ji Y., Wu B.A., 2018. K-PdM: KPI-oriented machinery deterioration estimation framework for predictive maintenance using cluster-based hidden markov model. IEEE Access, 6, 41676-41687.

Mohamudally N., Peermamode-Mohaboob M., 2018. Building an anomaly detection engine (ade) for iot smart applications. Procedia Computer Science, 134, 10-17.

Uhlmann E., Pontes R.P., Geisert C., Hohwieler E., 2018. Cluster identification of sensor data for predictive maintenance in a Selective Laser Melting machine tool. Procedia Manufacturing, 24, 60-65.

Wegener K., Gittler T., Weiss L., 2018. Dawn of new machining concepts: compensated, intelligent, bioinspired. Procedia CIRP, 77, 1-17.

Zhang J., Wang P., Yan R., Gao R.X., 2018. Deep Learning for improved system remaining life prediction. Procedia CIRP, 72, 1033-1038.

Zuev D., Kalistratov A., Zuev A., 2018. Machine Learning in IT Service Management. Procedia Computer Science, 145 (2018) 675–679.

Schmidt B., Wang L., 2018. Predictive Maintenance of machine tool linear axes: a case from manufacturing industry. Procedia Manufacturing, 17, 118-125.

Pandey S.K., Mishra R.B., Triphathi A.K., 2018. Software bug prediction prototype using bayesian network classifier: a comprehensive model. Procedia Computer Science, 132, 1412-1421.

Chakravorti N., Rahman M.M., Sidoumou M.R., (...), Gosewehr F., Wermann J., 2018. Validation of PERFoRM reference architecture demonstrating an application of data mining for predicting machine failure. Procedia CIRP, 72, 1339-1344.

Cortes C., Vapnik V., 1995. Support-vector networks. Machine Learning, 20 (3), 273-297.

Kotsiantis S. B., Zaharakis I., Pintelas P., 2007. Supervised machine learning: a review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, 3-24.

Marasco D.E., Kontokosta C.E., 2016. Applications of machine learning methods to identifying and predicting building retrofit opportunities. Energy and Buildings, 128, 431-441.

Lamirel J.-C., Al Shehabi S., Hoffmann, M., François C., 2003. Intelligent patent analysis through the use of a neural network. Proceedings of the ACL-2003 Workshop on Patent Corpus Processing, 20, 7–23.

Markovic D., Petkovic D., Nikolic V., Milovancevic M., Petkovic B., 2017. Soft computing prediction of economic growth based in science and technology factors, Phys. Stat. Mech. Appl. 465, 217–220.

Choi J., Jang D., Jun S., Park S., 2015. A predictive model of technology transfer using patent analysis. Sustainability, 7 (12), 16175-16195.

Searle, J. R., 1980. Minds, brains, and programs. Behavioral and Brain Sciences 3 (3), 417-457.

Samuel A. L., 1959. Some studies in machine learning using the game of checkers. BM Journal of

Research and Development archive, 3 (3), 210-229.

Praveenkumar T., Saimurugan M., Krishnakumar P., Ramachandran K.I., 2014. Fault diagnosis of automobile gearbox based on machine learning techniques. 12th Global Congress on Manufacturing and Management (GCMM), pp. 2092-2098. December 08-10, Vellore (India).

Fumeo E., Oneto L., Anguita D., 2015. Condition based maintenance in railway transportation system based on Big Data streaming analysis. INN Conference on Big Data, pp. 437-446. August 08-10, San Francisco (California, USA).

Ali S.M., Hui K.H., Hee L.M., Salman Leong M., Al-Obaidi M.A., Ali Y.H., Abdelrhman A.M., 2018. A comparative experimental study on the use of machine learning approaches for automated valve monitoring based on acoustic emission parameters. 3rd International Conference on Mechanical, Manufacturing and Process Plant Engineering (ICMMPE), UNSP 012032. November 22-23, Penang (Malaysia).

Li X., Wei L., He J., 2018. Design and implementation of equipment maintenance predictive model based on machine learning. 2nd annual international conference on cloud technology and communication engineering, UNSP 012001. August 17-19, Nanjing (China).

Koltsidopoulos P.A., Dawood T., Thies P.R., 2018. Data insights from an offshore wind turbine gearbox replacement. EERA DeepWind'2018, 15th Deep Sea Offshore Wind R&D Conference, art. no. 012003. January 17-19, Trondheim (Norway).

Elshrif M., Rizzo S.G., Betz F.D., Margineantu D.D., Zaki M.J., Chawla S., 2018. Embeddings for the Identification of Aircraft Faults (MERIT). 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), art. no. 8448927. June 11-13, Seattle (Washington, USA).

Zanelli D., López E., Pavez C., Pedreros J., Jain J., Avaria G., Moreno J., Bora B., Davis S., Soto L., 2018. Analysis of deterioration in a plasma focus device. XX Chilean Physics Symposium, November 30-December 2, Santiago (Chile).

Katona A., Panfilov P., 2018. Building predictive maintenance framework for smart environment application systems. 29th International DAAM Symposium "Intelligent Manufacturing & Automation", October 24-27, Zadar (Croatia).

Morales F.J., Reyes A., Cáceres N., Romero L.M., Benitez F.G., Morgado J., Duarte E., Martins T., 2017. Historical maintenance relevant information road-map for a self-learning maintenance prediction procedural approach. Building up Efficient and Sustainable Transport Infrastructure 2017 (BESTInfra2017), September 21-11, Prague (Czech Republic).

Sun C., Zhang Z., He Z., 2011. Research on bearing life prediction based on support vector machine and its application. 9th International conference on demage assessment of structure (DAMAS), July 11-13, Oxford (UK).

Mobley R. K., 2002. An Introduction to Predictive Maintenance. 2nd ed. Butterworth-Heineman.

APPENDIX

Table A: Extracted Documents From Scopus And Web of Science

ID	Source	Title	Authors	Years	Source Title	Cit.	Country
1	SCP	Research on bearing life prediction based on support vector machine and its application	Sun C., Zhang Z., He Z.	2011	Journal Of Physics: Conference Series	27	China
2	WoS	Fault diagnosis of automobile gearbox based on machine learning techniques	Praveenkumar T., Saimurugan M., Krishnakumar P., Ramachandran K.I.	2014	12th Global Congress On Manufacturing And Management (GCMM - 2014)	9	India
3	WoS	On combining machine learning with decision making	Tulabandhula T., Rudin, C.	2014	Machine Learning	1	USA
4	WoS	Machine Learning for Predictive Maintenance: A Multiple Classifier Approach	Susto G.A., Schirru A., Pampuri S., McLoone S., Beghi A.	2015	IEEE Transactions On Industrial Informatics	43	Italy
5	WoS	Condition Based Maintenance in Railway Transportation Systems Based on Big Data Streaming Analysis	Fumeo E., Oneto L., Anguita D.	2015	Inns Conference On Big Data 2015 Program	22	Italy
6	SCP	Hidden semi-markov models for predictive maintenance	Cartella F., Lemeire J., Dimiccoli L., Sahli H.	2015	Mathematical Problems In Engineering	14	Belgium
7	WoS	A Fault Diagnostic Method for Position Sensor of Switched Reluctance Wind Generator	Wang C., Liu X., Liu H., Chen Z.	2016	Journal Of Electrical Engineering & Technology	3	South Korea
8	WoS	A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests	Wu D., Jennings C., Terpenny J., Gao R.X., Kumara S.	2017	Journal Of Manufacturing Science And Engineering-Transactions Of The Asme	23	USA
9	SCP	Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators	Wang D., Tsui K.L., Miao Q.	2017	IEEE Access	23	Hong Kong, China
10	SCP	A Novel Multimode Fault Classification Method Based on Deep Learning	Zhou F., Gao Y., Wen C.	2017	Journal Of Control Science And Engineering	11	China
11	SCP	Ensemble machine learning and forecasting can achieve 99% uptime for rural handpumps	Wilson D.L., Coyle J.R., Thomas E.A.	2017	PLoS ONE	3	USA
12	WoS	Predictive Maintenance of Power Substation Equipment by Infrared Thermography Using a Machine-Learning Approach	Ullah I., Yang F., Khan R., Liu L., Yang H., Gao B., Sun K.	2017	Energies	1	China
13	SCP	Historical maintenance relevant information road-map for a self-learning maintenance prediction procedural approach	Morales F.J., Reyes A., Cáceres N., (...), Duarte E., Martins T.	2017	IOP Conference Series: Materials Science And Engineering	0	Portugal, Spain
14	WoS	A Robust Prescriptive Framework and Performance Metric for Diagnosing and Predicting Wind Turbine Faults Based on SCADA and Alarms Data with Case Study	Leahy, Kevin, Gallagher, Colm, O'Donovan, Peter, Bruton, Ken, O'Sullivan, Dominic T. J.	2018	Energies	6	Ireland
15	SCP	Data Insights from an Offshore Wind Turbine Gearbox Replacement	Koltsidopoulos Papatzimos, A., Dawood, T., Thies, P.R.	2018	Journal Of Physics: Conference Series	1	UK
16	SCP	Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting	Bedi, J., Toshniwal, D.	2018	IEEE Access	1	India
17	WoS	IDARTS - Towards intelligent data analysis and real-time supervision for industry 4.0	Peres, Ricardo Silva, Rocha, Andre Dionisio, Leitao, Paulo, Barata, Jose	2018	Computers In Industry	1	Portugal
18	WoS	Vehicle Remote Health Monitoring and Prognostic Maintenance System	Shafi, Uferah, Safi, Asad, Shahid, Ahmad Raza, Ziauddin, Sheikh, Saleem, Muhammad Qaiser	2018	Journal Of Advanced Transportation	1	Pakistan
19	WoS	A Comparative Experimental Study on the Use of Machine Learning Approaches for Automated Valve Monitoring Based on Acoustic Emission Parameters	Ali, Salah M., Hui, K. H., Hee, L. M., Leong, M. Salman, Al-Obaidi, M. A., Ali, Y. H., Abdelrhman,	2018	3rd International Conference On Mechanical, Manufacturing And	0	Malesia

			Ahmed M.		Process Plant Engineering (ICMMPE 2017)		
20	SCP	A smart framework for the availability and reliability assessment and management of accelerators technical facilities	Serio, L., Antonello, F., Baraldi, P., (...), Gentile, U., Zio, E.	2018	Journal Of Physics: Conference Series	0	Italy
21	SCP	Analysis of Deterioration in a Plasma Focus Device	Zanelli, D., López, E., Pavez, C., (...), Davis, S., Soto, L.	2018	Journal Of Physics: Conference Series	0	Chile
22	SCP	Building An Anomaly Detection Engine (ADE) for IoT Smart Applications	Mohamudally, N., Peermamode-Mohaboob, M.	2018	Procedia Computer Science	0	Mauritius
23	SCP	Building predictive maintenance framework for smart environment application systems	Katona, A., Panfilov, P.	2018	Annals Of DAAAM And Proceedings Of The International DAAAM Symposium	0	Austria, Russian Federation
24	SCP	Cluster identification of sensor data for predictive maintenance in a Selective Laser Melting machine tool	Uhlmann, E., Pontes, R.P., Geisert, C., Hohwieler, E.	2018	Procedia Manufacturing	0	Germany
25	SCP	Dawn of new machining concepts: Compensated, intelligent, bioinspired	Wegener, K., Gittler, T., Weiss, L.	2018	Procedia CIRP	0	Switzerland
26	SCP	Deep Learning for Improved System Remaining Life Prediction	Zhang, J., Wang, P., Yan, R., Gao, R.X.	2018	Procedia CIRP	0	USA
27	SCP	Design and Implementation of Equipment Maintenance Predictive Model Based on Machine Learning	Li, X., Wei, L., He, J.	2018	IOP Conference Series: Materials Science And Engineering	0	China
28	SCP	Embeddings for the Identification of Aircraft Faults (MERIT)	Elshrif, M., Rizzo, S.G., Betz, F.D., (...), Zaki, M.J., Chawla, S.	2018	2018 IEEE International Conference On Prognostics And Health Management, ICPHM 2018	0	Qatar, Italy, USA
29	WoS	Gaussian Process Operational Curves for Wind Turbine Condition Monitoring	Pandit, Ravi, Infield, David	2018	Energies	0	Scotland
30	WoS	K-PdM: KPI-Oriented Machinery Deterioration Estimation Framework for Predictive Maintenance Using Cluster-Based Hidden Markov Model	Wu, Zhenyu, Luo, Hao, Yang, Yunong, Lv, Peng, Zhu, Xinning, Ji, Yang, Wu, Bian	2018	IEEE Access	0	China
31	SCP	Machine Learning in IT Service Management	Zuev, D., Kalistratov, A., Zuev, A.	2018	Procedia Computer Science	0	China, Russian Federation
32	SCP	Predictive Maintenance of Machine Tool Linear Axes: A Case from Manufacturing Industry	Schmidt, B., Wang, L.	2018	Procedia Manufacturing	0	Sweden
33	SCP	Software Bug Prediction Prototype Using Bayesian Network Classifier: A Comprehensive Model	Pandey, S.K., Mishra, R.B., Triphathi, A.K.	2018	Procedia Computer Science	0	India
34	SCP	TELS: A Novel Computational Framework for Identifying Motif Signatures of Transcribed Enhancers	Kleftogiannis, D., Ashoor, H., Bajic, V.B.	2018	Genomics, Proteomics And Bioinformatics	0	UK, USA, Saudi Arabia
35	SCP	Validation of PERFoRM reference architecture demonstrating an application of data mining for predicting machine failure	Chakravorti, N., Rahman, M.M., Sidoumou, M.R., (...), Gosewehr, F., Wermann, J.	2018	Procedia CIRP	0	UK, Germany
36	SCP	Data analysis and feature selection for predictive maintenance: A case-study in the metallurgic industry	Fernandes, M., Canito, A., Bolón-Canedo, V., (...), Praça, I., Marreiros, G.	2019	International Journal Of Information Management	0	Portugal, Spain
37	SCP	Data science for vibration heteroscedasticity and predictive maintenance of rotary bearings	Lee, C.-Y., Huang, T.-S., Liu, M.-K., Lan, C.-Y.	2019	Energies	0	Taiwan

INTEGRATION OF PROCESS MINING TECHNIQUES IN SIMULATION RESULTS ANALYSIS

Irīna Šitova[(a)], Jeļena Pečerska[(b)]

[(a), (b)] Riga Technical University, Riga, Latvia

[(a)]irina.sitova@rtu.lv, [(b)]jelena.pecerska@rtu.lv

ABSTRACT

The research is carried out in the area of analysis of simulation results. The aim of this research is to explore the applicability of process mining techniques, and to introduce the process mining techniques integration into results analysis of discrete-event system simulations. As soon as the dynamic discrete-event system simulation (DESS) is based on events list or calendar, most of simulators provide the events lists. These events lists are interpreted as event logs in this research, and are used for process mining. The information from the events list is analysed to extract process-related information and perform in-depth process analysis. Event log analysis verified applicability of the proposed approach. Based on the results of this research, it can be concluded that process mining techniques in simulation results analysis provide a possibility to reveal new knowledge about the performance of the system, and to find the parameter values providing the advisable performance.

Keywords: discrete-event system simulation, process mining, queueing system, simulation results analysis

1. INTRODUCTION

Simulation is one of operations research approaches that is realized in form of software tools. A simulation is the imitation of a complex system that gives a possibility to display and describe the behaviour of a system in a detailed way (Pidd 2004). In the present paper, the discussion is limited to the DESS approach, further referenced as "simulation". Simulations are most effectively used for the analysis of dynamics of complex artificial material systems. This choice is a reasonable one, as soon as this approach aims to investigate business, manufacturing, service and other processes. Researchers simulate processes to gain insight into the operation of systems. In the process of simulation, random factors influencing the system are taken into account, as well as their changes over time. Researchers obtain information about the system under consideration after experiments with a verified, calibrated and validated model (Law and Kelton 2000).

Technically the approach is implemented with the software tools; most of them simulate artificial event logs. The event log in form of events lists or events calendars is a mean of model entity management. Simulation is an important tool to explain how process performance indicators react in the face of controllable factors and environmental factors (Banks et al. 2010). Any changes of the system state during the process execution are recorded in the simulation events lists. Simulation results after running models of this type make it possible to obtain estimates of various performance measures: productivity and throughput measures, resource utility measures, and service level measures (Merkurjevs et al. 2008). The obtained simulation results are used for understanding the behaviour of the system, to formulate forecasts, to compare alternatives or to solve the optimization tasks of system parameters. Events lists usually are not analysed as a simulation outcome.

Process mining is described by its authors as a tool to extract non-trivial and useful information from process execution logs. Process mining uses event data to extract the information. This discipline is built on process model-driven approaches and data mining, and is used to support process improvements (van der Aalst 2011).

Thus the correspondence with the process mining tool in the area of the research, as well as in using natural or simulated event logs, and implementation goals – process improvement – is detected. The event logs in process mining are used as the source of information about the process and may be aimed on simulation models discovery. In DESS the models are based on the information about the process, models are realised through simulation of the events, organized as events lists, and produce simulation results.

The paper proposes to use a combined approach to the analysis of the outputs, both simulation results and events lists, of the simulation model, as shown in Figure 1.

The findings in the area, authors' previous work and the basics of the proposed approach are provided in the following sections.

2. RELATED WORK

There are a considerable number of articles and projects concerning application of process mining techniques for analysis of various process types.

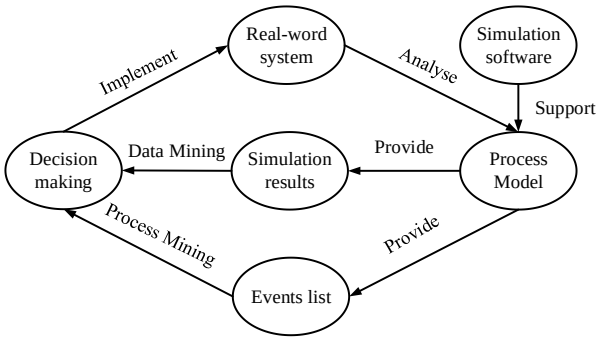


Figure 1: Process Mining for Simulation Outcomes Analysis

Some of them deal with efficient resource utilization in manufacturing processes, business process organization in warehousing, insights in human behaviour from realization of smart house processes and business rules discovery from manufacturing processes. Some of the most characteristic publications are considered in this section.

2.1. Analysis of Resource Behaviour

Authors in (Nakatumba and van der Aalst 2009) developed an approach that applies process mining techniques to provide detailed insights analysing resource behaviour. They investigate the relationship between workload and services time using linear regression analysis on historic data from Process-Aware Information Systems (PAISs). Researchers note a strong connection between workload and performance of workers which is confirmation of the "Yerkes-Dodson Law of Arousal". The authors used real-life logs to validate their approach using implemented new plug-in in ProM. The authors showed how better resource description helps to create more realistic and precise simulation models and provides better work allocation.

2.2. Analysis of Warehouse

Haav and Kalja in (Haav and Kalja 2014) presented the application of automatic process mining for solving a real warehousing problem. The goal of the research was to map a specific business problem to the relevant process mining problem definition, perform the analysis and translate the technical results into business insights for the evaluation. The goal was reached on the basis of the case study results. The case study was performed on data mined from the Enterprise Resource Planning (ERP) system of a logistics company, rendering warehousing services. The authors applied the developed methodology for extracting, transforming and loading the data using ProM software. Two process mining algorithms – Alpha and HeuristicMiner were used for data analysis. Thanks to the experiment the company, which provided data, received insights that helped to adjust the strategy and operations of logistic enterprise.

2.3. Analysis of Daily Life

The method described in (Tax et al. 2018) provided extracting insights in (un)healthy living habits from smart home environment data. To deal with overgeneralizing process models abstraction of the events to a higher-level interpretation was made in order to reach better understanding for analytics. For abstraction from sensor-level to human activity purpose framework using supervised learning methods that is based on the XES IEEE standard for event logs were developed. This enables direct analysis of process models showing real human behaviour. In confirmation of the correctness of the methodology the authors demonstrate the added value on the example of the case study.

2.4. Analysis of Semiconductor Manufacturing

Another research in the area (Khemiri et al. 2018) showed an example of use of data mining algorithm for discovery of business rules in semiconductor manufacturing process inspired by process mining ideas. Authors explained the importance of the topic with the lack of knowledge of the operating rules in industrial companies, which leads to not the best possible production performance. The authors, using the case study results, proved that the developed method helps to gain knowledge about the existing rules, to detect problems in the system by using process modelling and serve as a base to knowledge capitalization.

2.5. Current Findings and Future Development

Comparing the approaches from the researches described above, one can conclude that the process mining is aimed at:

- Data-based facilitation of process models creation or improvement (Nakatumba and van der Aalst 2009);
- Using ProM tool for data filtering, analysis and visualization (Haav and Kalja 2014);
- Extracting event log from real life process (Tax et al. 2018);
- Revealing business rules from the output decision trees (Khemiri et al. 2018).

The goal of process mining is to use event data to extract process-related information (van der Aalst 2011). The DESS approach deals with the simulated processes that are producing event data. The objective of creation and use of simulation models is in-depth analysis of the simulated processes and obtaining knowledge about these processes. The generic goal of the simulation studies is transformation of obtained knowledge into actions to improve the performance indicators or structure of the investigated processes. There is a clear similarity of the goals of process mining and DESS, as well as the common subject of the research – processes – and model-oriented nature of process mining. Thus a logical motivation

arises for adaptation of process mining techniques for simulation results analysis and simulation-based process analysis for process mining purposes.
Taking these similarities into account and expecting potential findings, the authors of the present article in the fourth section propose the approach to simulation results analysis that is based on process mining techniques.

3. THE AUTHORS' PREVIOUS WORK

The present research is a logical continuation of authors' previous research (Šitova and Pečerska 2018). Regarding this it is shortly described below.

3.1. The Developed Approach

The authors (Šitova and Pečerska 2018) provided a review of approaches in the area of analysis of simulation results by using data mining techniques. The goal of the research was to explore the applicability of data mining techniques in the area of simulation result analysis and to introduce an application scheme of data mining techniques for the analysis of simulation results. After analysing the reviewed scientific publications (Brady and Bowden 2001, Brady and Yelling 2005, Painter et al. 2006, Feldkamp et al. 2015, Kibira et al. 2015, Feldkamp et al. 2016), as well as several books (Dunham 2003, Witten and Frank 2005, Han and Kamber 2006, Cassandras and Lafortune 2008) devoted to simulation, data mining and integration of these technologies, the authors had developed their approach. The approach includes the analysis of simulation results using several data mining techniques in the proposed order. As a result of the theoretical study, a two-stage approach was formulated, combining the fundamental principles of data farming and knowledge discovery. In the developed approach, each of the used technologies has its own purpose. Data farming, the purpose of which is to get results from a simulation model that corresponds to some real process. Knowledge discovery, the purpose of which is to find patterns in a simulation model state variable behaviour, in order to uncover knowledge which would otherwise be hidden.
A variety of data mining techniques, including correlation analysis, clustering and several visualization mechanisms of results, were used for the knowledge discovery. The developed approach was applied to the analysis of experimental data of a simple DESS model.
The authors hypothesized, tested and confirmed that data mining techniques may provide a better interpretation of simulation output as well as visualization of outputs. Also, the authors showed that data mining reveals not only trivial knowledge from simulation output. As a result of data mining techniques application in simulation results analysis, the knowledge and decision rules were obtained from simulation results coupled with the relevant visualization, which increases the simulation efficiency. It provides a more versatile output analysis and deals with a potentially huge amount of simulation output data.

The overall scheme of the developed approach application, which reflects data farming and knowledge discovery phases, is shown in Figure 2.

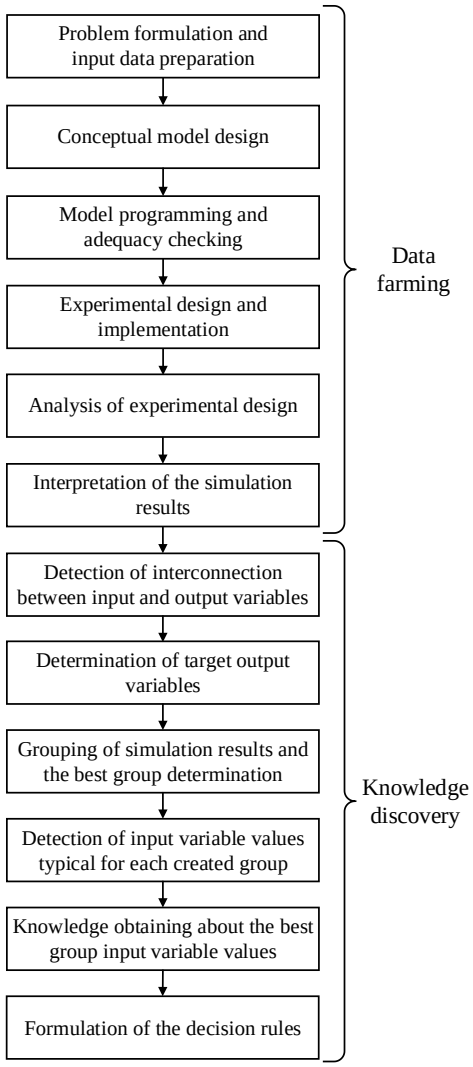


Figure 2: The Developed Approach Application Scheme

3.2. Case Study: Queueing System Simulation and Results Analysis

To achieve the goal of the research (Šitova and Pečerska 2018), the process was consistently implemented as follows. First comes data farming realization: the simulation model of the system under consideration was constructed, simulation experiments were designed, replicated and results were obtained. Second – knowledge discovery realization: the simulation results were analysed using data mining techniques. The domain of acceptable values of experimental factors was revealed and visualised.
To obtain data for mining, 28 model experiment scenarios had been developed with the realistic experimental factor values. Each of the scenarios had been replicated 25 times, providing 700 experiment results sets.
The general scheme of knowledge discovery consists of three steps:

- Correlation analysis of a relationship between input and output variables;
- Clustering algorithm application against target output variables;
- Results visualization with box charts, histograms, radar charts.

Based on the experimental results, the decision principles for the case study problem have been formulated for detecting the input variable values for experimental factors that are specific for each cluster and in particular for the best cluster. The values for the input variables that are experimental factors of the best cluster were defined. Thus the main goal of the case study – the best performance of a queueing system – has been achieved.

In spite of satisfactory results and findings in (Šitova and Pečerska 2018) authors noticed that a whole layer of "additional" results – the events lists data – is not taken into account during the process of knowledge discovery. Therefore, in the present study, the authors decided to conduct knowledge discovery based on data from events lists while data farming remains unchanged. Due to this, the concept of the developed approach has changed as shown in Figure 3. The relevant approach is described in fourth section, as well as the case study is shown in the fifth section.

4. PROCESS MINING IN THE SIMULATION RESULTS ANALYSIS

The present research is a continuation of the previous research described in third section (Šitova and Pečerska 2018). The approach proposed in this research still consists of two stages – data farming and knowledge discovery. However, after analysing the scientific papers (Nakatumba and van der Aalst 2009, Haav and Kalja 2014, Tax et al. 2018, Khemiri et al. 2018, van der Aalst 2012) and books (van der Aalst 2011, van der Aalst 2016) dedicated to process mining and application of process mining techniques, the authors made changes in their approach. These changes affect knowledge discovery stage.

The proposed approach is based on the process mining discipline formulated by van der Aalst in (van der Aalst 2011). The authors consider that not using an additional source of information from DESS models, such as events list, is inconsistent. In this paper, the authors deal with the conventional DESS approach. In most cases, the simulation technique involves the creation and management of events lists to organize and track entities in the simulation model. The term “entity” is used here to designate a unit of traffic. Entities instigate and respond to events (Schriber et al. 2014). Technically, there are several types of events lists involved into the running the simulations. Further, the authors do not delve into the type of events list used; it is assumed that it is possible to use information as necessary. The event logs may be interpreted in a way that is conventional in DESS – a list of particular event notices, ordered by time of occurrence and aggregated if necessary. The typical structure of the DESS model events list is provided as an example in Table 1.

Table 1. The Fragment of the Events List Created by DESS Software

Simulation time	Entity’s ID	Event type
…		
481.56	46	Arrival
481.56	46	End of Job on Gate
481.56	46	End of Job at Decision_0
481.56	46	Moved to Server_3 for service
482.45	47	Arrival
482.45	47	End of Job on Gate
482.45	47	End of Job at Decision_0
482.45	47	Moved to Server_4 for service
483.23	45	End of Job at Server_2
…		

While running the simulations, the events list is managed: events are created, sorted, executed and deleted according to the model structure and logic. For the purposes of the current research, the events lists provide the source for the event log data. Some trivial data saving, sorting and processing actions are not introduced in this paper. These actions are applied to transform the conventional events list into event log.

The concept of the events list-based approach is available in Figure 3. The stages of approach implementation include both data farming and knowledge discovery phases.

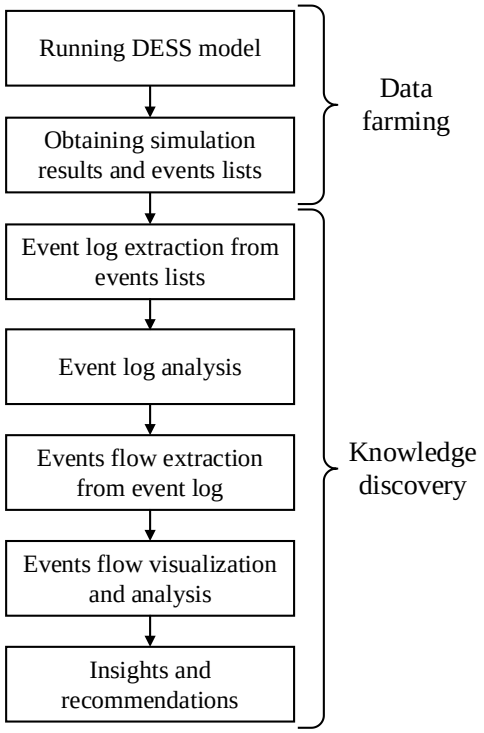


Figure 3: The Stages of the Events List-Based Approach

During data farming stages, the queueing system DESS model provides both simulation results and process history as the record of events list. The events list provides the necessary base for event log extraction, analysis and further processing. The data export from events list provides the possibility to save events flows by event type. For example, in a simple queueing system G/G/*n* (Kendall's notation), there are arrival events, service finish events, as well as other event types, but a particular events list containing only arrival events notices can be used, as well as any particular events list can be constructed. Thus the conventional events list of a simulation model is transformed into particular event logs and further into events flows.

The events flows are introduced as the chronological sequence of the events of the same type, each event is associated with a particular model entity and simulation time. The example of events flow is a sequence of entity arrivals in a service system.

As soon as events are categorized and events flows are described, the knowledge discovery may be implemented not only through visualization, but also through in-depth analysis of interaction of events flows.

The standard procedure for analysis of events flows cannot been proposed, as it depends on the discoveries in the previous stages. The relevant example of events flow analysis is provided in the case study at the following section.

In the proposed approach events list entries provide additional options for simulation result analysis, as well as extending the approach applicability from analysis of the processes in simulation models to analysis of the processes in real systems.

5. CASE STUDY

The following case study tends to conduce to revealing knowledge from simulation results by applying the process mining techniques.

5.1. Problem Statement

One of the fundamental phases of the simulation procedure is an appropriate analysis of simulation results. However, analysing results commonly events list is not taken into account. The events list keeps all the data concerning the changes of the system state. The event log may be easily extracted from events list and transformed into events flow protocol. Further in the text it is assumed as performed action and the existing of event log is fixed. In order to analyse data from the event log process mining techniques could be used. A simple queuing system with a queue of infinite capacity, exponential distribution of interarrival times, variable average value of this exponential distribution at different daytime, random service time, queue-length based probability of arriving customers not served is an object of the case study. Interarrival times and service times are independent random variables. Service discipline is first in first out. Service time is a function of the demand for product. The variable demand for product is used for evaluation of the revenue and described as empirical distribution. Dropout rule is introduced as a function of the queue length. Kendall's notation G/G/*n* queue is relevant for this system. Thus the system performance is not stable, and indicators – service quality and revenue – are conflicting. As a result, it is necessary to formulate a trade-off approach to achieve the best possible performance.

The detailed analytical study of system performance measures is complicated. For the purposes of performance analysis, a discrete-event simulation model of a system is created. Process mining techniques are applied to determine the parameter values for obtaining such performance.

5.2. The Summary of the Research Approach

In the proposed approach data from the events list is used for enriching opportunities in the area of simulation result analysis for getting extra insights from events list. The research approach combines the fundamental techniques of data farming and knowledge discovery. In the data farming phase, the simulation model is constructed, simulation experiments are designed, replicated and results obtained. Knowledge discovery phase is based on process mining ideas and implemented as follows. The conventional events list of a simulation model is transformed into particular event logs and further into events flows. From that insights and recommendations concerning acceptable values of experimental factors are revealed and visualized.

5.3. Comparison of Case Studies

In the previous research (Šitova and Pečerska 2018) the set of system parameters values was obtained by applying data mining techniques to simulation results from several replications. The simulation model in this case study used for data farming was the same as in the current research. Further this case study is referred as Case Study 1 while the current case study as Case Study 2.

Taking into account the fact that the same model was analysed during the knowledge discovery phase, it is possible to make a relevant comparison of both case studies. Comparative information is provided in Table 2. A detailed description of Case Study 2 realization is provided in the following sections.

5.4. Simulation Model of a Queuing System

For the purposes of performance analysis of the system, a discrete-event simulation model is used, providing the estimates of the relevant performance measures.

Outputs of the simulation model are statistics of server utilization, customer time spent in the system, a volume of satisfied demand, number of services and dropouts, and some other typical performance measures that are described in (Šitova and Pečerska 2017). There are two types of dropouts – working-schedule based and dropout rule-based. There are two experimental factors used for experimentation – number of servers and working schedule.

Table 2. Comparison of Case Studies

Features	Case Study 1	Case Study 2
Type of analysed simulation outputs	Time-persistent and observational simulation results (e.g. number of dropouts; queue length statistics)	Events flow (e.g. flow of served customers; flow of dropouts because of gate being closed)
Techniques applicable for analysis	Data mining techniques: correlation analysis, clustering and visualization	Process mining techniques: events list, event log and events flow extraction and analysis, visualization
Target performance indicator	Composite objective function	Integral value of dropouts flow

The conceptual model of the queueing system (QS) under consideration is shown in Figure 4. The model is created in accordance with the recommendations from (Robinson 2015).

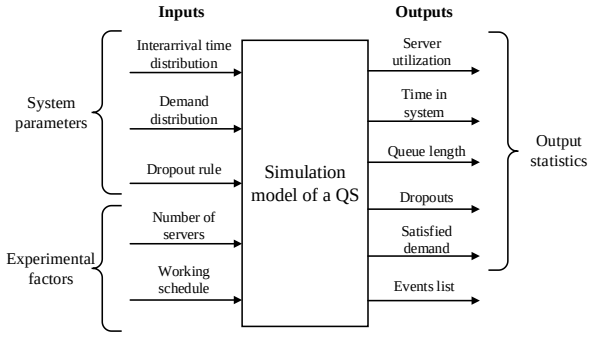


Figure 4: The Conceptual Model for the Case Study

Since this study is a continuation of Case Study 1 the same simulation model is used. Simulation model is created step by step as suggested in (Martin 2018).
The goal of the previous study was achieved, i.e. the combination of input variable values and adjustable working schedule, providing the best values of the composite objective function, was found. Therefore, the simulation model was configured as proposed in the work (Šitova and Pečerska 2018). However, these suggestions concerning experimental factors were not strictly defined. Case Study 1 resulted in proposed number of servers ranges from 2 to 3, the selected number – 3, number of service working hours ranges from 13 to 19, and the selected working schedule includes working times 07:00–15:00 and 16:00–24:00.
Initial simulation model layout as proposed in Case Study 1 is shown in Figure 5. The initial model parameters are selected as proposed in Case Study 1. After running the simulation model, results were

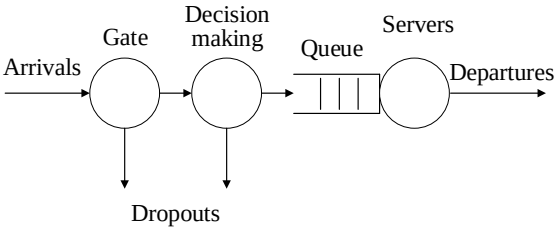


Figure 5: Initial Simulation Model

obtained in the form of events list, which were used for the further analysis.

5.5. Log data transformation: Events list, Event log, Events flow

For the analysis of simulation results, the MS Excel is used as a powerful tool with available visualization. Resulting data from the simulation model in form of events list were uploaded to MS Excel. All further operations were held in it.
Number of possible occurring events in created simulation model is finite. In this case there are 7 different event types. Each event has its ID and description, this information is provided in Table 3.

Table 3. Possible Event Types in Simulation Model

Event's ID	Event description
1	Arrival
2	Crossing gate
3	Start of service
4	End of service
5	Facing closed gate
6	Facing unsatisfying queue size
7	Departure

Events list contains information about simulation time, entity's ID and event type. The full set of events in the list includes records about all cases, associated with them events with certain timestamps i.e. each entity in the system and each change of its transitions in time. Here case, event and timestamp are interpreted as process mining terms in (van der Aalst 2011). Decomposition of the QS process into cases and events is shown in Figure 6.

Case ID	Event ID	Time-stamp
I	1	10:01
	2	10:01
	3	10:01
	4	10:04
	7	10:04
…	…	…
555	1	14:22
	2	14:22
	6	14:22
	7	14:22
…	…	…
999	1	23:07
	5	23:07
	7	23:07
…	…	…

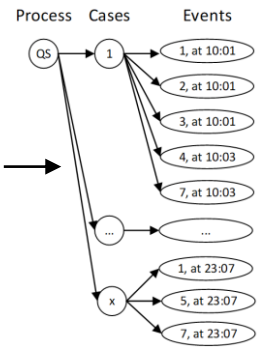


Figure 6: Decomposition of the QS Process

Results in the form of events list displayed as in Table 1 are transformed into event logs according to defined event types played in Table 3. As a result, logs of all event types occurring at a specific time are obtained. After events categorization events flows are described. As soon as the events flows are the chronological sequence of the events of the same type, seven particular events flows are obtained.
Events flow example, namely flow of arriving customers in the system, is shown in Figure 7.

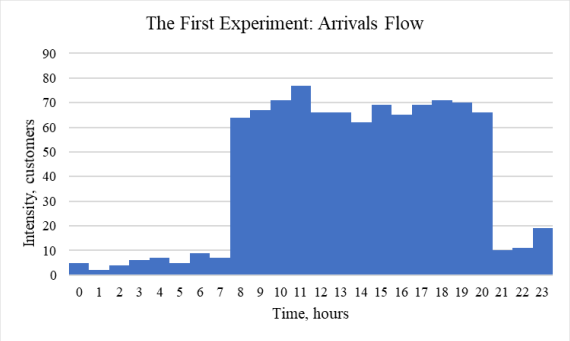


Figure 7: The Flow of Arriving Customers in the System

Analysis of created events flows, and experiments associated with them are discussed in the next subsection.

5.6. Analysis of Events Flows

Analysing arrival flow in Figure 7, it is clear that an inefficient working schedule was chosen for the initial simulation model (hereinafter is referenced as the first experiment). In the first experiment the gate is being closed at 00:00–07:00 and 15:00–16:00, i.e. queueing system is not working. While a decision not to work at night-time is a reasonable one, an hour in the middle of the day is inappropriate. During this hour a large number of potential customers arrives, which are forced to leave the system immediately. Given this fact, the second experiment is conducted. For this experiment the working hours of the system are changed. Based on Figure 7 the working hours are changed to 08:00–21:00. Figure 8 shows changes in dropouts flow after applying the new working schedule.
Figure 8 shows an approximately 25 percent decrease in the number of dropouts because of the gate being closed in the second experiment. Moreover, the total amount of working hours decreased from 16 to 13. This led to the conclusion that the efficiency of the QS after changing working schedule increased.
To detect the best number of servers, the third experiment is conducted. This time the simulation model contained 2 servers. To analyse this experimental factor impact, two events flows were analysed simultaneously. Changes in dropouts flow and served customer flow after changing number of servers from 3 to 2 respectively are displayed in Figure 9 and Figure 10.

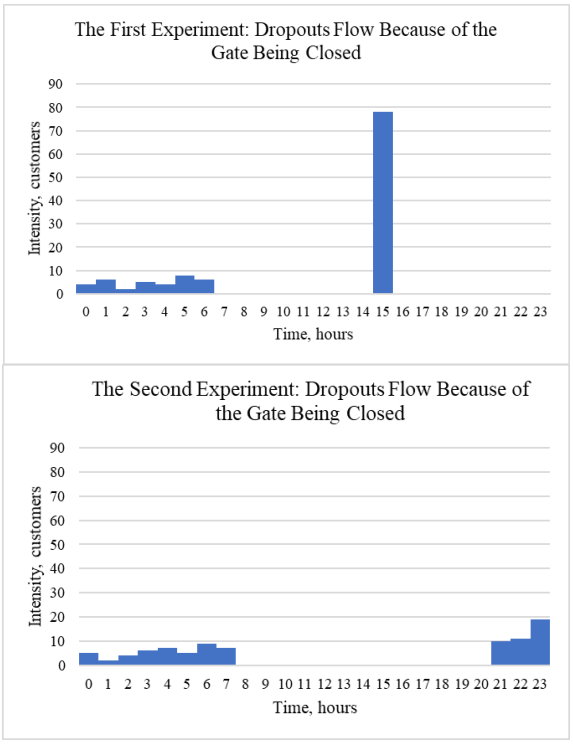


Figure 8: The First and the Second Experiments: Changes in Dropouts Flows

Figure 9 and Figure 10 show an approximately 20 percent drop in the number of served customers.
A swift decrease in the number of served customers is a critical factor for the service system.

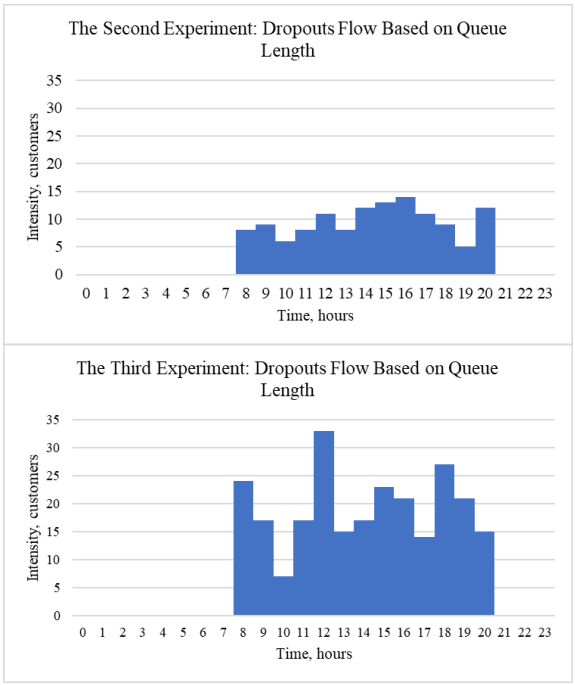


Figure 9: The Second and the Third Experiments: Changes in Dropouts Flows

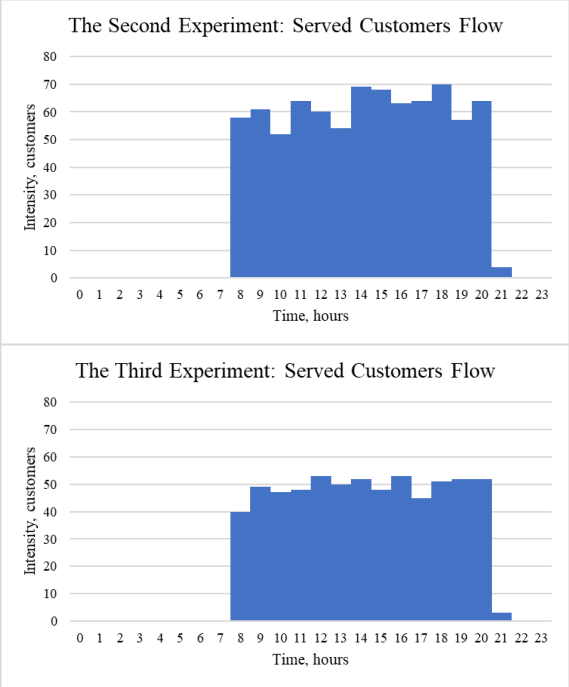


Figure 10: The Second and the Third Experiments: Changes in Served Customers Flows

Based on this, the authors decide that 3 is a required number of servers for efficient performance of this system.

5.7. Analysis of Experiments Results and Recommendations

Information summary about experimental factors in all the experiments is shown in Table 4.

Table 4. Comparison of Experimental Factors

Experi-ment Nr.	Working hours	Working schedule	Nr. of servers
1	16 hours	07:00–15:00; 16:00–24:00	3
2	13 hours	08:00–21:00	3
3	13 hours	08:00–21:00	2

According to the analysis of simulation results based on events flows, it may be concluded that the experimental factors in the second experiment provide the improved results: 25 percent decrease in the number of dropouts and 20 percent increase in the number of served customers.

On the basis of the results of the case study the following recommendations are made for efficient operation of the queueing system: number of servers is 3, there are 13 working hours and the selected working schedule includes working time from 08:00 to 21:00.

Thus the goal of the case study – to find experimental factors values leading to the best performance of a queueing system with a help of process mining techniques – has been achieved.

6. CONCLUSIONS

The experiments performed in the research confirming applicability of the proposed approach helped authors to answer the main question of the research "Is it possible to reveal insights and formulate recommendations from the simulation results by applying the process mining techniques?" positively.

Data transformation from events list of a simulation model into particular event logs and then into events flows helps to reveal the new knowledge about the performance of the system and to find the parameter values providing the advisable performance. The number of dropouts and the number of served customers are key factors in decision-making about the best performance of a queueing system.

To conclude, the developed events list-based approach is applicable to extracting process-related information and performing in-depth process analysis of queuing systems as well as other simulation-based projects.

In spite of satisfactory results of the research there is one limitation. The research did not include application of process mining algorithms. Thus the authors plan to continue research in the field of analysis of simulation results. In the future works, the authors are going to use process mining algorithms to verify their applicability in simulation results analysis.

ACKNOWLEDGMENTS

This work has been supported by Riga Technical University doctoral research grant program for academic year 2018/2019.

REFERENCES

Banks J., Carson J.S., Nelson B.L., Nicol D., 2010. Discrete-event System Simulation, 5th ed. Upper Saddle River, NJ. USA: Prentice Hall.

Brady T. F., Bowden R. A., 2001. The effectiveness of generic optimization routines in computer simulation languages. Proceedings of the 2001 Industrial Engineering Research Conference. Dallas, Texas.

Brady T. F., Yelling E., 2005. Simulation data mining: A new form of computer simulation output. Proceedings of the 2005 Winter Simulation Conference, pp. 285–289. https://doi.org/10.1109/WSC.2005.1574262.

Dunham M., 2003. Data Mining: Introductory and Advanced Topics. Pearson Education, Inc. ISBN-13: 9780130888921.

Cassandras C. G., Lafortune S., 2008. Introduction to Discrete Event Systems, Second edition., Boston, MA, USA: Springer. https://doi.org/10.1007/978-0-387-68612-7.

Feldkamp N., Bergmann S., Strassburger S., 2015. Knowledge Discovery in Manufacturing Simulations, Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. https://doi.org/10.1145/2769458.2769468.

Feldkamp N., Bergmann S., Strassburger S., Schulze T., 2016. Knowledge discovery in simulation data: A case study of a gold mining facility. Proceedings of the 2016 Winter Simulation Conference. https://doi.org/10.1109/WSC.2016.7822210.

Haav H.-M., Kalja A., 2014. Business Process Mining in Warehouses: a Case Study. Proc. of the 11th International Baltic Conference, Baltic DB&IS 204. Tallinn: Tallinn University of Technology Press, 387−394.

Han J., Kamber M., 2006. Data Mining: Concepts and Techniques, Second Edition, San Francisco, CA, USA: Elsevier Inc.

Khemiri, A., Hamri, M. A., Frydman, C., Pinaton, J. Improving business process in semiconductor manufacturing by discovering business rules, Proceedings of the 2018 Winter Simulation Conference, pp. 3441-3448. https://doi.org/10.1109/WSC.2018.8632509.

Kibira D., Hatim Q., Kumara S., et al., 2015. Integrating data analytics and simulation methods to support manufacturing decision making. Proceedings of the 2015 Winter Simulation Conference. https://doi.org/10.1109/WSC.2015.7408324.

Law A.M. and Kelton W., 2000. Simulation modeling and analysis. 3nd ed. Mc Graw Hill Higher Education.

Martin N., Depaire B., Caris A., 2018. A synthesized method for conducting a business process simulation study. Proceedings of the 2018 Winter Simulation Conference. 276-290. 10.1109/WSC.2018.8632298.

Merkurjevs J., Pečerska J., Tolujevs J., 2008. Simulation-Based Analysis of Logistic Systems, Humanities and Social Sciences. Latvia, vol. 4, no. 57, pp. 27–48.

Nakatumba J. and van der Aalst W. M. P., 2009. Analyzing Resource Behavior Using Process Mining. In Business Process Management Workshops, volume 43 of Lecture Notes in Business Information Processing, pages 69–80. Springer. ISBN 978-3-642-12185-2.

Painter M. K., Erraguntla M., Beachkofski B., et al., 2006. Using simulation, data mining, and knowledge discovery techniques for optimized aircraft engine fleet management. Proceedings of the 2006 Winter Simulation Conference. https://doi.org/10.1109/WSC.2006.323221.

Pidd M., 2004. Systems Modelling: Theory and Practice. Chichester: John Wiley & Sons Ltd.

Robinson S., 2015. A tutorial on conceptual modeling for simulation, Proceedings of the 2015 Winter Simulation Conference. https://doi.org/10.1109/WSC.2015.7408298.

Šitova I., Pečerska J., 2018, Approach to Integration of Data Mining Techniques in Simulation Results Analysis. Information Technology and Management Science. vol. 21, pp. 86–92. https://doi.org/10.7250/itms-2018-0014.

Šitova I., Pečerska J., 2017. A concept of simulation-based SC performance analysis using SCOR metrics, Information Technology and Management Science. vol. 20, pp. 85–90. https://doi.org/10.1515/itms-2017-0015.

Schriber T. J., Brunner D. T., Smith J. S. Inside discrete-event simulation software: how it works and why it matters. Proceedings of the 2014 Winter Simulation Conference. https://doi.org/10.1109/WSC.2014.7019884.

Tax N., Sidorova N., Haakma R., van der Aalst W. M. P., 2018. Mining Process Model Descriptions of Daily Life through Event Abstraction In: Bi Y., Kapoor S., Bhatia R. (eds) Intelligent Systems and Applications. IntelliSys 2016. Studies in Computational Intelligence, vol 751. Springer, Cham. https://doi.org/10.1007/978-3-319-69266-1_5.

van der Aalst W. M. P., 2011. Process Mining - Discovery, Conformance and Enhancement of Business Processes. Springer. https://doi.org/10.1007/978-3-642-19345-3.

van der Aalst W. M. P., 2012. Process Mining: Overview and Opportunities. ACM Trans. Management Inf. Syst., 3(2):7. https://doi.org/10.1145/2229156.2229157.

van der Aalst W. M. P, 2016. Process Mining: Data Science in Action. Springer. https://doi.org/10.1007/978-3-662-49851-4.

Witten I. H., Frank E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, San Francisco, CA, USA: Elsevier Inc.

AUTHORS BIOGRAPHY

Irīna Šitova is a Doctoral student at the Information Technology Institute, Riga Technical University (RTU). She received her M. sc. ing. degree in Information Technology from RTU in 2018. Currently, she is a Scientific Assistant at the Department of Modelling and Simulation of RTU. Her interests include discrete-event simulation, decision making, data mining, machine learning and process mining.

Jeļena Pečerska is a Doctor of Information Technology in the field of system analysis, modelling and development. She has been working for Riga Technical University since 1979. At the moment, she is an Associate Professor at the Department of Modelling and Simulation of Riga Technical University. Professional interests include methodology of discrete-event simulation, combined simulation, supply chain modelling, practical applications of discrete-event simulation and discrete-event simulation in education. She is a member of the Latvian Simulation Society.

SIMULATION AND ANALYSIS OF A MULTI-SCALE TUMOR MODEL USING AGENT CLUSTERED NETWORK

Ghazal Tashakor [a], Remo Suppi [a]

[a] Universitat Autònoma de Barcelona, Department Computer Architecture & Operating Systems, School of Engineering, Campus Bellaterra,
Cerdanyola del Vallès, Barcelona, 08193, Spain

[a] Ghazal.Tashakor@caos.uab.cat [a] Remo.Suppi@uab.cat

ABSTRACT

The increasing application of network models to translate and analysis of biological systems discusses the necessity of novel methodological and informatics insights for dealing with biological complexity.

Today, using tools from graph theory to simulate the dynamical system and to understand the behavior of a biological network system such as tumor growth is unavoidable.

Perhaps the most significant level of network analysis comes from mining the network measures and configuration data which allow us to explore deeper in multi-scale and multi-level biological models.

This paper presents a graph agent-based tumor model which allows us to mine network measures and visualizing the evolving behavior of tumor in molecular and cellular levels. Besides, in this paper, we have applied three applicable techniques to discover and classify subgraphs in a growing network of agents for the use of a cluster computing system.

Keywords: biological networks, biological informatics, graph, tumor, multi-scale modeling, agent-based modeling (ABM)

1. INTRODUCTION

The study of complex networks involves physics, mathematics, chemistry, biology, social sciences, and information sciences. These networks are commonly represented by directed or undirected graphs which include sets of nodes representing objects such as people or groups of people, cellular and molecular entities, computers or any other thing.

These objects joined together in pairs by edges in the concept of linking and presenting the type of relationship.

The network could be the Internet, the World Wide Web, social networks, information networks, neural networks, metabolic networks, and protein-protein interaction networks (Estrada2006).

However, the representation of complex systems as a network is not enough for the study of a certain problem, because it gives very limited information about the structure of the system in the real world. This limitation guides the study by expanding the network scale and levels to the developing platform for agent-based modeling on networks.

Agent-based modeling of complex network someway drives the emergence of the agent mining field. Agents can support and enhance the knowledge discovery process in many ways.

For instance, agents can contribute to data selection, extraction, pre-processing, and integration, and they are an excellent choice for peer-to-peer parallel, distributed, or multi-source mining. In cases where precise contact network data is unavailable, an alternative is to mine (Cao2009). In this paper, we have proposed and implemented a network simulation system using agent-based modeling and probabilistic finite state machine based on a tumor growth scenario to extract in-depth knowledge.

2. BIOLOGICAL NETWORK COMPLEXITY

Tumor evolution is a complex multi-scale process. The evolution depends on both the molecular growth factors and also on cellular growth factors effective from the cell microenvironment. The molecular growth factors include genetic mutation, gene expression, cell adaptability, robustness but, the cellular growth factors covers from the multiple metabolites and nutrient gradients. On the other hand, tumor cells can mechanically interact with other tumor cells as well as with various other stromal cells, such as fibroblasts, macrophages, and immune cells.

In the beginning, analyzing tumor evolution and metastasis were only taken into consideration at the gene or protein scale, but recently the impact of this evolution at the cellular scale and level also has been considered since the progression of most tumors depends strongly upon the interaction of the cells and the cellular architecture of the host tissue.

Evolution of cell is often modeled using a simulation of the specific cellular process such as cell growth, division, death, or movement. These simulations are determined sequentially by comparing cell status, cell age, nutrients level, the number of cell neighbors, or the

configuration of cell membrane receptors (Rejniak2011).
Another approach involves cell interactions with external factors that are taken a concentration of metabolites or ECM degradation modeled using the neural networks, signaling pathways or protein networks (Rejniak2011).

2.1. Modeling in biological network informatics

Models can be constructed based on existing knowledge of molecular interactions, from the relationship between data profiles, or based on mapping of data onto knowledge-based networks. They have been developed to assist with a variety of decisions (Railsback2019).
Knowledge-based modeling is tackling complex data for studying complex metabolism. The aim is providing support for laboratory test ordering or designing a decision support system for a scientist. The knowledge-based modeling contains the rules and associations of compiled data which mostly is mining from the rules.
Unlike knowledge-based models, non-knowledge-based modeling uses a form of artificial intelligence to allow the computer to learn from past experiences or to recognize patterns in the clinical data. Two types of non-knowledge-based systems are neural networks or genetic algorithms. There is no need for writing rules or configuring expert parametric inputs, but since the system cannot explain the reason it uses the data the way it does, most clinical does not like to use them because of no reliability. This method always relies on existing data, and therefore encounters data security and big data deficiencies will be the consequences.
It is possible to build a Knowledge-based model which will cover multiple scales from the genotype and various biochemical reactions to the details on cell morphology, and the probable behavior pattern of millions of individual cells is interacting with the other cells to form the whole tumor tissue.
Such a model may obtain structural complexity which is comparable with the biological network, but It will be less effective and computational (Railsback2019).
Therefore, it is more desirable to find a way to integrate and bridge independent sub-models rather than build a single mega-model that encompasses all the complexity of tumor development.
This integration may be in terms of separate models that consider distinct parts of the tumor evolving process or the same process but on different scales.
Extracting data from the final discrete model and doing analysis for taking the decision needs studying network informatics modules and algorithms.
Network Informatics is an interdisciplinary science based on informatics, network science, and other related scientific disciplines (Zhang2016).
Network Informatics aims to understand and investigate the structure, properties, and organization of information in the network.
The scope of network informatics covers theories, algorithms, and software of network informatics; mechanisms and rules of flow and organization of information in the network; theory and methodology of dynamics, optimization and control of information networks; network analysis of information networks; factors that affect organization and communication of information, etc. (McGillivray2018).
The integrative network-based analysis aims at identifying coordinated changes in molecular processes. In this sense, network-based analysis of high-throughput data provide the means for generating biologically meaningful hypotheses and for extracting behavior patterns from experiments to unveil the underlying regulatory mechanisms (Railsback2019).

3. AGENT-BASED COMPLEX SYSTEM BIOLOGY

Agent-based models based on complex system biology are dynamic networks of many interacting agents. Since there has not been yet established a general framework for designing, testing, and analyzing bottom-up models of cellular automata or agent-based models, recent advances in modeling have come together in a broad strategy called pattern-oriented modeling.

This strategy provides a unifying framework for discovering the organization of agent-based complex systems and may lead to merging algorithmic theories of the relation between adaptive functioning and behavior in complex systems (Grimm2005).

3.1. Pattern-Oriented Agent-Based Modeling

Pattern-oriented modeling can reduce uncertainty in model parameters in two ways. First, it helps make models structurally realistic, which usually makes them less sensitive to parameter uncertainty (Grimm2005).
Second, the realism of structure and mechanism of pattern-oriented models helps parameters interact in ways similar to interactions of real mechanisms (Grimm2005).
In this case, a technique which is known as “inverse modeling” aids to scale the parameters. The technique is finding values which reproduce multiple patterns simultaneously. In a complex biologic model such as the tumor growth analysis scenario, this inverse modeling can help the scientist to find essential values and profiles.

3.2. Graph-Structure Agent-Based Modeling

Graph algorithms have been used to characterize inter-connectivity and more detailed relationships between nodes. So this method can facilitate the modeling of biological network and its fundamental biological concepts such as cellular pathways and genes expression profiles (Aittokallio2006).
Once a biological network has been represented as a graph, the conventional graph-driven analysis work-flow involves (Aittokallio2006).

- Evaluating the specificity of the model predictions using graph evolving behavior patterns

- The shortest path length of indirectly connected nodes
- Computing the local graph properties such as the number and complexity of clustering subgraphs
- Centrality measures and statistics

Mapping agents to the nodes in the graph-structure model coordinate the assignments of values to their variables in such a way that maximizes their aggregation. Agents work as states, locations or even sometimes as controls of all the variables that map to the nodes.

3.3. Metabolic Network

A reaction in a metabolic network can be described as a weighted directed edge in a directed graph where nodes are the chemicals and edges are the reactions. There is a lack of a well-developed theory for the structural analysis of directed graphs, so two alternative representations of a metabolic reaction are usually utilized (Klamt2009).

One is a bipartite graph, and the other is a substrate graph. The bipartite graph is used when modeling relations divided into two different classes of objects such a parent and child. Substrate graph is more useful to study chemical reactions while two nodes are connected if the corresponding chemical compounds take part in the same reaction. Global characterization of the metabolic network can be carried out by obtaining the mean of the average sub-graph centrality. Small subgraphs capture specific patterns of interconnection characterizing the biological networks at the local level (Pavlopoulos2011). Analyzing local characterization of biological networks improves our understanding of evolving diseases such as metastatic cancer.

4. TUMOR AGENT-BASED MODEL

At the agent-directed modeling and simulation of tumor growth, tumor cells are affected, inflamed and turn quiescent. Based on these critical factors we have simulated the tumor growth behavior and measurements such as the tumor volume, density and also we calculate the number of dead, inflamed and tumor-derived cells. The strategy begins with the initial identification of a minor population of cells with the characteristics of "tumor-initiating cancer cells." They will be assumed inflamed or dead under the influence of angiogenic switch factors. In Figure 1 we have illustrated the different states and their variation in three different colors.

4.1. Graph Agent-Based Modeling using Python

In our first paper (Tashakor2018) we have developed a preliminary tumor agent-based model in a bipartite graph architecture considering that in life science data analysis such as tumor almost everything is about connections and dependencies. We used Mesa (Tashakor2018) framework to develop our python agent-based model. Each agent nested in a single-node and changes in three states under the influence of the neighbor nodes. The process goes on until the tumor agent's volume appears as metastasis. Two agent's connection which has shown by the graph edge changes color after the inflammation.

Following on the development of the model and for studying and analyzing the behavior of the tumor model under different conditions, we needed to explore the relevant data of the model using an extensive range of parametric executions. Our first static preliminary model was a limited scale model. To advance the initial idea, we proposed a computational workflow for simulating a multiscale tumor model. The graph-based methodology nested in agent-based modeling aids us to exploit evolving analysis more accurate and on a larger scale and by generating different patterns.

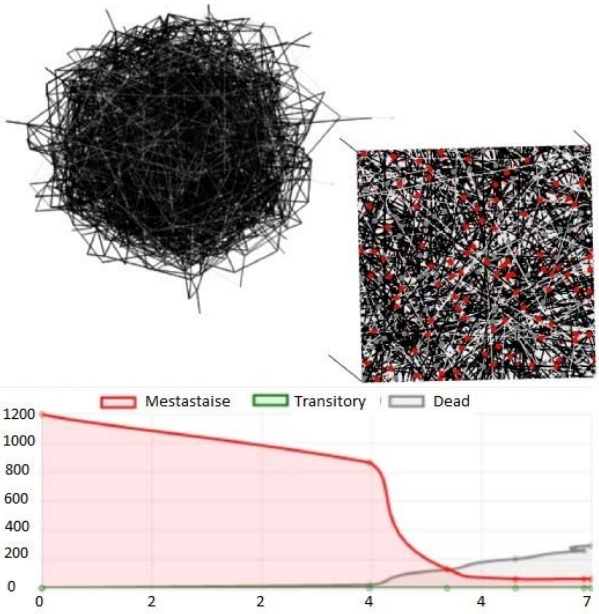


Figure 1: Graph agent-based visualization of tumor growth to 1200 cells

We have deployed two complementary graph-driven methods for analyzing and estimating probable growing network patterns. The selected methods are presented in the context of network analysis in Python using Mesa (Masad2015) and NetworkX (Aric2008) packages. The amount of data which is extracted from the methods address some problems in cell biology.

4.2. Visualization

As you can see in Figure 1, the strength of using Python with packages such as Mesa and NetworkX, first is an interactive data visualization which allows the user such as an oncologist to see the model running in the browser and second is setting up the data, parameters, figures, and plotting interactively while exploring the model data space. Python data visualization provides strong support for integration with several technologies and higher programming productivity across the development life cycle in

comparison to just using agent-based modeling software.

The plot in Figure 1 shows the change scale of inflammatory (red line) because of angioprevention interfere. The grey line in the plot shows the increasing number of dead cells. In Figure 1, besides showing the ratio of the dead cells to the inflamed cells, we have also been able to demonstrate different tumor growth behavior upon the useful laboratory condition from the angiogenic switch. Furthermore, the efficient multi-level visual exploration of multi-scale tumor model for simulating cell life cycle is one of the most influential achievements of the Python method modeling in Mesa framework using NetworkX.

4.3. A Computational Workflow for Tumor Evolving Analysis

Figure 2 shows four steps (a, b, c, and d) for the scenario of simulating tumor growth model and evolving analysis. The first step is simulating an initial tumor by setting up initial input features including normal cells and cancer cells. The second step is nesting the first graph in an agent-based model which feeds from a probabilistic finite state model of angiogenic switch that acquires the acute inflammation.

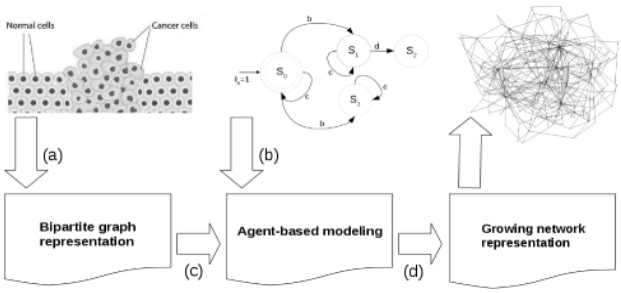


Figure 2: Graph agent-based visualization of tumor growth to 1200 cells

Forwarding the graph agent-based model to the growing module is the final step of the workflow to represent a growing network. The following is the description of each step in the content of reasons and applied techniques of the simulation workflow.

a) Bipartite graph representation: The dynamic bipartite graph is used to simulate disease spread in the behavior stated way except on a very large scale. We have simulated an initial tumor by setting up the graph in two different class of data (Normal Cells, Cancer Cells) to create a bipartite graph model as the static scale of the tumor.

b) Finite state machine: The fundamental biological aspect of the probabilistic finite state model of our tumor growth comes from the acute inflammation based upon the critical factors involved such as an angiogenic switch. Tumor angiogenesis is essential for tumor growth and maintenance. The threshold of angioprevention factor compares with the assessment values of transition probability which is selected by oncologist interactively. The result of the comparison works as a trigger to change the state of the cells from their current state to the proliferation state or the inflammation state. Afterward, under the influence of evolving angiogenic switch values, the inflammation state may turn to the progression state and metastasis happens. Since the stem cell quiescence is a way to control the inflammation in the tumor microenvironment, we have simulated probable immune state at the state model by targeting inflammation using angioprevention and stop cancer cells from moving to the proliferating state. The results in the plot show the reduction of the number of inflamed cells in comparison to the dead cells which is illustrated in Figure 1 for 1200 cells.

c) Agent-based modeling: We have created our new Tumor model in the style and structure of Mesa to facilitate our distributed execution. It is a library for Agent-Based Modeling in Python. The environment of an agent-based model in Mesa is handled by the model class to define the space where the agents evolve. The space range is from a simple square grid to a complex spatial area. The environment represents a scheduler which manages the agents at every time step. For studying the behavior of the system under different conditions, we need to collect the relevant data of the system while it is running which the Data Collector class is defined for this task. Running the model with different string point in Mesa determines in a Batch Runner class. Data collection and batch running are implemented in the appropriately-named analysis modules (Tashakor2018).
Each agent nested in a cell and changes in as many as states it needs under the influence of the neighbor nodes in our tumor growth model. The process goes on until the tumor agent's volume appears as metastasis.

d) Growing network representation: The main available network analysis algorithms are concerned with the network community identification (Hu2018). It means we need to evolve the growing cellular biological network to the molecular level and to develop the model to the other scale like zooming in to get more profound insights.
For moving from the cellular level of the tumor to molecular level, we need to transform our growing bipartite graph to a scale-free molecular format which is a better simulation for cellular interaction and molecular interconnections.

We used a growing network graph with probability (p) of tumor growth using NetworkX for adding a node one at a time with a link to the initial nodes in the premier bipartite graph.

Then by implementing a redirected modeling which is started with 200 initial random parameters and increases to 1200 sets, we reproduced multiple patterns simultaneously.

These multiple patterns keep us from the difficulty of building models that are too complex and uncertain. Also, it helps to reproduce an array of probable patterns without tuning of parameter values taken from the scientist. In Figure 6, we have presented three probable tumor behavior patterns for four redirected growing tumor model.

Practically, passing through the graph-driven analysis work-flow and changing four parameters as features subsets which were particularly uncertain, shows relatively independent effects on different outputs from the patterns. The model could be calibrated manually and also independently for generating big data-set of tumor behavior patterns.

In more deep layers of a growing network module, using network analysis algorithms which are explained in section 5 helps us to extract important microenvironment information from the model.

5. METHODS AND RESULTS OF NETWORK ANALYSIS

Use of a cluster computing system for analyzing and extracting information allows us to scale the very large complex models and also to run many replicates of the large parametric simulation.

To implement a discrete model from a dynamic coherent agent-based model and distribute the subsets among the clusters, we need to discover the complex network community structure of the agent-based model and classifying the total number of agents based on detected scales. The growing agent network model is one of the most complex models for distribution because since the agents behave in a stochastic, nonlinear manner over time, so there is a need to discover a class of the agents based on a kind of similarity such as their scale or states at their community.

In order to achieve the goals described above, practically we selected three main algorithms and techniques in network analysis for discovering, classifying and sub graphing our agent network to distribute the agents. All the techniques implemented in python using NetworkX and other packages.

5.1. Power-Law Distribution for ABM

Discovering a complex network community structure is an important challenge. Many advanced algorithms have been proposed to detect community structures in complex networks, but most have limitations. The limitations include supporting the large-scale network discovery, overlapping communities, large multiple parametric dependencies, specific structures. Therefore still cannot generate stable partitions (Liu2016).

Some methods proposed to detect the community utilizing one kind of network representation like topological measures. For example, spectral clustering (SC) for discovering the community in the graph network can effectively cluster networks, but finding the critical factor to affect the graph clustering in biological networks is very difficult by using this method because the algorithm for this task needs to overcome the problem of data representation of heterogeneous information in multi-scale biological models.

Considering the heterogeneous cell-agent population as a scale-free or very large scale network model in paper (Sugimori2015), they have proposed using Power-Law for growth of a heterogeneous population of the cells.

Power-laws are contrary to traditional Gaussian averages in that they demonstrate correlated phenomena. Examples of power laws are over eighty types of natural and social power-law phenomena in different fields. In biology, tumor growth is one of the examples (Sugimori2015).

Their idea is that a power-law encodes the high frequency of single cell identical profile. With the same hypothesis, we have cut off our graph agent-based tumor model using an algorithm for growing graphs with power-law degree distribution and approximate average clustering to achieve this identical cell profile. The algorithm joined to the growth model with an extra step that each random edge is followed by a chance of making an edge to one of its neighbors too (and thus a triangle).

5.2. Egocentric Network Analysis

Many computational methods have been developed to extract genome profiles or pathway information from biological or clinical networks for identifying subnetworks and hub nodes which have an essential role among other nodes and at the neighborhoods.

For example analysis of protein-protein interaction (PPI) networks helped to understand better the complex biology of specific disease complex system including cancer. Gene expression data have been collected from PPI networks on cancer tumors for developing algorithms. These algorithms functionally express gene profiles into some common modules which could be shared in a subset of the different type of cancers and tumors.

Genes are mapped to the corresponding proteins for representing a network graph of PPI to achieve the best-translated modules.

For modeling a multi-scale tumor network graph, we need to estimate many analysis techniques and modules which could translate the most relevant expression of PPI networks.

Egocentric network analysis techniques and modules frequently use in the social network, but there are some proposed new methods in bioinformatics such as EgoNet in (Yang2014) which present their method based on egocentric network-analysis techniques, to search and prioritize disease subnetworks and gene markers from a large-scale biological network. They have developed an algorithm to identify significant sub networks that are functionally associated with diseases, as well as predict clinical outcomes.

From a new perspective, we also used an ego-network module using NetworkX for subnetworking our graph agent-based tumor growth model.

Figure 3 shows the visual egocentric network format of our tumor growth model which is identifying seven subgraphs in the tumor model. Each subgraph has its different ego node which has its agent's role also. Each ego node works as a hub node among the other nodes in the subgraph.

Since we implemented the egocentric algorithm in our agent-based growing model, each ego node shows a different color based on its agent state. Six red agents are the metastasis, and one purple agent is inflamed but not yet become metastasis. Finding these seven hub nodes was made possible by growing the cells in the power-law distribution algorithm.

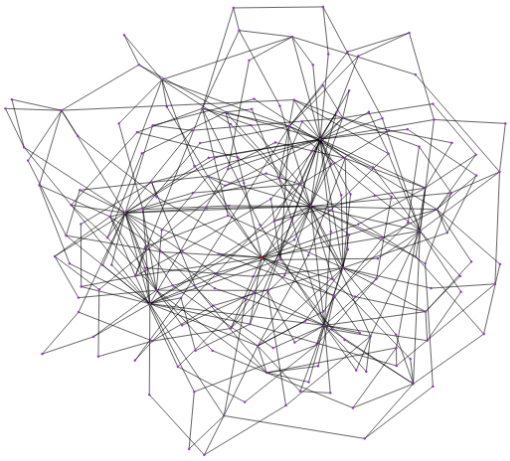


Figure 3: Egocentric visual graph of an agent-based tumor model

5.3. Discovery of Sub-Graph Centrality

One of the most critical tasks in the analysis of protein-protein interaction (PPI) is to predict a group or cluster of transiently interacting proteins that together can accomplish a biological function. These groups can be mapped to specific subgraphs in the network (Shen2012).

Characterizing nodes in a network is according to the number of closed walks starting and ending at the node. Each closed walk is associated with a connected subgraph, and the measure counts the times that a node takes part indifferently connected subgraphs of the network. The node behaves like a hub. They have called this measure the "sub-graph centrality" (SC) for nodes in a network.

Since molecular sub-typing could be done based on gene expression patterns and it helps for tumor-derived cell classification (Goodspeed2016), we used the sub-graph centrality measure as a benchmark for assessing essential genome profile of the tumor-derived cells. Each genome profile in molecular level is related to the transiently interacting proteins and identify as hub genes.

Assessing the essential genome profile of the tumor-derived cells could be an optimal partitioning function for classification of tumor cells at the large multi-scale model clustering and analysis.

Accordingly, an exclusive characteristic distribution for tumor-derived cells is illustrated by the result of the sub-graph centrality measuring in Figure 4.

These measures are labeled as genome profile of tumor- derived cells.

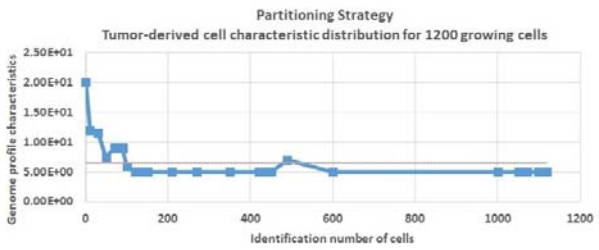


Figure 4: Tumor-derived cell characteristic distribution for 1200 growing cells

In Figure 4, from another point of view, we can see the genomic evolution of the tumor-derived cells which causes biological function such as changing the state of the neighbor cells in the subgraphs. In each run, we mine a dictionary of nodes with sub-graph centrality measures as the values. Sub-graph centrality value of a tumor-derived cell is the sum of weighted closed walks of all lengths starting and ending at the cell.

At the molecular level, tumor-derived cell acts as a hub gene and key pathways among the others so their genome profile could be the critical factor to affect the graph clustering and data classification of heterogeneous information in multi-scale models.

In this experiment, we could identify 30 tumor-derived cells among 1200 cells, but not all of them are a metastasis. With the help of Egocentric networking, we can identify how many of them are a metastasis. Also based on the obtained value of sub-graph centrality, we can estimate the scope of influence of them on the others. We have identified six metastasis cells in the egocentric graph of our model.

Finally, figure 5 shows six prepared clusters from Q1 to Q6 based on the obtained subgraphs in section 5.2. They are considered as the tumor-derived cell clusters, and six metastasis hub genes of 1200 growing cancer cells are considered as hub nodes which create the clusters.

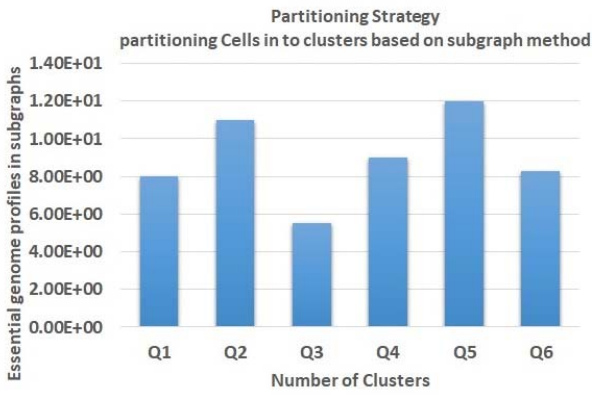


Figure 5: Partitioning Cells into clusters based on sub-graph centrality

Each cluster includes all the neighbors of a determined distance less than the intended radius of the hub node and the connected neighbors who are in the same class of the genomic profile collect in the same cluster.

Tumor cells in cluster number five (Q5) have the most effective behavior and essential genome profile which probably can accomplish an aggressive function among the others or become the genetic reason of the metastasis.

Figure 6 is illustrated the three redirected behavior pattern representations of tumor-derived cells in four different tumor model during parametric executions.

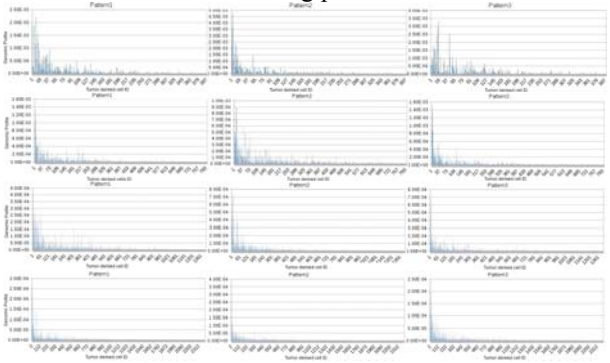


Figure 6: Three probable behavior patterns for four redirected growing tumor model based on tumor-derived cell distribution

This figure shows that each tumor could grow into different pattern based on the probability of tumor-derived cells distribution. Each pattern shows the analyzing of the tumor-derived cells variation by identifying genomic profiles of those cells.

6. CONCLUSION

In this paper, we developed a multiscale graph agent-based model. The model uses for extracting and analyzing cell information from the growing network of a tumor.

The dynamic behavior of tumor-growth seems interesting to oncologists and scientist since they can study the probable predictive power of pathways in the cellular network of the tumor.

Also, it shows that creating a discrete model from a growing multi-scale agent-directed simulation that each agent is in the transition to several possible states and is in communicate with own neighborhood is an important challenge in dynamic modeling and simulation.

This paper presents a graph workflow simulation system for modeling, growing behavior, and sub-graphing analysis which practically allows the scientist to distribute their large scale parametric models distribute on clusters.

ACKNOWLEDGMENTS

This publication is supported under contract TIN2017-84875-P, funded by the Agencia Estatal de Investigación (AEI), Spain and the Fondo Europeo de Desarrollo Regional (FEDER) UE and partially funded by a research collaboration agreement with the Fundacion Escuelas Universitarias Gimbernat (EUG).

REFERENCES

Estrada E, Rodríguez-Velázquez JA. Subgraph centrality and clustering in complex hyper-networks. Physica A: Statistical Mechanics and its Applications. 2006 May 15;364:581-94.

Rejniak KA, Anderson AR. Hybrid models of tumor growth. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2011 Jan;3(1):115-25.

Railsback SF, Grimm V. Agent-based and individual-based modeling: a practical introduction. Princeton university press; 2019 Mar 26.

McGillivray P, Clarke D, Meyerson W, Zhang J, Lee D, Gu M, Kumar S, Zhou H, Gerstein M. Network analysis as a grand unifier in biomedical data science. Annual Review of Biomedical Data Science. 2018 Jul 20;1:153-80.

Aittokallio T, Schwikowski B. Graph-based methods for analysing networks in cell biology. Briefings in bioinformatics. 2006 Sep 1;7(3):243-55. [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. science. 2005 Nov 11;310(5750):987-91.

Tashakor G, Luque E, Suppi R. High Performance Computing for tumor Propagation Agent-based Model. In: XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017)

Tashakor G, SuppiR. : Agent-based model for tumor-analysis using Python+Mesa. In: European Modeling and Simulation Symposium, pp. 248-253, EMSS (2018)

Masad D, Kazil J. MESA : an agent-based modeling framework. In: Proceedings of the 14th Python in Science Conference, pp. 53-60, SCIPY(2015)

Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart.: Exploring network structure, dynamics, and function using NetworkX, In: Proceedings of the 7th Python in Science Conference, pp. 11–15. SciPy (2008)

Yang R, Bai Y, Qin Z, Yu T. EgoNet: identification of human disease ego-network modules. BMC genomics. 2014 Dec;15(1):314.
Hu P, Niu Z, He T, Chan KC. Learning Deep Representations in Large Integrated Network for Graph Clustering. In2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) 2018 Sep 26 (pp. 101-105). IEEE.
Shen R, Goonesekere NC, Guda C. Mining functional subgraphs from cancer protein-protein interaction networks. BMC systems biology. 2012 Dec;6(3):S2.
Goodspeed A, Heiser LM, Gray JW, Costello JC. Tumor-derived cell lines as molecular models of cancer pharmacogenomics. Molecular Cancer Research. 2016 Jan 1;14(1):3-13.
Cao L, Gorodetsky V, Mitkas PA. Agent mining: The synergy of agents and data mining. IEEE Intelligent Systems. 2009 May;24(3):64-72.
Sugimori M, Hayakawa Y, Boman BM, Fields JZ, Awaji M, Kozano H, Tamura R, Yamamoto S, Ogata T, Yamada M, Endo S. Discovery of power-law growth in the self-renewal of heterogeneous glioma stem cell populations. PloS one. 2015 Aug 18;10(8):e0135760.
Liu W, Pellegrini M, Wang X. Detecting communities based on network topology. Scientific reports. 2014 Jul 18;4:5739. Zhang W. Network informatics: A new science. Selforganizology. 2016;3(2):43-50.
Klamt S, Haus UU, Theis F. Hypergraphs and cellular networks. PLoS computational biology. 2009 May 29;5(5):e1000385.
Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, Schneider R, Bagos PG. Using graph theory to analyze biological networks. BioData mining. 2011 Dec;4(1):10.

AUTHORS BIOGRAPHY

GHAZAL TASHAKOR is Research Fellow and Ph.D. Candidate for trainee research staff position (PIF) at the Computer Architecture & Operating Systems Department (CAOS) of the University Autònoma de Barcelona. Her research interests include high performance simulation, agent-based modeling, and their applications, service systems, integration, resource consumption, and execution time. Her e-mail address is ghazal.tashakor@caos.uab.cat

REMO SUPPI received his diploma in Electronic Engineering from the Universidad Nacional de La Plata (Argentina), and the Ph.D. degree in Computer Science from the Universitat Autonoma de Barcelona (UAB) in 1996. At UAB, he spent more than 20 years researching on topics including computer simulation, distributed systems, high performance and distributed simulation applied to ABM or individual-oriented models. He has published several scientific papers on the topics above and has been the associate professor at UAB since 1997, also a member of the High- Performance Computing for Efficient Applications and Simulation Research Group (HPC4EAS) at UAB. His email address is remo.suppi@uab.cat.

SIMULATION-BASED ANALYSIS OF INVENTORY LEVELS FOR LOW DEMAND SPARE PARTS IN A COOPERATIVE INVENTORY POOLING-SYSTEM

Yannic Hafner[(a)], Christian Looschen[(b)], Johannes Fottner[(c)]

[(a),(b),(c)]Technical University of Munich, Chair of Materials Handling, Material Flow, Logistics, Boltzmannstr. 15, 85748 Garching, Germany

[(a)]yannic.hafner@tum.de, [(b)]christian.looschen@tum.de, [(c)]j.fottner@tum.de

ABSTRACT

Due to their systemic relevance, a high availability of intralogistics systems is crucial. This requires the stockpiling of capital-intensive spare parts. Strategies are therefore needed to reduce the severe capital commitment for spare parts management. No practical recommendations and solutions for this problem are available however. This contribution describes strategies and procedures for a cooperative inventory pooling-system and our model of a simulation-based analysis of inventory levels for low demand spare parts in these systems. We determine that companies can gain benefits and greatly lower their inventory level and associated costs when cooperating in spare parts management. Furthermore, we investigated effects of changing network compositions while cooperating over time.

Keywords: spare parts management, cooperative inventory-pooling, discrete event simulation, inventory optimization

1. INTRODUCTION

Intralogistics systems are the basis for internal material flows and value creation of producing or distributing companies. Therefore, a high availability of these systems is crucial. This requires the stockpiling of capital-intensive spare parts. The majority of the necessary spare parts are expensive and characterized by sporadic demand, which leads to an increased and often obsolete stock of spare parts (Huiskonen 2001). The annual material costs for the spare parts are up to approx. 2.5 % of the purchase price of an installation (Gallagher et al. 2005). Hence strategies are needed to reduce the severe capital commitment for warehousing of these necessary spare parts.

A cooperative inventory pooling-system is one promising solution to increase the efficiency of spare parts management. This contribution deals with a simulation-based analysis of inventory levels for low demand spare parts in a cooperative inventory pooling-system.

2. STATE OF SCIENCE AND TECHNOLOGY

An optimal inventory level can be calculated using analytical and simulation-based models and is the subject of numerous scientific studies. In this contribution, we focus on literature that deals with the inventory management for cooperating inventory pooling-systems. Paterson et al. (2009) summarize and review several publications on inventory models with lateral transshipments, a synonym for inventory pooling. These use different key characteristics to differentiate between the models introduced in the different papers and articles. These are, among others, the characteristics of the depicted system (number of echelons, number of warehouses, number of items), the ordering timing and policy, the transshipments and pooling types. Considering only models in which characteristics are most similar to ours, we can limit the relevant publication reviewed by Paterson et al. to models with reactive lateral transshipments in a single echelon logistics system with complete pooling. The review of the inventory in these systems is continuous and a (S-1, S) reordering policy is used. Full pooling of the models inventory is assumed.

Kukreja et al. (2001) define an analytical model for the required network stocking level that minimizes inventory costs while respecting the desired Service Level of the cooperation. A Poisson distribution is used to derive a lower bound for the stocking level. All companies in the network have identical spare part demands.

Wong et al. (2006) also use an analytical model to measure the performance of their single-echelon pooling system with *n* different companies sharing repairable items. The performance is measured by the system's Service Level, the expected total downtime for a company and the expected number of lateral transshipments.

While many other authors consider a zero lateral transshipment time, Wong et al. consider a non-zero lateral transshipment time that is distributed exponentially. The repair time of items, which includes the shipment time to the repair facility, is also modelled as exponentially distributed. The authors show, that delayed lateral transshipment have a big impact on the expected number of backorders and that this has to be considered when deciding how many and where items are to be stocked.

Karsten and Basten (2014) investigate a joint inventory pooling and the influence on the cost for spare parts management. Multiple companies, each having a Poisson demand process for expensive, low demand items, are storing every spare part at a single mutual location. The authors calculate the necessary on-hand stock, if n companies are pooling their spare parts. Using cooperative game theory, they derive a function for the expected storage costs and a function for allocating the costs throughout the cooperation of companies so that each company benefits economically.

A stochastic cost model for joint spare part inventories is described by Wang (2012). The model optimizes the inventory control of the required spare parts and the machine inspection interval. Wang uses an algorithm with stochastic dynamic programming to find optimal joint solutions.

Fritzsche (2012) considers further cost estimations in a cooperative inventory pooling-system. Assuming a dynamic failure rate and using the MFOP (maintenance free operating period) process, the author shows that using dynamic failure rates instead of constant ones to describe the failure of spare parts results in the reduction of costs and improvement of operational stock planning. This is achieved by MFOP, which guarantees multiple maintenance free periods and the use of memory learning systems to optimize the calculation of the failure.

State of science and technology show a wide range of analytical and simulation-based models to find an optimal joint inventory level. Karsten and Basten (2014) extend their model using cooperative game theory, in consideration of the behavior of the different members. All models have in common that the inventory pooling describes a situation in which the companies are cooperating before buying necessary spare parts and without already acquired spare part inventories. The situation of companies starting a new cooperation with existing inventories is not considered. An optimal solution will be reached in the future, but interested companies have no way to investigate their benefits and evaluate the effect of the cooperation beforehand. New companies might want to join the cooperation, whereas other companies leave. New spare parts can be pooled and necessary costs reduced or other spare parts can't be stockpiled economically in the network anymore. This research gap is addressed in our contribution.

3. CONSIDERED SYSTEM AND STRATEGIES

The simulation depicts a network with n different companies, one central internal supplier, one external supplier and one central network platform as a control unit.

We consider two different representative spare parts (*SP1*, *SP2*) with varying demands, with not every company having a need for each spare part. The demand of a company is characterized by the failure of the parts inside the companies' machines. The demands of the companies are independent from each other.

Every spare part is stored centrally at one location, since delivery times between companies are negligible in our network. It is assumed that the time between ordering a spare part and receiving it from a cooperation partner is equal to the time for disassembling the machine to access the faulty part. To reduce necessary deliveries, the spare parts are stored at the company with the highest demand, which acts as a central internal supplier. A one-for-one replenishment policy (S-1, S) is implemented. Failed parts are removed from the system and not repaired.

The procedure for ordering a spare part is shown in Figure 1. If a spare part demand arises, the company that needs the spare part places the order in the central network platform. The platform then processes the order and checks which spare part is needed and whether the order can be fulfilled from the cooperative inventory. Orders are hereby processed on a first-come, first-served basis. If a spare part is available, the storage location will be checked. Spare parts stored at the ordering company can be taken directly from the stock. A new part will immediately be reordered by the platform from the supplier to ensure a desired Service Level. This part is sent to the company that acts as the central internal supplier. The delivery time for this reorder is t_s. If the spare part is available at a network partner, it will be handled as an internal order and delivered from this location before a new spare part is reordered.

If an order cannot be filled from the system inventory, the platform inquires if an already ordered spare part of the same kind is close to delivery to the central internal supplier. If the remaining delivery time of the needed spare part is smaller than an accepted downtime t_{dt}, the part is rerouted to the company which placed the order and is finally delivered to this company in need, before a new spare part will be reordered. The accepted downtime is hereby shorter than the total delivery time for reordering a spare part from the central supplier ($t_{dt} << t_s$). If no part is available in the network inventory and there is no sufficient open order, an emergency order with penalty charges is placed with an external supplier that can supply the spare part within the accepted downtime t_{dt} of the manufacturing process. Ordering costs are accounted once the spare part arrives at the company.

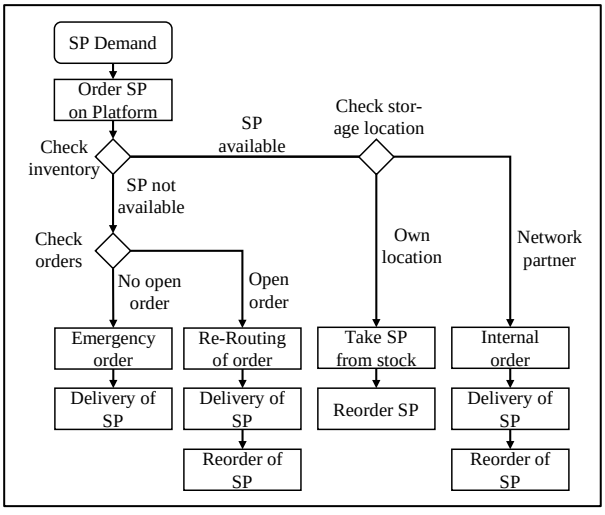


Figure 1: Procedure when ordering a spare part

4. SIMULATION MODEL

We followed the typical proceeding of building and evaluating a simulation, generalized by the VDI-3633 (VDI 2014) guideline of the Association of German Engineers, to transfer the considered system into a simulation model in the simulation environment *Tecnomatix Plant Simulation*.

The simulation model consists of different modules and elements, the main ones being the *company*, the *central network platform*, the *external supplier*, the *central internal supplier* and the considered *spare parts*.

A company module consists of a component representing the inventory, the machines used by the company, a receiving area and a distribution area. Methods control the flow of goods in and out of the inventory, the distribution of incoming goods inside the company and the distribution of goods out of the company. Another method in the module manages the company's order procedure, while it is presumed that the company has a limitless storage capacity. The machines of a company are depicted as a working station, which can hold multiple (spare) part elements. A method inside the machine module calculates the time until the failure of the part. A formula for the Weibull distribution with the form parameter β and the scale parameter η is used to calculate this time.

A method controlling the machine output destroys the spare part element after its time in the machine has run out and then initiates a spare part order.

Imbedded inside the module for the network platform are the methods controlling the spare parts order process. These can be classified into methods controlling the replacement of the stocked part, the network internal delivery, the order process from the central internal supplier, the order process from the external supplier and the rerouting of a nearly completed order from one company to another. Included in the network platform is also a method calculating and updating the Service Level of the cooperative inventory pooling-system every twelve hours. For evaluation, we take track of the current demand of the companies and record the inventory level in the pooling-system. The Service Level is calculated by dividing the number of total deliveries minus the number of emergency deliveries from the external supplier by the total number of deliveries.

The modules for the central supplier and for the external supplier are similarly built. Both have methods for creating a spare part element when ordered and for distributing these spare parts to the ordering companies. The production or acquisition of a spare part is represented by the creation of a spare part element inside a buffer module. The element stays in the module for the whole delivery time. As the delivery time for reordered spare parts to the central internal supplier (t_s) is longer than the ones coming from the external supplier (t_{ex}), the central internal supplier module documents the information of current reorder procedures at the supplier. In terms of real and simulation, the time for delivering the spare parts is of no consequence.

The input variables are:

- the experiment number e_n
- the spare part type *SPn*
- the number of companies in the cooperation nr_c
- the number of warehouses nr_w
- the desired Service Level SL_d
- the inventory level of the whole cooperation to be investigated I
- the Weibull form-parameter β
- the Weibull scale-parameter η in days
- the delivery time of an internal delivery from the central internal supplier to another company t_{id} in days
- the delivery time of a reordered spare part from the supplier t_s in days
- the delivery time for an emergency order from the external supplier t_{ex}
- the tolerated machine downtime t_{dt}
- the number of parts needed by a company for their machines nr_p equivalent to the average demand per characteristic lifetime d_{cl}

Additional variables are the internal delivery costs for orders from the central internal supplier C_{id}, the spare part costs for a spare part ordered from the central internal supplier C_s, the spare part costs for ordering the spare part from the external supplier for emergency orders C_e and the inventory holding costs C_h. In addition, the input control allows the user to choose if the number of companies changes over time. This means that an additional company module is included into the simulation model, having the attributes set in the specific tables inside the simulation control, or a company is removed from the model.

The output of the simulation is as follows:

- the Service Level of the cooperative inventory pooling-system over time
- the cumulated total costs for each company in the cooperation
- the costs per company for each year
- the number of total deliveries divided into the separate delivery types for the whole network and for each company
- the time every part is installed in the machine module (depicting the time between demands)

At the beginning of the simulation, after the matching of the input variables, the number of companies nr_c specified in the input are created and the spare part elements are distributed over the number of warehouses nr_w with the companies with the highest demand assigned the central internal supplier. Subsequently, elements are created according to the number of parts nr_p in the machines of the companies. Failure times are calculated using the Weibull distribution function and assigned to the parts.

Once the set-up is complete, the simulation follows the previously described procedures (Section 3).
The simulation model needs to be validated and verified before experiments can be conducted. In a first step we used a structured walkthrough to verify that the implementation does not have any mistakes and that the model and the running processes matched the considered system description (Section 3 and Section 4). A second step included the carrying out of experiments to verify, that the distribution of the times between demands in the simulation model concur with the Weibull distribution chosen for the depiction of demand in the network. The results of the experiments showed a consistency of the analytical and simulated distributions. The simulation model is therefore regarded to be valid.

5. EXPERIMENTS

The conducted experiments aimed to answer the following questions

1. What is the needed inventory level to reach a desired Service Level of 95 % in different representative cooperative inventory pooling-systems warehousing low demand spare parts?
2. What is the effect during an increase of spare parts demand d_{cl} for certain inventory levels?
3. What are the costs and the amount of necessary spare parts per company for companies that are part of the cooperative inventory pooling-system compared to a stand-alone solution?
4. What is the effect over time when a new company with existing inventories joins or leaves the cooperative inventory pooling-system?

Due to the randomly distributed spare part demand of each company in our cooperative inventory pooling-system we need to determine the number of simulation runs for each experiment. We aim to receive a confidence interval for the average Service Level of 1 % with a confidence level of 95 %. By testing a number of simulation runs we achieved this confidence interval with 500 runs for every experiment.
The first experiments were used to analyze the needed inventory levels for two different spare part types (*SP1*, *SP2*) to reach a desired Service Level of at least 95 % with different network sizes. *SP1* represents a typical gear motor and *SP2* a conveyor belt that is subject to high wear and tear. Effects and costs are analyzed during an increase of the spare part demand d_{cl}. The parameters for the simulation are given in Table 1. The Design of Experiments was used to create an experimental design with an experiment number of 270 for each spare part type.

Table 1: Parameters for experiments analyzing the needed inventory level for a desired Service Level of 95 %

Parameter	*SP1*	*SP2*
nr_c	2-10	2-10
nr_w	1	1
SL_d	95 %	95 %
I	1-10	1-10
β	1.2	1.2
η	3 years, 5 years, 10 years	0.5 years, 1 year, 1.5 years
t_{id}	0	0
t_s	35	28
t_{ex}	1	1
t_{dt}	1	1
d_{cl}	1	1

The cost function takes inventory holding costs C_h in percent per day, costs for an internal delivery C_{id}, costs for the spare part served from the cooperation C_s and costs for the spare part from an external supplier C_e, which is three times C_s, into account. The values for these costs are based on analyzed industry data sets and are shown in Table 2.
Inventory costs are distributed among the companies of the network; all other costs are accounted when a spare part is ordered.

Table 2: Parameters for the spare parts related costs for experiments analyzing the occurred costs for companies that are part of the cooperative inventory pooling-system

C_{id}	C_s	C_e	C_h
80 €	2105 €	6315 €	15 %

The second excerpt of experiments was used to analyze the reaction of the networks' Service Levels over time if a company joins or leaves the cooperative inventory pooling-system. Therefore, a network of four companies with one central internal supplier was set up. After roughly three years, a company joins or leaves the network with its own inventory and a machine part already in use. In this scenario, all companies have the same demand for each spare part of the type *SP1* with a characteristic lifetime of three years.
The simulation time for all conducted experiments was 15 years.

6. RESULTS

In this section, we present and discuss the results obtained by the experiments described in section 5.

6.1 Inventory levels for a Service Level of 95 % in different representative cooperation networks

A cooperative inventory pooling-system is highly interesting for companies that aim to reduce their required inventory levels and therefore, the necessary capital commitment costs to serve a specific Service

Level. As a basis for comparison, Table 3 shows the required inventory levels of a single company not cooperating at different demands d_{cl} to reach a 95 % Service Level for spare part *SP1* and *SP2*.

Table 3: Results for analyzing the required inventory levels for achieving a desired Service Level of 95 % at different demands d_{cl} of a single company for spare part *SP1* (η = 3 years) and *SP2* (η = 1 year)

Demand d_{cl}	*SP1*		*SP2*	
	Inventory Level	Service Level in %	Inventory Level	Service Level in %
1	1	98.77	1	95.6
2	1	95.77	2	99.58
3	2	99.76	2	98.46
4	2	99.55	2	97.43
5	2	99.13	2	95.57
6	2	98.81	3	99.21
7	2	98.32	3	98.81
8	2	97.91	3	98.35
9	2	97.28	3	97.77
10	2	96.65	3	97.03

For *SP1*, an inventory level of one is sufficient to satisfy a demand d_{cl} of up to two parts. A higher demand of ten parts requires the stockpiling of one more part. *SP2* has a significant shorter characteristic lifetime which leads to an inventory level of up to three parts.
The results presented in Table 4 and Table 5 show the maximum number of companies with a demand d_{cl} of one each that can cooperate to guarantee a 95 % Service Level with an identified inventory level. As expected, the number of companies able to be supplied, and the achieved Service Level rise with the number of spare parts on stock.

Table 4: Results for analyzing the required inventory levels for achieving a desired Service Level of 95 % in different sized cooperation networks for spare part *SP1*

Characteristic Lifetime in years	Inventory Level	Max. number of companies	Service Level in %
3	1	2	95.59
3	2	10	96.72
3	3	10	99.69
5	1	3	95.43
5	2	10	98.61
5	3	10	99.91
10	1	5	95.64
10	2	10	99.79
10	3	10	99.96

The jump between the maximum number of companies for an inventory level of one and two for *SP1* with a characteristic lifetime of three years is prominent in Table 4. A maximum of two companies can be supplied with an inventory level of one spare part and a Service Level of 95.59 % will be reached. Warehousing just one more spare part allows the cooperative inventory pooling-system to supply up to ten companies, which results in an even higher Service Level of 96.72 %.
Table 5 shows comparable results: An inventory level of a maximum of four spare parts is sufficient to supply a cooperative inventory pooling-system of ten companies.

Table 5: Results for analyzing the required inventory levels for achieving a desired Service Level of 95 % in different sized cooperation networks for spare part *SP2*

Characteristic Lifetime in years	Inventory Level	Max. number of companies	Service Level in %
0.5	2	2	97.78
0.5	3	5	97.31
0.5	4	10	97.09
1	2	5	95.77
1	3	10	97.2
1	4	10	99.45
1.5	2	5	98.1
1.5	3	10	98.91
1.5	4	10	99.89

Comparing the results of Table 4 and Table 5 show that an increasing inventory level is necessary for spare part *SP2* to satisfy the demand of the same number of companies. Seeing that the form-parameter β is the same for *SP1* and *SP2* the increase can be explained with a reduction of the average characteristic lifetime of *SP2*, as is to be expected.
With an increasing inventory level, the average amount of spare parts per company needed to guarantee a desired 95 % Service Level is reduced (Figure 2). Supplying ten companies with two spare parts *SP1* results in 0.2 spare parts per company that have to be financed and stored at a time compared to an inventory level of one for a single company. Similar results can be determined for spare part *SP2*, where 0.3 spare parts per company are required.

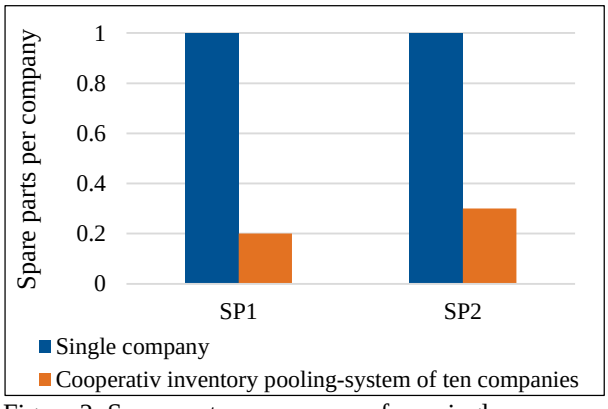


Figure 2: Spare parts per company for a single company and a cooperative inventory pooling-system of ten companies

6.2 Effects when increasing the spare part demand for certain inventory levels

Figure 3 shows the effect when increasing the spare part demand for *SP1* (η = 3 years) and *SP2* (η = 1 year) for certain inventory levels and completes the previous investigations. Each company faces the same demand of one spare part.

In the following, we note that the Service Level for *SP2* is more dependent on the inventory level as the characteristic life time is much shorter.

Furthermore, a demand for *SP1* and an inventory level of one lead to a Service Level of 98.5 %. A comparable Service Level can be achieved having two spare parts on stock for six companies. An inventory level of three leads to a Service Level of nearly 100 % in a cooperation of ten companies.

For *SP2*, we achieve a comparable Service Level of 95 % for one company warehousing one spare part or ten companies warehousing three spare parts.

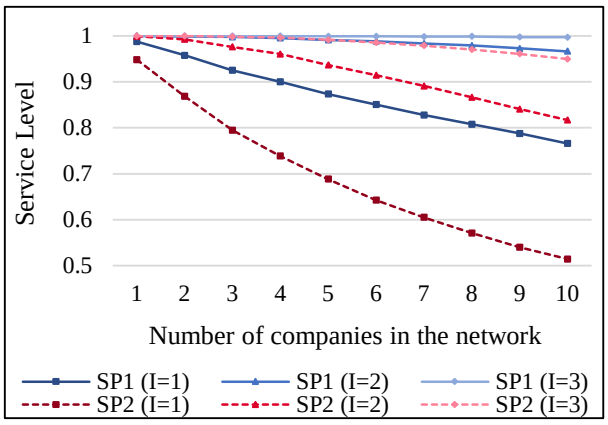


Figure 3: Service Level over networks' spare part demand for *SP1* and *SP2* for an inventory level of one to three

6.3 Costs evaluation of the cooperative inventory pooling-system

Joining a cooperative inventory pooling-system leads to a reduced amount of spare parts per company. However, the decision of joining such a system is cost-dependent. Necessary costs need to be examined. An initial cost function was therefore derived.

Using this simple cost function provides a first overview of the incurred cumulated costs for each company over a time of fifteen years (Figure 4). The referred representative network consists of five companies (C1 to C5) warehousing two spare parts *SP1* that result in a Service Level of 99.13 %. All companies face a spare part demand of one *SP1* with a characteristic lifetime of three years. The spare parts are stored centrally at Company 1 to fulfill occurring demands.

For the described scenario, savings of about 30 % to 40 % compared to a single company (SC) without a cooperative inventory pooling-system can be realized while the Service Level stays at this high level. The cost-saving effect increases with the number of cooperating members. However, a positive effect can already be seen when two companies are cooperating.

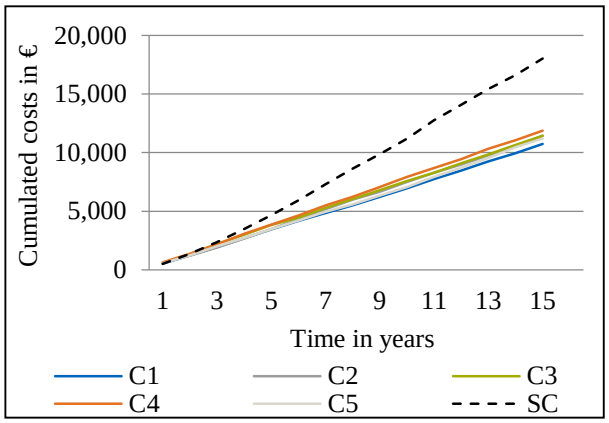


Figure 4: Cumulated costs for a cooperative inventory pooling-system of five companies warehousing *SP2*

6.4. Changing the network composition while cooperating

In another excerpt of experiments, we analyzed the effects of changing the network composition while cooperating. We analyzed the resulting Service Level over time. The results showing the Service Level for a System with

- no changes (NC),
- a joining company without its own inventory (J-0SP),
- a joining company with an own stock level of one spare part (J-1SP),
- a joining company with an own stock level of two spare parts (J-2SP) and
- a leaving company (LC)

are shown in Figure 5.

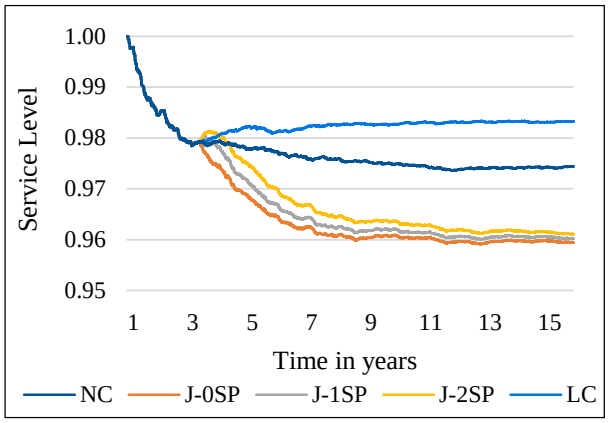


Figure 5: Service Level over time when changing the network composition while cooperating

A system with no changes as reference shows a Service Level of about 97.4 %. Companies joining the cooperation lead to an increased spare parts demand; in our experiments; this results in a growth of 25 %. The Service Level over time therefore drops to 96 %. The difference between companies joining with or without an own stock of spare parts is a delayed drop of the Service

Level. For a short period of time, the actual Service Level increases when joining due to the own stock. This effect is enhanced when the company joins with two spare parts.
When a company leaves the cooperation, the demand is reduced. This leads to an increase of the Service Level over time to about 98.4 %.

CONCLUSION AND OUTLOOK

In this contribution, we considered a cooperative inventory pooling-system for low demand spare parts. We conducted a series of simulation experiments to compare cooperation to the stand-alone solution based on required inventory levels and costs to achieve a desired Service Level.
We were able to show that companies can lower the amount of necessary spare parts per company greatly when cooperating. A cooperation size of ten companies leads to a reduction of required spare parts per company of 70 % to 80 %. A first cost function was derived and showed, that economic benefits of 30 % to 40 % can be gained in the long-term.
Furthermore, the effects of changing the network composition while cooperating were examined. It can be noted that a joining company has a low negative long-term effect on the cooperation. The effect is delayed when bringing an own inventory. Companies leaving the cooperation slightly raise the Service Level over time.
By way of next research, we suggest implementing a detailed cost function that considers date of payment and the resulting free cash flow individually for each company and for the cooperation.
In addition, recommendations for companies that are interested in a cooperative inventory pooling-system that faces different demands for specific spare parts need to be developed and derived.

REFERENCES

Fritzsche R., 2012. Cost adjustment for single item pooling models using a dynamic failure rate: A calculation for the aircraft industry. Transportation Research Part E 48, pp. 1065-1079.

Gallagher T., Mitchke M. D., Rogers M C, 2005. Profiting from spare parts. The McKinsey Quaterly.

Huiskonen J., 2001. Maintenance spare parts logistics. Special characteristics and strategic choices. International Journal of Production Economics 71 (1-3), S. 125–133.

Karsten F., Basten R., 2014. Pooling of spare parts between multiple users. How to share the benefits? European Journal of Operational Research (233-1), S. 94–104.

Kukreja A., Schmidt C. P., Miller D. M., 2001. Stocking Decisions for Low-Usage Items in a Multilocation Inventory System. Management Science 47 (10), S. 1371–1383.

Paterson C., Kiesmuller G. P., Teunter R. H., Glazebrook K., 2009. Inventory models with lateral transshipments: a review. Eindhoven University of Technology.

Schillinger R., Wortmann B., Buß D., 2015. Ressourcenschonende Chemieparklogistik. In: Voß P., 2015, eds. Logistik – eine Industrie, die (sich) bewegt, Wießbaden, pp. 173 -190.

Shen Z. M., Coullard C., Daskin M. S., 2003. A joint Location-Invenotry Model. Transportation Science 37(1), pp. 40-55.

Verein Deutscher Ingenieure (VDI), 2014. VDI-3633 Part 1: Simulation of systems in materials handling, logistics and production – Fundamentals. Beuth Verlag, Berlin.

Wang W., 2012. A stochastic model for joint spare parts inventory and planned maintenance optimisation. EURO Excellence in Practice Award 2001 216 (1), S. 127–139.

Wong H., Cattrysse D., Van Oudheusden D, 2004. Inventory pooling of repairable spare parts with non-zero lateral transshipment time and delayed lateral transshipment, European Journal of Operation Research (165), pp- 207-218.

YANNIC HAFNER has been working as a research associate at the Chair of Materials Handling, Material Flow, Logistics at the Technical University of Munich, since 2017. His research deals with digital platforms for cooperative inventory pooling-systems.

CHRISTIAN LOOSCHEN graduated from the Technical University of Munich. His master's thesis deals with a simulation-based analysis of inventory levels for low demand spare parts in cooperative inventory pooling-systems.

JOHANNES FOTTNER is professor and head of the Chair of Materials Handling, Material Flow, Logistics at the Technical University of Munich. His research interests include investigating innovative technical solutions and system approaches for optimizing logistical processes.

STUDY ON FUEL ECONOMY AND GENERATING ENERGY OF 48V MHEV WITH TWO MOTORS

Seongmin Ha [(a)], Hyeongcheol Lee[*(b)]

[(a)]Department of Electric Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea
[(b)] Division of Electrical and Biomedical Engineering, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea

[(a)] haha4100@hanyang.ac.kr, [(b)] hclee@hanyang.ac.kr

ABSTRACT
In this paper, 48V MHEV vehicle with P0 + P4 structure is studied. In the previous paper, the best fuel economy was obtained when the energy obtained by regenerative braking was used only for EV driving. In an actual driving environment, a power generating function by an engine is essential. Therefore, not only regenerative braking energy but also additional energy from power generation is used for EV travel. Finally, the fuel economy is evaluated according to the amount of generating energy and control tendency of generating.

Keywords: Mild hybrid electric vehicle(MHEV), Fuel economy, Generating energy

1. INTRODUCTION

In the previous paper, various fuel economy comparisons of 48V MHEV of P0+P4 structure were studied. The best fuel economy was achieved when all the electrical energy from regenerative braking was used to EV driving from the same motor specifications. (S, HA. 2017). Also, when the ratio of two motor power is different, the BSG only has power to start the engine and the higher the power of the Rear-Axle motor, the better fuel economy. Because the EV operating area is wider and regenerative braking energy is high. (S, HA. 2018).

In the actual driving environment, control is not easy to use regenerative braking energy only for EV driving. In fact, the vehicle is not driven in a defined scenario (ex: FTP 75 cycle). And there is additional electricity consumption, such as the use of air conditioners. Therefore, power generating functions using engines and BSG are essential. In this paper, several control methods of generating are divided to define simulation cases. The effect of generating on fuel economy is analyzed by comparing generating energy quantity and efficiency for each simulation case.

2. 48V MHEV SYSTEM MODELING

In this paper, 48V MHEV with P0 + P4 structure is modeled as shown in Figure 1. The models of the engine, LDC and 48 V battery applied the same specifications as in the previous paper. And the Rear-Axle motor was 15kW and the BSG was 5kW(S, HA. 2018).

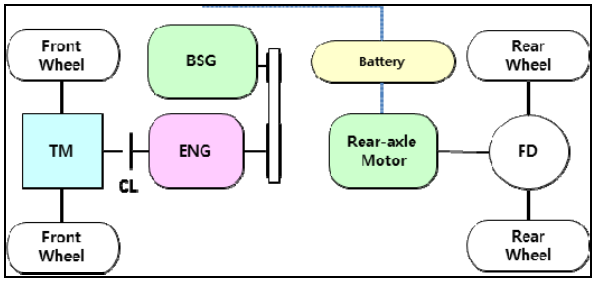


Figure 1: Target Vehicle

2.1. Engine model

Table 1 shows the specification of the engine of the target vehicle.

Table 1: Engine specification

Engine parameter	
Type	Inline 4-cylinder
Fuel type	Gasoline
Maximum torque (Nm)	168.7
Maximum power (kW)	99.54

Figure 2 represents maximum torque and BSFC(Brake Specific Fuel Consumption) at engine torque and speed conditions. As shown in Equation (1), the engine model outputs the engine output torque determined by the host controller (HCU) to the torque value between the maximum torque in the throttle maximum open state and the engine friction torque in the closed state.

$$T_{eng} = \min(T_{wott}(w_{eng}), \max(T_{HCU,eng} + T_{eng,idle}, T_{ctt}(w_{eng}))) \quad (1)$$

Where,

$$T_{eng.idle} = \max(k_p(w_{idle} - w_{eng}), 0)$$

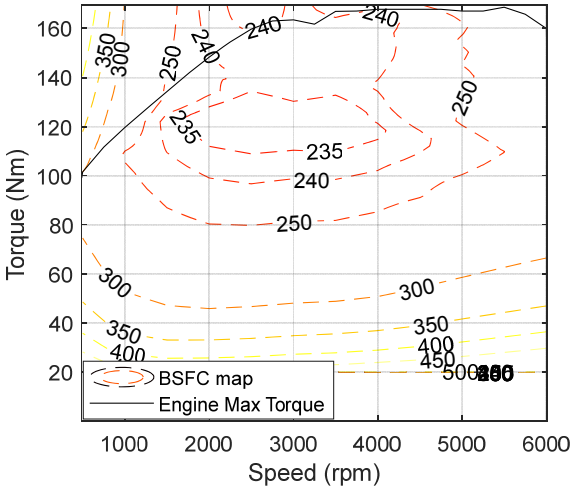


Figure 2: Engine maximum torque and BSFC map

2.2. Motor model

The motors are classified into two types according to their mounting positions. The BSG is connected to the engine by a belt. The rear-axle motor is connected to the rear reduction gear. The specifications of the two motors are as shown in Table 2.

Table 2: Electric motor specification

Electric motor parameter		
Type	BSG	Rear-axle motor
Max power (kW)	5	15
Max torque (Nm)	31.8	47.7
Base Speed (RPM)	1500	3000
Max Speed (RPM)	16000	12000

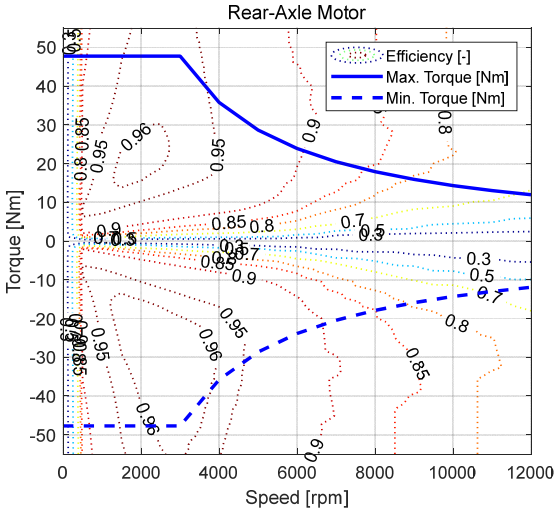


Figure 3: Max/min torque and Efficiency map of Rear-Axle motor

The drive motors mounted on the front and rear wheels have the same model structure, and the corresponding motor model reflects the motor torque command value determined by the HCU in the output. The output torque of the motor model is limited to the maximum drive or braking torque possible at the current motor speed. The torque of the motor is expressed by Equation (2), and the power of the motor has a relationship as shown in Equation (3).

$$T_{mot} = \min(T_{\max}(w_{mot}), \max(T_{HCU.mot} - T_{\max}(w_{mot}))) \quad (2)$$

$$P_{mot} = T_{mot} \times w_{mot} \times h^{k}(T_{mot}, w_{mot}) \quad (3)$$

Where,

$$k = \begin{cases} 1 & (T_{mot} \times w_{mot} < 0) \\ -1 & (T_{mot} \times w_{mot} \geq 0) \end{cases}$$

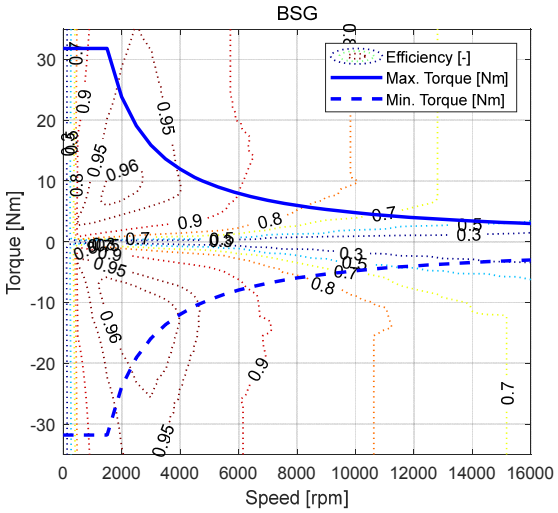


Figure 4: Max/min torque and Efficiency map of BSG

2.3. LDC model

LDC (Low Voltage DC-DC Converter) supplies power to a 12V electric field load. The 12V total field load power is 217W and the LDC efficiency is 0.95, both of which are fixed values. The LDC output power is given by Equation (4).

$$P_{LDC} = \frac{P_{Load}}{h_{LDC}} \quad (4)$$

2.4. Battery model

The 48V battery cell voltage is calculated from the open circuit voltage of the battery cell, the cell internal resistance, and the current. The temperature is assumed to be maintained at 40 ° C, and the internal resistance is determined according to the direction of the current and the SOC as shown in equation (5).

$$V_{bat.cell} = V_{OCV}(SOC) - R_{\text{int}}(SOC, \text{sgn}(I_{bat})) \times I_{bat} \quad (5)$$

Where,

$$R_{\text{int}} = \begin{cases} R_{discharge} & (\text{sgn}(I_{bat}) \geq 0) \\ R_{charge} & (\text{sgn}(I_{bat}) < 0) \end{cases}$$

The current of the 48V battery is calculated by the equation (6) with the sum of BSG power, rear-axle motor power and LDC power divided by the battery voltage.

$$I_{bat} = \frac{P_{mot.bsg} + P_{mot.rear} + P_{LDC}}{N_{cell} \times V_{bat.cell}} \quad (6)$$

Nominal voltage is 3.7V, and capacity is 23Ah. 13 battery cells are connected in series.

3. SUPERVISORY CONTROL ALGORITHM

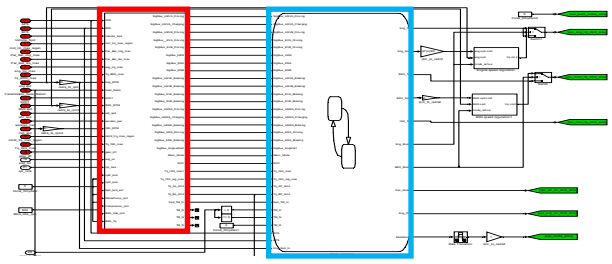


Figure 5: Supervisory control algorithm(Power distribution(red)/Mode decision(blue) algorithm)

As shown in the figure 5, the upper control algorithm was developed using Simulink, and consists of the mode decision algorithm and the power distribution algorithm of each mode. Details of driving mode and power distribution can be found in previous paper(S, Ha. 2017).

4. SIMULATION

4.1. Simulation case

The simulation case is defined as nine, as shown in Table 3. The method of control is divided into Constant torque and OOL(Optimal Operating Line) control. And they are divided into three generating energy for each. In addition, case without control of generating is included for fuel economy comparison.

Table 3: Simulation case by Generating control method

Control method : Constant torque	Control method : OOL torque
No Gen. (A)	
-10 Nm (B)	T_{ool} (F)
-7.5 Nm (C)	0.95 T_{ool} (G)
-5 Nm (D)	0.9 T_{ool} (H)
-2.5 Nm (E)	0.85 T_{ool} (I)

The Constant torque control method uses the BSG to generate electricity at a constant torque when the engine speed is greater than 800 rpm. The OOL control method runs the engine at the OOL torque when the engine speed is more than 800 rpm. In the OOL control method, run the engine at the OOL torque when the engine speed is more than 800 rpm. The BSG generates electricity from the driver's demand torque minus the OOL torque.

4.2. Simulation Result

The vehicle speed in all simulation cases is not significantly different from the speed of FTP75 cycle as shown in Figure 6. In Figure 7, SOC simulation results for all cases. The Final SOC for all cases is between 55% and 56%. Therefore, the fuel economy effects of the final SOC are ignored.

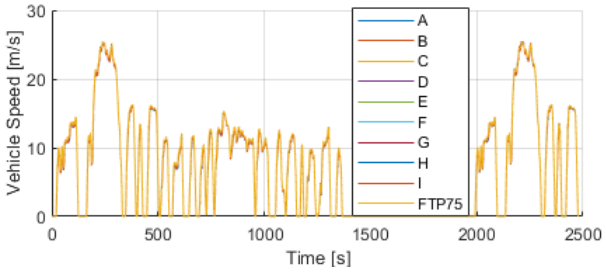


Figure 6: Simulation result - Vehicle speed

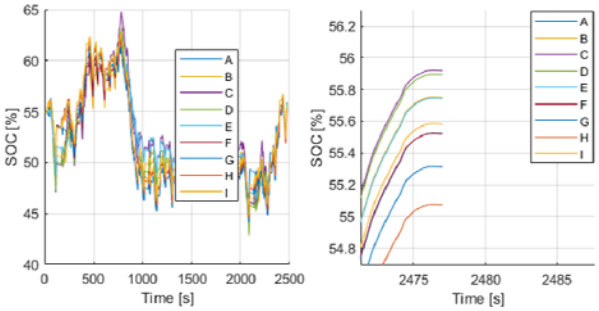


Figure 7: Simulation result - SOC of 48V battery

D

Table 4: Simulation result - Fuel economy

Simulation case	Fuel economy (km/l)	Fuel economy Improvement rate (%)
A	19.15	0
B	20.84	8.86
C	20.62	7.66
D	20.07	4.79
E	19.43	1.48
F	21.73	13.49
G	21.67	13.17
H	21.55	12.57
I	21.27	11.08

The fuel economy results for each case are shown in Table 4. Based on the results of Case A, the fuel economy improvement rate of Case B to I was calculated. Figure 8 is the engine operating point of Case A and is the basis for comparison with other cases. Figure 9 is for constant torque control (B to E) and Figure 10 is for OOL torque control (F to I).

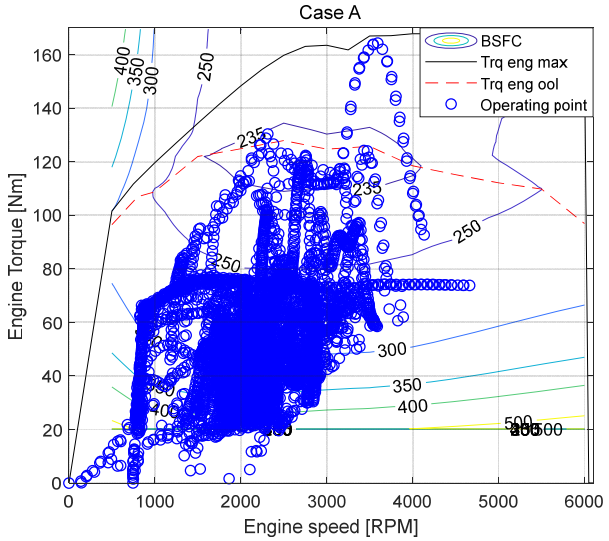


Figure 8: Engine operating point (Case A)

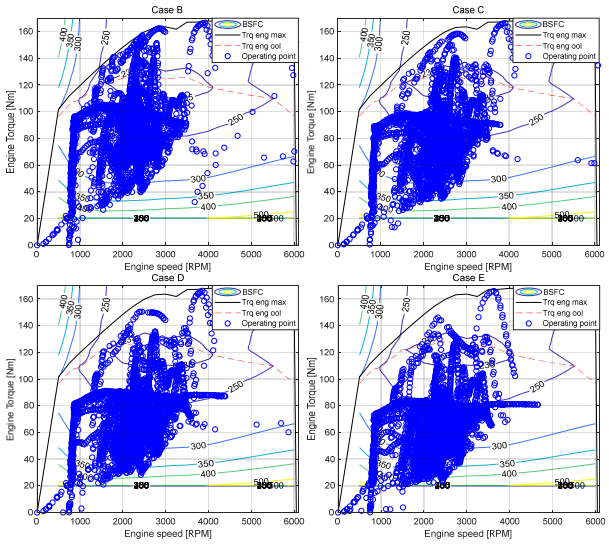


Figure 9: Engine operating point (Case B~E)

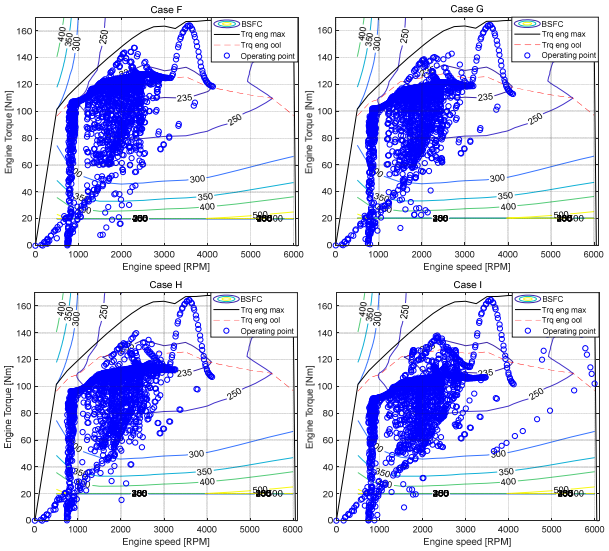


Figure 10: Engine operating point (Case F~I)

4.3. Analysis of simulation result

In Table 5, the power generation energy for each simulation case was compiled and calculated as shown in Equation (7) below.

$$E_{gen} = \int_0^T V_{BSG}(t)' \left| I_{BSG}(t) \right|' k_{gen}\, dt \qquad (7)$$

Where,

$$k_{gen} = \begin{cases} 1 & (Generating\ On) \\ 0 & (Generating\ Off) \end{cases}$$

Table 5: Generating energy

Simulation case	Generating energy (kJ)
A	0
B	1798
C	1746
D	1338
E	647
F	1661
G	1651
H	1636
I	1577

More power generation energy can be obtained from cases where power control has been carried out with high torque. However, the increase in torque and the power generation energy are not directly proportional. During EV operation, power generation is not possible by turning off the engine and using the BSG. Therefore, if the power generation torque is increased, the 48V battery can be charged faster and the EV mode can be driven more. So the time for BSG to generate electricity is reduced.

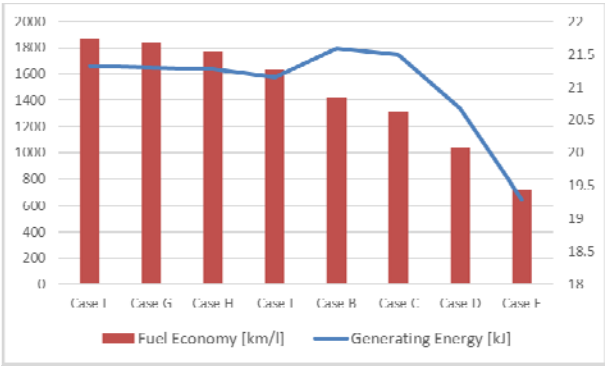


Figure 11: Fuel economy and Generating Energy

Looking at Figure 11, the case of OOL torque control (F~I) has a higher fuel economy result than that of Constant torque control (B–E). In particular, Case I has lower power generation energy but higher fuel economy than Case B and C. As Figure 12 shows, case C is caused by the engine running at a lower efficiency point than I. Figure 12 shows the engine operating points of Case E with the lowest fuel economy improvement and Case A without power generation control. Engines tend

to be less efficient at low torque. The BSG's power generation operation increases the operating torque of the engine to increase the efficiency of the engine, and uses the power generation energy to drive the EV to improve fuel economy.

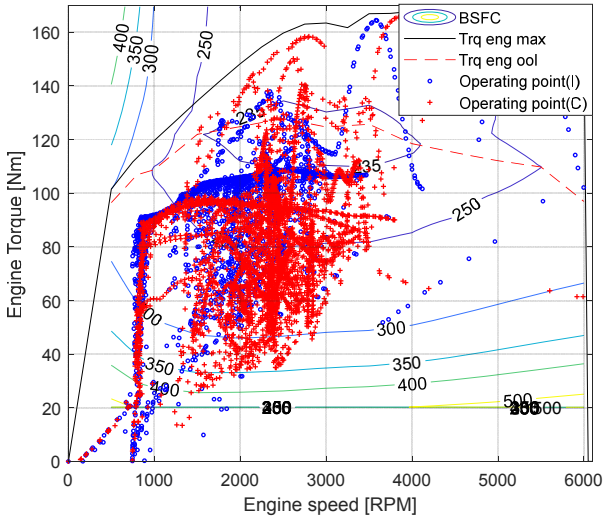


Figure 12: Operating point of Case I and C

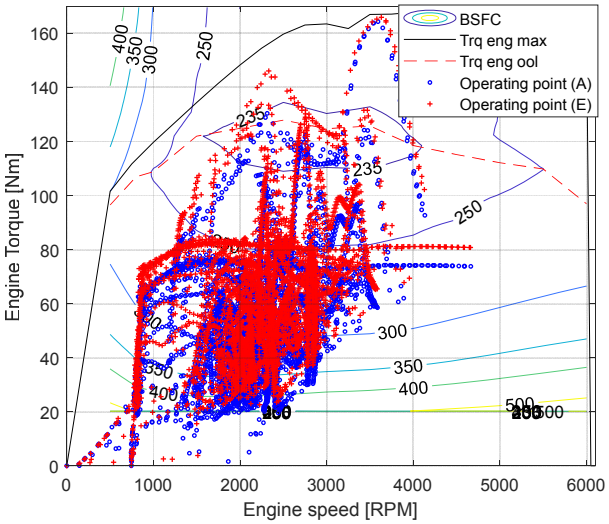


Figure 13: Operating point of Case A and E

5. CONCLUSION

In this paper, it was confirmed that the fuel economy improvement effect can be obtained when EV driven with the additional electrical energy obtained through the development of BSG in the 48V MHEV of P0+P4 structure. It is also important to obtain additional electrical energy through generating, but it is more important to run the engine at a more efficient operating point. However, since the 48V MHEV system has relatively low power, verification of the application of the OOL control in the actual vehicle, not in the simulation, is required.

ACKNOWLEDGMENTS

This work was supported by the Industrial Strategic technology development program, 20000578, A development of optimal eco driving system in HEV/PHEV based on connected vehicle environment funded By the Ministry of Trade, industry & Energy(MI, Korea).

REFERENCES

S, Ha., T, Park., W, Na,. and H, Lee,."Power distribution control algorithm for fuel economy optimization of 48V mild hybrid vehicle," Proc. of the Int. Conference on Modeling and Applied Simulation, 2017, ISBN 978-88-97999-91-1, 185-190.

S, Ha., W, Na,. and H, Lee,."Study on economy and regenerative braking of 48V mild hybrid vehicle with two motors," Proc. of the Int. Conference on Modeling and Applied Simulation, 2018, ISBN 978-88-85741-07-2, 22-26.

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.

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.

AVIO-REFUELING PROCESS SIMULATION IN AN AIRPORT ENVIRONMENT

Pasquale Carotenuto[(a)], Stefano Giordani[(b)], Jacopo Ponticelli[(b)]

[(a)]Istituto per le Applicazioni del Calcolo (IAC) – Consiglio Nazionale delle Ricerche, Italy
[(b)]Dipartimento di Ingegneria dell'Impresa, Università di Roma "Tor Vergata", Italy

[(a)]carotenuto@iac.cnr.it, [(b)]stefano.giordani@uniroma2.it, [(b)] jacopo.ponticelli23@gmail.com

ABSTRACT
The process of aircraft refueling has crucial impact in the performance of an airport. It is in fact of common knowledge that one of the most important indicators for benchmarking an airport is the punctuality of flights departure. To assure high results, the airplane service activities such as passengers boarding, baggage handling and aircraft refueling must not delay one another and the overall departure time. The scope of the proposed study is to produce an instrument capable of simulating the process of the aircraft refueling in the airport environment and to consider different scenarios and evaluate their impact in the overall performance. This tool has significant relevance for the company whom process we have analyzed, allowing it to be able also to evaluate easily and in a short period of time complex changes in the process.

Keywords: aircraft refueling, process simulation, scenarios evaluation, airport.

1. INTRODUCTION

The airport is the site we are referring to, with attention to the areas used for the handling of aircrafts and service vehicles, correlated to each other in order to guarantee the effective and efficient results for aircraft landing and take-off operations. These areas include the take-off and landing runway, taxiways and lay-bys. The management phase of an aircraft transiting in an airport environment is a topic of particular interest at a scientific level and the literature offers several studies on the different activities that are carried out in this field (Ruamchat *et al.* 2017).
First of all, there is the handling of the aircraft in the moments following the landing, with the objective of transferring the aircraft from the runway to the lay-by assigned to it from the control tower, taking into account especially the traffic conditions. This extremely delicate phase must be carried out guaranteeing the maximum possible rapidity and the least bulk of the road conditions; especially for the larger types of aircraft, a parking area must be guaranteed as close as possible to the landing strip, also considering the available roads that often have to be modified to adapt to the ever-increasing size of the aircraft. Given the delicacy and importance of this phase, there are several studies in the literature that analyze it and propose different mathematical models able to efficiently coordinate these operations (Samà, D'Ariano, Corman, & Pacciarelli 2018), and which also propose models that can serve as an instrument for the optimal allocation of the stalls (Guépet, Acuna-Agost, Briant, & Gayon 2015) as an instrument serving the control tower.

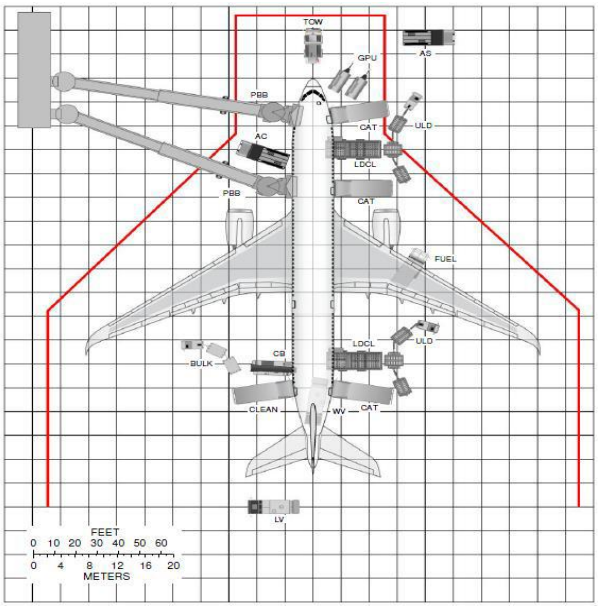


Figure 1: The "Ramp Area": diagram to illustrate the complexity of the integrated system (K.Wing, Cloutier, & Felder, 2015)

After parking the aircraft at a lay-by, also referred as "Ramp Area" (see Figure 1), the operations performed by service providers such as catering restoration, cleaning services, baggage loading and unloading, passenger descent and ascent, and refueling are activated. This area is often congested. It is due to the fact that all the airline companies are day by day more focused on being in the lay-bys the least possible amount of time, and so all the different service providers are forced to execute their activities in a small portion of time, causing a lot of traffic between themselves. The person responsible for coordinating all these operations in the ramp area is called "Ramp Agent". He/She has the role of assuring two conditions during the work of the service providers:

- Rapidity, making sure that the operators do not interfere one another.
- Safety & Security.

In the literature, many studies have focused their attention on the single activities done in the ramp area and on the overall efficiency of the coordination of these. For example, Kierzkowski & Kisiel (2017) study the passengers boarding procedure and how the human factor is crucial in the implementation of one methodology or another; others studies propose the development of simulation tools to support the analysis, for example, of the baggage handling phase and its performance (Cavada, Cortés, & Rey, 2017), the optimization of the operators used for aircraft refueling in airport parking area (Carotenuto et *al*. 2019),, or proposing refueling optimization tools capable of finding the optimum number of avio-refuelers and the best routes, given a specific market demand (Babić, 1987).
None of these, however, keeps in consideration the contemporaneity of the distinct activities, resulting in a probable underestimation of the complexity of the system. The ramp agent comes in handy right where the human factor is of primary importance and without whom the coordination of these activities would result poorly managed (Wing, Cloutier, & Felder, 2015).

2. SYSTEM DEFINITION

The proposed study was developed for a company operating at an important airport in Italy, in charge of refueling aircraft at a Ramp-Area (we will call it BlueSupply for data confidentiality). At the airport, BlueSupply owns two cisterns capable of storing 200,000L of jet-fuel and a tank truck fleet with an overall store capacity of 240,000L. It is the company duty to restore the product in the cisterns ordering it from the referenced refinery according to the estimated demand on a daily base.
The following factors in particular make the complexity of the activities done by the company grow:

- Jet-fuel decantation;
- Bridger arrival rhythm;
- Quality standards;
- Customs controls;
- Distance of the ramp areas from the trucks stand by area;
- Strict deadlines.

2.1. Jet-fuel decantation

The handled fuel-oil product is called Jet A1, and it is a type of fuel-oil used in the aviation field. For its nature, it is crucial for safety and security reason that jet fuel is left still for a significant amount of time every time it is moved from any source to any destination. In particular, the referenced periods are of two kinds:

1. 2 hours after transferring jet fuel into the cisterns.
2. 10 minutes every time it is moved from or to the tank trucks.

It is not sufficient to have the right capacity of storing, but what is crucial is the right choice of the times when the operators replenish the trucks from the cisterns.

2.2. Bridger arrival rhythm

With the word bridger, we are referring to the truck that carries the product ordered from the refinery to the depot. In the scenario that is presented in the referenced depot, every bridger has a capacity of approximately 35,000L and the number of bridgers arriving every day varies accordingly to the estimated demand.
Given the decantation time durations, it is fundamental to choose the right cistern in which transfer the product brought, because if a correct choice is not taken, the situation where both cisterns are inaccessible might occur.

2.3. Quality standards

As well as for the decantation time periods and the correct management of the bridgers, a considerable amount of effort is spent in the activity of assuring high levels of quality of the jet-fuel. What must be verified in different parts of the supply chain is the respect of the following factors:

- Absence of water in the product;
- Absence of debris (clear and bright);
- Correct density;
- Correct electric conductibility;

The achieving of determined standards allows the company to assure the minimum possible risk in terms of accidentality inside the airport (episodes of combustion) and during the flights (episodes of water icing inside the aircraft tanks at high altitude). A lot of time is spent in the verification of this standards.

2.4. Customs controls

Given the position of the reference company's depot, which is outside the airport area used for the aircrafts handling and for the carrying out of the airplanes services (called "Air Side"), it is a strict obligation that of assuring high standards of security through the customs controls of everyone going into this area from outside. This results in generating lot of traffic in this area for high levels of demand.

2.5. Distance of the ramp areas from the trucks stand by area

As for the airport structural and procedural limitations, the refuelers are obliged to come back to stand by in an area far from the lay-bys waiting there for the request for refueling of an airplane captain. This limitation shrinks the overall throughput capacity given the fact that every operator spends an average time of 4 minutes every time he has to go from the standby area to the ramp area, and vice versa.

2.6. Strict deadlines

As it is widely known, flight departing times are a factor of extreme relevance for an airline company

performance, and for this, it is not tolerated any delays due to the services activities. For this reason, not only the refueling operations must assure high standards of quality in terms of service and product, but they have to be made in the shortest possible amount of time and as soon as possible after the request of the captain.

3. CURRENT SCENARIO

In the airport we are referring to, it is not present only the company whom process we have analyzed, but there are two other companies (owning two other depots) that also perform the aircraft refueling activity.
The first two most important aspects characterizing the system that have been considered are the following:

1. The three existing companies do not share the clients but each one of them has a unique set of airline companies to serve. A client can anyway be refueled by another company in situations of emergency.
2. With the current resources at hand, company BlueSupply is not capable of meeting all the demands of its clients. For this reason, it gives in outsourcing a part of its set of clients to the other two company.

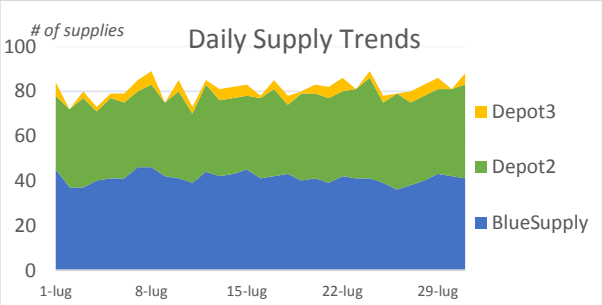


Figure 2: Daily Supplies Trend Distribution

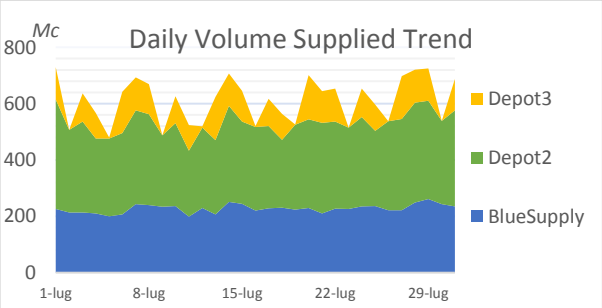


Figure 3: Daily Volume Supplied Trend Distribution

Figure 2 and Figure 3 show the number of supplies and the amount of jet-fuel refueled, respectively, day by day during a typical month. Figure 2 shows that BlueSupply (in blue in the chart) is capable of meeting only about half of the demand of its clients.

4. SIMULATION APPROACH

The proposed study wants to offer a tool capable of simulating the as-is process in a way that it can be used at a strategic level for evaluating how the system reacts to the change of exogenous and structural factors previously considered at a managerial stage.
For our scope, we utilized the discrete events simulation software Arena (Kelton, Sadowski, & Zupick, 2014) that, thanks to a user-friendly interface, permits also to non-expert users to interact with it and evaluate different scenarios. Moved by the above reasons, we have followed the following steps:

- Development and implementation of the simulation model for the as-is process and validation of it;
- Evaluation of the impact that changes in exogenous factors can bring to the performance;
- Evaluation of the impact that changes in structural factors can bring to the performance.

5. THE AS-IS MODEL

For the representation of the process and for having a better vision of the overall system simulation, we have split the whole process in six different areas that are linked together in the model. These are:

- Aircraft arrival and stand-by;
- Depot management;
- Customs controls;
- Tank trucks stand-by area;
- Truck choice decision process;
- Refueling process.

The process simulation has been made only on real data obtained by the reports written by the operators. The implemented model read the data from an external spreadsheet such as excel, access or any another compatible software, and reproduce the real system performance accordingly to these.

5.1. Aircraft arrival and stand-by

In this section (see Figure 4, Appendix A), it is represented the process of aircraft arrivals. We have represented in the simulation model what is of major interest for our purpose that is the arrivals of the refueling requests accordingly with the refueling report data written by the operators. The first part of this area is the one responsible for reading the above-mentioned data; in particular, each entity generated will read the specifications of each airplane that consists of:

- Flight identification number;
- Quantity of fuel requested (which will be communicated only upon arrival of the tank to the aircraft);
- Historically requested quantity of fuel;
- Landing time;
- Historical arrival time of the tank alongside the aircraft;
- Historical communication time of the quantity to be supplied;
- Expected departure time (checking that refueling does not cause delay);
- Scheduled refueling area.

After reading the data from the excel spreadsheet, it is simulated the phase of identification of the queues generate by the different arrival times. It is necessary to identify the part of flight plan that the operators analyze to operate accordingly to it. The reason behind this is to replicate how the operators estimate the time when a refueling request will arrive so to be in the best position possible and react to it as fast as possible.

5.2. Depot management

In this section of the model (see Figure 5, Appendix A), every action that takes place inside the depot is represented, so the following activities are simulated:

- Operators shift coordination, accordingly with the real work force existing;
- Tank trucks management, considering their stand-by and their load from the cisterns;
- Bridgers arrival coordination;
- Jet-fuel decantation.

As regards the management of the bridgers and the decantation of the fuel, a crucial activity is represented: it is that of the coordinator of the warehouse, responsible for deciding which tank to load and when, in order to ensure the possibility of having at least one of the tanks free (not in decantation) with enough product inside. Since the company lacks a written procedure for managing this task, it is entrusted to the depot supervisor experience the above-mentioned choice, and so does the model, approximating the real decisional process.

5.3. Customs controls

This section of the simulation model represents the process of customs control (see Figure 6, Appendix A) that occurs every time an operator goes into the apron coming from its outside. What is here important is the distinction of two recurring moments in the day, depending on the traffic that the operators find at the gates. It was in fact analyzed that in two specific time periods (on average from 8:00 AM to 9:00 AM, and from 1:00 PM to 2:00 PM) the congestion causes the average controls duration to rise from 2.5 minutes to 10 minutes.

5.4. Tank truck stand-by area

In this section (see Figure 7, Appendix B) it is represented the phase where the operators wait in standby the arrival of the communication of the permit of the refueling of an aircraft. It is due to the airport internal procedure that they have to wait in this area and that it is not allowed to wait near the ramp areas unless they wait only for a little amount of time, for example going from a previously refueled airplane to the next.

5.5. Truck choice decision process

This is the most crucial section of the simulation model, representing the decisional process undertook by the operators every time they finish an activity (see Figure 8, Appendix B). It is necessary for letting the simulated process be as flexible as the real one. It replicates the decision of the operators in choosing between approaching a ramp area in advance, going from the area of the last refueling activity to another one without passing for the stand-by area, or come back to the depot to load the tank trucks.

It is built accordingly to a survey made amongst the operators at the depot with which it was our intent to understand their behavior in approaching the process and in choosing to do an activity rather than another.

5.6. Refueling process

In this last section (see Figure 9, Appendix B) we represent the refueling activity in a ramp area. It is done when the flight captain has submitted the request, and accordingly to the quantity reported by the operators. For the truck's pump output rate, it has been utilized the specification of the truck owned by the company to be close to reality as much as possible.

Once the refueling is done, the model reacts by writing a "report" (in excel format) of the flight refueled in order to:

1. Report if the flight has been refueled on time or if has been caused some delay;
2. Report data necessary to extrapolate the performance of the process and to validate the model.

5.7. Validation

Once the model was run, it was possible to compare the output simulation data with the existing output data of the real case, in order to validate the model. With this validation in mind, given the available data it was thought to use two KPIs (Key Performance Indicators) of the system, which are capable to identify the performance in two aspects: the timing of the supplies, referring to the essential punctuality of this activity in the airport environment, and resource occupancy, specifically the fraction of time the operators use to carry out operations and activities aimed at the refueling itself (such as nearby parking, reporting, and fuel transfer to the aircraft) over the total available time at their disposal. As we will see later, the core business of the company is not centered on the performance identifiable by these two indices but, for the purposes of comparison, they are particularly interesting as they allow us to assess the validity of the simulation in terms of approximation of behavior and decisions implemented.

These two KPIs were identified:

- Average of the gaps (delta) between the expected departure of a flight and the moment the aircraft is released (end of refueling).
- Total utilization rate.

5.7.1. Average of the gap between the expected departure of a flight and the refueling finish time.

Denoting with δ_i the gap between the expected departure of a flight (P_i) and the end of refueling for that same flight (FR_i), in which the subscript i identifies the i-th flight:

$$\delta_i = P_i - FR_i \quad (1)$$

The KPI can now be defined as the average between the δ_i of all the n served flights:

$$\Delta = \frac{1}{n}\sum_{i=1}^{n} \delta_i \quad (2)$$

5.7.2. Total utilization rate.

It is defined as the ratio between the total time spent for refueling operations and the total available time. This does not indicate how much each operator has been operating during the day, even operationally uninteresting, but rather how much time is spent on "useful" operations to the final objective; the lower it is, the more it indicates a poor management of the dynamics inside the warehouse (such as the timing of loading of tankers). Denoting with AR_i the starting time of refueling flight i, keeping in mind that each operator's turn sees him present for t minutes a day, and that there are 5 operators per day divided into as many shifts for 7 days a week, the total utilization level GU is equal to:

$$GU = \frac{\sum_{i=1}^{n}(FR_i - AR_i)}{t*5*7} \quad (3)$$

5.7.3. Results.

After the introduction of the two indicators we can see in Figure 10 how the results obtained from the simulation are extremely like the ones obtained by the real report, indicating the simulation as a good approximation of the reality.

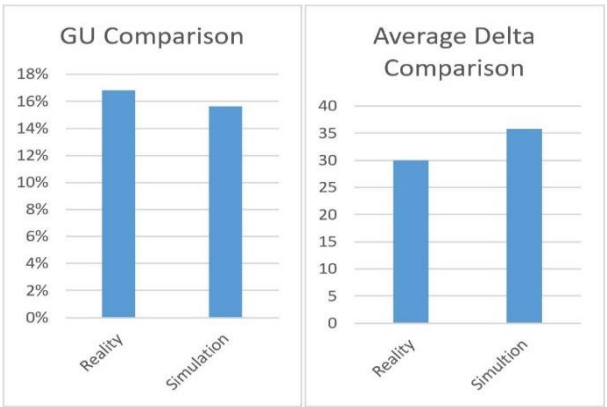


Figure 10: As-Is process simulation validation

6. TO-BE SCENARIOS ANALYSIS

The core scope of the proposed study is to generate a decision-making support tool capable of evaluating the impact of different changes in the system on the overall performance. For this reason, we have analyzed different possible changes in the operations and how they can change the performance; to evaluate this, we have simulated 4 different to-be scenarios and compared it with the indicators above-mentioned. These hypotheses include 2 changes in the structural factors such as:

- Doubling the flow rate of pumps for the loading of the cisterns;
- Possibility of loading two tank trucks at the same time.

The two other changes regard the demand and in particular:

- Customer resumption test;
- Capacity estimation based on the entire customer list.

As we can see from Figure 11, the overall performance is unchanged. It is because what we have changed is a non-critical factor in a way that an alteration of it does not affect the system capacity.

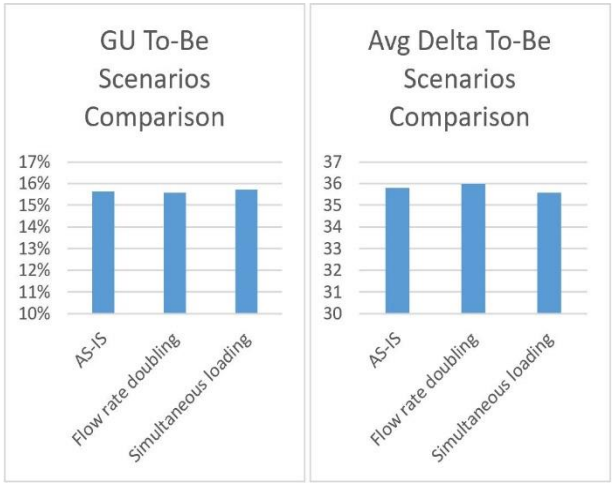


Figure 11: Structural To-Be scenarios analysis

Figure 12: New clients list scenario comparisons

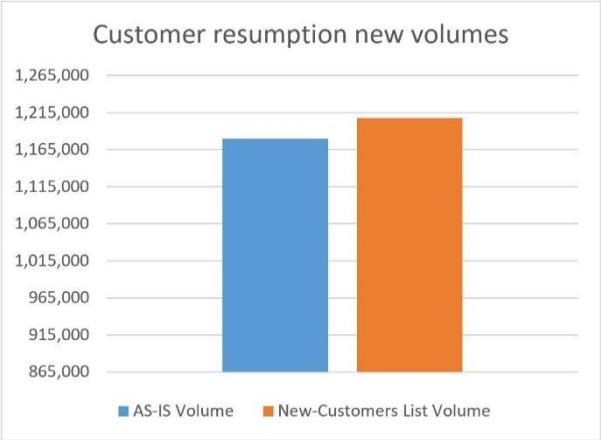


Figure 13: Customer resumption new volumes

As it is shown in Figure 12, also the customer resumption test leads to the same indicators. Despite this, what is more interesting is the fact that given the same indicators, the volume of the refueled product has increased without varying the resources (Figure 13).

7. CONCLUSIONS

The simulation model proposed is itself very complex, being constituted by a very large number of modules, representative of the activities of the process, and including the simulative representation of the behavioral rules used by the operators on field.

During its use, the structure of the model must be always kept in mind; in fact, it tries to approximate the real system with high flexibility miming all the behavior rules adopted by the operators. If it is not properly used, by narrowing operator's behavior rules there is the risk of returning results and performance distant from reality. In the construction of the model, it was necessary to resort to the insertion of "rules" identified on the field by surveys, necessary to analyze the behavior of operators in extreme situations, non-frequent but possible. It was therefore natural to deepen their use in different scenarios, where it might be necessary to insert new behavioral rules or modify those already tested. By correctly implementing the rules acquired in the field, the simulation was particularly fast and efficient, allowing the analysis of various scenarios in a short time.

From the obtained results, we can conclude that the model can give excellent feedbacks and can be used as a decision-making support tool. By the last analysis we can highlight the possibility to increase the served demand carrying out an in-depth analysis on the own list of served airplane.

This last consideration focuses the attention on the two critical points of the model:

- the possibility of stiffening the process if not correctly approached;
- the lack of a database suitable for such an analysis.

Therefore, considering the results obtained, it is possible to follow two evolving paths:

1. The first concerns the continuous development and improvement of the proposed instrument, so as to enhance its characteristics as an excellent managerial tool that allows the rapid and efficient evaluation of strategic or tactical changes, providing the results that validate or refute the reasons of the proposed investments or changes in question.
2. The second one concerns the extension of this decision support tool towards an operational tool in which the model is no longer thought only as a strategic level tool, but rather as an operational verification tool and, in a more advanced phase, as a tool for optimization and preliminary allocation of resources.

Following this second path, we can evaluate the possibility of use it as an optimization tool. In fact, we can consider to use the simulation tool together with a scheduling model. In this scenario a scheduling model, analyzing the results previously obtained by the simulation, could be able to identify an optimal use of the available resources trying to improve what is be done.

REFERENCES

Babić O. (1987). Optimization of refuelling truck fleets at an airport. *Transp. Res. Part B* 21(6), 479–487.

Carotenuto P., Giordani S., Salvatore A. & Biasini A. (2019). Resource planning for aircraft refueling in airport parking area. *Transp. Research Procedia*, (37) 250–257. DOI 10.1016/j.trpro.2018.12.190.

Cavada J.P., Cortés C.E., & Rey P.A. (2017). A simulation approach to modelling baggage handling systems at an international airport. *Simulation Modelling Practice and Theory* 75, 146–164.

Guépet J., Acuna-Agost R., Briant O., & Gayon J.P. (2015). Exact and heuristic approaches to the airport stand allocation problem. *European Journal of Operational Research* 246(2), 597–608.

Kelton W. D., Sadowski R., Zupick N., (2014) *Simulation with Arena*. McGraw-Hill Education, 6th ed., pp 635, ISBN-13: 978-0073401317.

Kierzkowski A., & Kisiel T. (2017). The Human Factor in the Passenger Boarding Process at the Airport. *Procedia Engineering* 187, 348–355.

Ruamchat K., Thawesaengskulthai N. & Pongpanich C. (2017). A method of prioritizing quality improvement in aviation refuelling services at airport. *Advances in Mechanical Engineering* 9(6), 1–12. DOI: 10.1177/1687814017706436.

Samà M., D'Ariano A., Corman F., & Pacciarelli D. (2018). Coordination of scheduling decisions in the management of airport airspace and taxiway operations. Transportation Research Part A: Policy and Practice 114, 398–411.

Wing A.K., Cloutier R.J., & Felder W.N. (2015). Ramp Area Support System: Limitations of Modeling Approach to an Airport Apron Area. *Procedia Computer Science* 44, 354–362.

APPENDIX A)

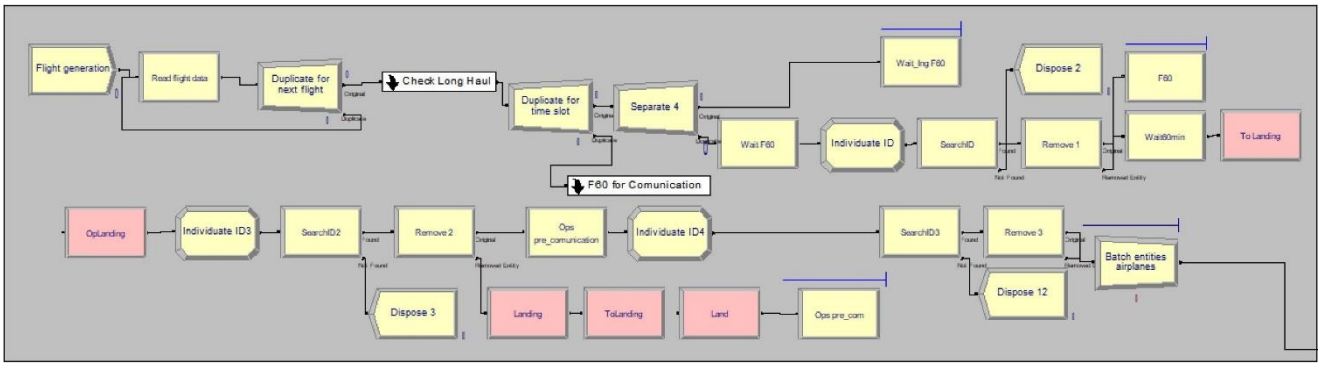


Figure 4: Airplanes arrival e stand-by

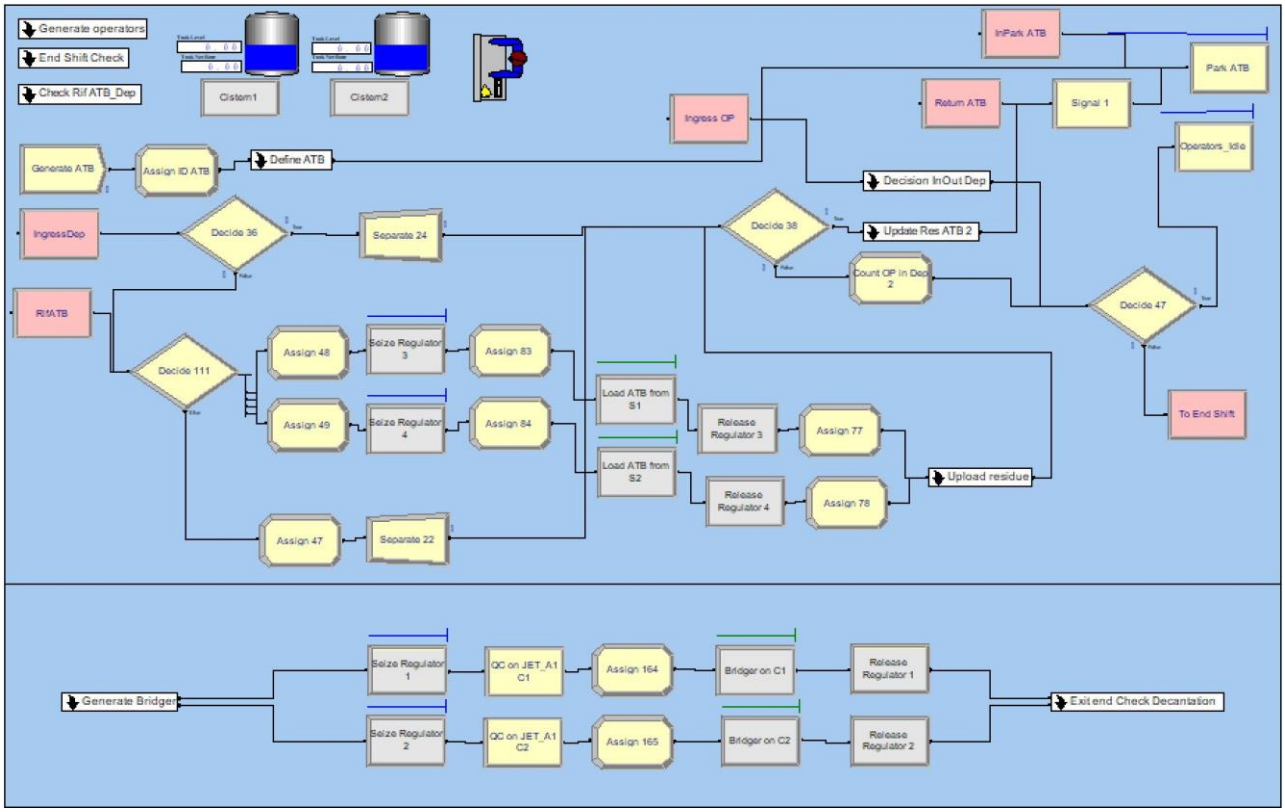


Figure 5: Depot management

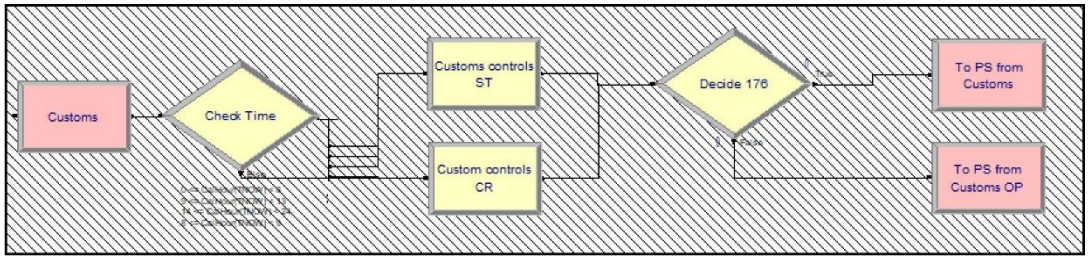


Figure 6: Customs controls

APPENDIX B)

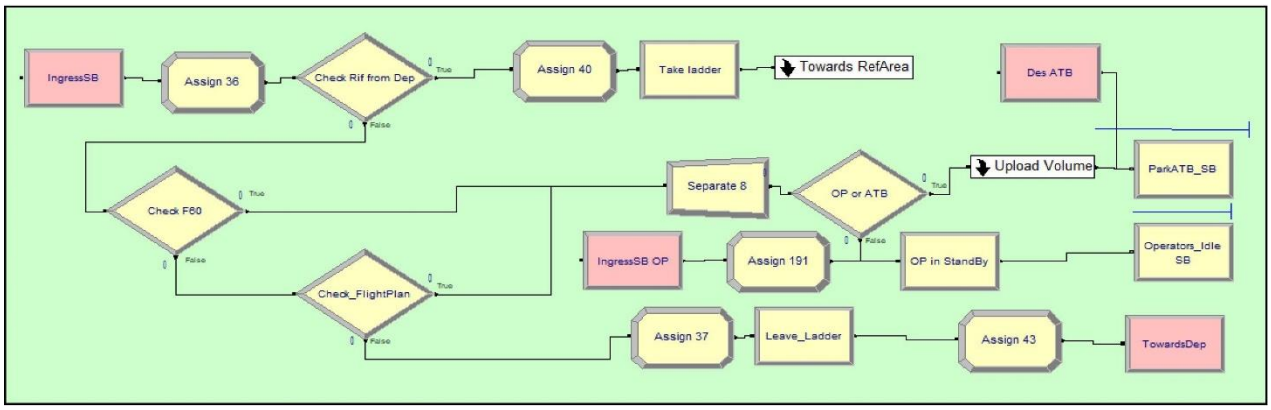


Figure 7: Tank truck stand-by area

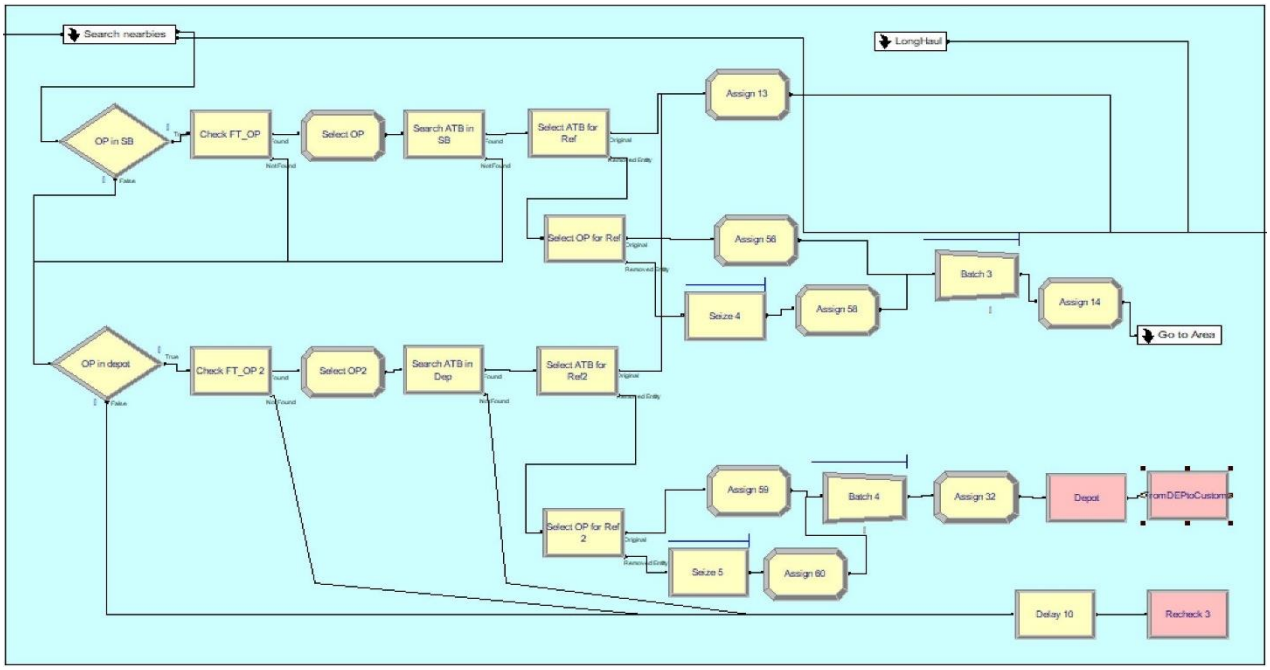


Figure 8: Truck choice decision process

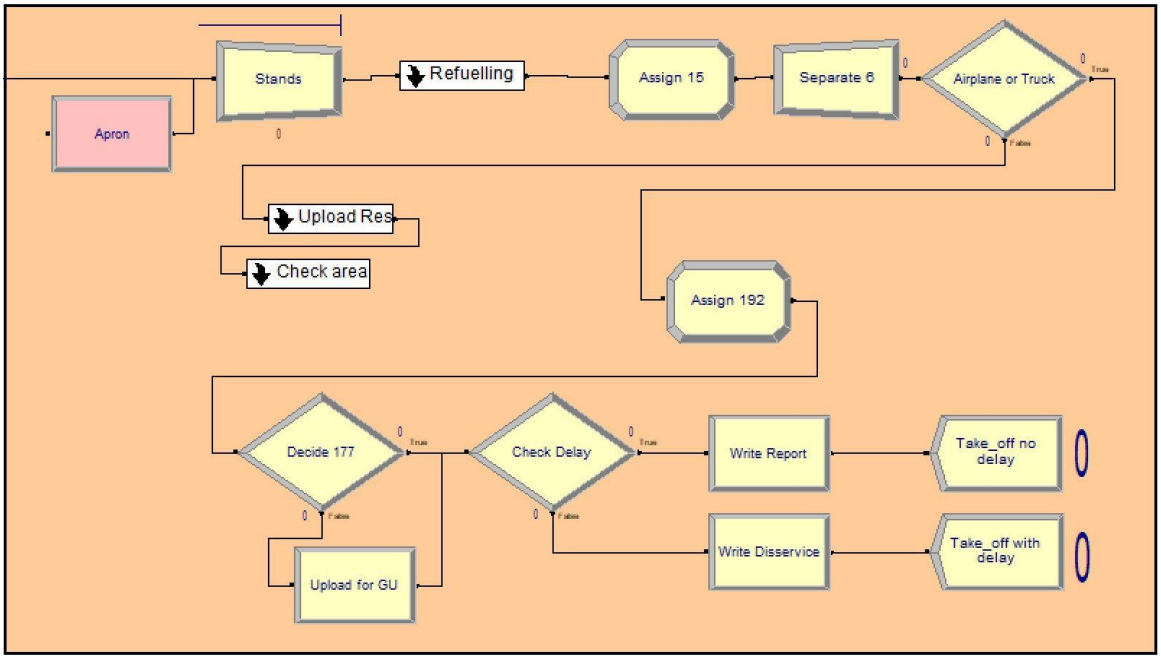


Figure 9: Refueling process

LEVERAGING ON THE DIGITAL TWIN FOR IMPROVING RETAIL STORE DAILY OPERATIONS MANAGEMENT

Yasmina Maizi[(a)], Ygal Bendavid[(b)] Janosch Ortmann[(c)]

IoT Lab., CRI^{2}GS
École des Sciences de la Gestion – Université du Québec à Montréal

[(a)]maizi.yasmina@uqam.ca, [(b)]bendavid.ygal@uqam.ca, [(c)]hortmann.janosch@uqam.ca

ABSTRACT
With the fast development of IoT technologies and the potential of real-time data gathering, allowing decision makers to take advantage of real-time visibility on their processes, the rise of Digital Twins (DT) has attracted several research interests. DT are among the highest technological trends for the near future and their evolution is expected to transform the face of several industries and applications and opens the door to a huge number of possibilities. However, DT concept application remains at a cradle stage and it is mainly restricted to the manufacturing sector. In fact, its true potential will be revealed in many other sectors. In this research paper, we aim to propose a DT prototype for in-store daily operations management and test its impact on daily operations management performances. More specifically, for this specific research work, we focus the impact analysis of DT in the fitting rooms' area.

Keywords: Simulation, RFID Technology, IoT, Retail, Digital Twin in organization

1. INTRODUCTION

As the vision of the Internet of Things (IoT) is progressively becoming a reality, the worldwide number of connected devices is expected to increase and reach 125 billion by 2030 (IHS 2017). The idea that any object can be equipped with technology to become a computing device, interacting autonomously, in real time, with its environment is happening in various sectors. Accordingly (Bain & Company 2018) suggests that the combined markets of the IoT (i.e., hardware, software, systems integration and services) will grow to about $520 billion by 2021, more than double of $235 billion spent in 2017. In line with this emerging trend, the concept of Digital Twin (DT) has been brought to the scientific community at the beginning of this millennium by Michael Grieves and lately formally described in (Grieves and al. 2017). The idea underlying a DT is that each object, process or system can be replicated by a digital representation (Gartner 2019a), therefore creating two similar systems (twins) connected by a permanent real-time data exchange. The physical system produces operational data captured in real time by using technologies such as Internet of Things (IoT) technologies. This information is stored in a database that, depending on the stage and structure of the digital twin, processes it and then feeds a simulation model that replicates or mimics physical system operations in real time. Then, several runs of simulation scenarios allow a decision maker to adjust physical system parameters as well as operational rules to improve the efficiency and effectiveness of the physical system.

The advent of Internet of Things (IoT) technologies has largely contributed to the development of DT and their application show a great potential for operations management in different sectors (Gartner 2019a). However, the development and use of DT is still at its cradle stage with limited fields of applications. Moreover, it is interesting to note that DT is and will be among the strongest technological trends rising in the near future. Additionally, the application of DT concept won't be limited to the manufacturing sector anymore and will quickly take off to be deployed in a wide range of sectors as healthcare operations management and organizations management. In fact, in a recent survey from Gartner Research (2019b), the authors suggested that "the digital tweens are entering mainstream" in all kinds of organizations, as an increasing number of companies that have implemented IoT will have deployed a DT in the next year. It is in this vein that we focus our research project and suggest a DT framework for daily retail operations management.

Moreover, The Internet of Things (IoT) is attracting growing interest in the apparel retail sector. With the help of (IoT) technologies, retailers can now take advantage of new possibilities offered by various automatic data capture technologies such as Radio Frequency Identification (RFID) for item identification and tracking or Bluetooth Low Energy (BLE) technologies, for tracking consumers within stores. Implementing these technologies provides real-time visibility into customer behaviour and in-store items movements; enabling retailers improve the performance of store operations management.

In this research work, we aim to develop a DT framework for apparel retail store daily operations management. We first determine the appropriate technologies that can be

implemented at different levels of the store. Second, we develop a discrete event simulation model to represent daily store operations. Third we capture the store data through a physical prototype existing in our living IoT Lab. Finally, we run several scenarios and analyze them to improve our understanding of in store operations management performances.

The remainder of this paper is organized as follows: In **section 2**, we explore the existing literature regarding: i) RFID/IoT technologies application in retail, ii) modelling retail store operations and iii) applications of DT concept. **Section 3** describes the design science research approach followed to investigate the business case of applying a DT in a retail environment. We first, describe the problem from the field and the objectives of our DT model. This is followed by the design and development of our DT prototype aiming to (a) capture real-time customers' and products movements in a store (b) run a simulation model that replicates daily operations (c) automatically integrate collected RFID/IoT data to the simulation model (c) analyze this data to improve daily store operations management. **In sections 4 and 5** we share the expected outcomes of our DT model, present concluding remarks and future work.

2. LITERATURE REVIEW

This section is dedicated to the following fields literature review: i) RFID/IoT technology in retail; ii) RFID/IoT suitable solutions for in store daily operations management; iii) modelling retail operations, in particular those related to data analysis approaches & simulation studies; iv) finally digital twin concept and its applications fields.

2.1. RFID/IoT Technology

In recent years, among the technological solutions proposed to increase operations efficiency of apparel retailers', retailers chose to rely on automatic data capture technologies such as (i) radio frequency identification (RFID) technologies, for products; (ii) Bluetooth Low Energy (BLE) and Wifi technologies for in-store customer's tracking, through the use of their mobile phone or smart watch (Zaino, 2016). The gained interest in these technologies can be easily explained through the motivation of remaining competitive (Roberti 2018) by using real-time information to: (a) improve the customer experience (e.g., interactive journey) and (b) improve operations management performance (e.g., reduce operating costs, improve inventory management). Today we find more and more implementations of "connected stores", especially in the consumer sector, including the most publicized: Amazon Go and its promise of "no lines, no checkouts, and no registers" that was finally opened to the public in Seattle in January 2018 (Deirdre, 2018).

2.2. RFID/IoT Suitable Solutions for in Store Daily Operations Management

More specifically, in the retail/apparel sector, a recent study by Zebra (2017) has shown a very strong trend in the adoption of IoT technologies. Indeed, 73% of surveyed managers confirm that they are ready to adopt IoT solutions by 2021. This trend is similar to the trend of embracing data analysis solutions captured by IoT devices, with nearly 60% of managers planning a budget for acquiring such solutions.

Regarding RFID technologies, Zaino (2016) brought a similar conclusion suggesting that: «major retailers in Europe, North America and South America are embracing the technology to track and manage apparel and footwear, in order to improve inventory accuracy and provide customers with an omnichannel «anytime, anywhere shopping experience». Indeed, since the first RFID initiatives in the retail sector such as Walmart, Gerry weber, Macys, Zara, Decatlon, Celio, Scalpers, Marks & Spencer, and Kaufhof are leaning towards such technologies to identify customer preferences and customer behaviour. Numerous business cases can be found in the web site of (RFID Journal).

However, if the implantations of IoT / RFID technologies are increasing in recent years, their application is limited to basic operations management processes (e.g., inventory management, shelve replenishment) or used for marketing purposes (e.g., real-time advertising). Therefore, we still do not have a comprehensive analysis or study on the impact of RFID/IoT technologies on the floor space utilization and fitting room utilization. This is especially important, since fitting room are a classic bottleneck as well as a privilege space to interact with the consumer and understand its preferences.

2.3. Modelling Retail Store Operations

Most of the research work related to simulation in the retail is dedicated to modelling supply chain operations. Some studies relate to in-store operations management. Among them: Miwa et al. (2008) provide a very interesting study where data used for building in-store operations' simulation model is gathered through point of sales system. They use a discrete-event simulation model built with Arena. They assume that point-of-sale data is gathered every day and derive store performance measures using simulation outputs. It is interesting to note that even though point-of-sale gathered data can provide interesting information regarding the number of customers purchased items, checkout queue statistics, they miss a part of crucial information related to customers' in-store experience and behaviours. Wang Huiru et al. (2018) present an agent-based simulation model to represent wine consumers' behaviour and understand customers' motivation to select a specific wine among different wine brands. Shandong et al. (2018) provide a thorough literature review on retail store operations and identifies five main axes attracting research attention and activities: demand forecasting, inventory management, assortment and display,

employee management and checkout of operations. One interesting concluding remarks of this research work is that research studies that deal with multiple store operations decisions, multiple themes and multiple products are rare. Moreover, traffic counters and real-time data are not efficiently used and open the door to several research opportunities. Tsai et al. (2015) provide a research work related to ours in which they study the impact of real-time analytics on checkouts, shelf replenishment and shelf allocation. They use a data mining approach to optimize shelf space allocation, taking into account customers' purchase and moving behaviours. Finally, Nervo and al. (2019) study the impact of RFID technology on retail operations and use a system dynamics simulation approach to assess the economic impact of item-level RFID technology on retail. They show that RFID technology can be used as a competitive advantage driver.

2.4. Digital Twin

Although the concept of digital twin is still at an infant stage, the literature shows several research papers revealing different terminologies used for describing the digital twin and suggest several definitions. However, the basic concept remains the same (Grieves and al. 2017). Werner Kritzniger et al. (2018) conducts a comprehensive literature review and classification of digital twin in manufacturing according to their level of integration. They provide three subcategories of DT, according to their level of data integration: (i) Digital Model where there is no automatic data exchange between the physical system and the digital system. The transfer of data is manually handled (ii) Digital Shadow, where there is an automatic data flow exchange from the physical system to its digital representation, but a manual data exchange from the digital system to the physical system (iii) finally, the digital twin, where automatic data exchange is in both directions. It is interesting to note that according to their study, only 18% of the studied papers really describe a digital twin and only one case study was implemented in a laboratory environment. Recently, Yu Zheng et al. (2019) changed this statistic by providing an application framework of DT and its case study, but still applied in manufacturing. Therefore, there is a significant amount of research opportunities related to digital twin conception and application in various sectors of the industry. In this research work, we propose to develop a digital twin prototype for organization suitable for in-store operations daily management. Moreover, with the help of the digital twin, we propose to take up the challenge discussed in subsection 2.2, namely, considering multiple store operations decisions, addressing multiple themes and multiple products.

3. RESEARCH APPROACH FOR BUILDING IN-STORE DIGITAL TWIN BUSINESS CASE

For this project we followed a Design Science Research (DSR) approach (Peffers, *et al.*, 2007). According to the scope of this paper, we focused on phases 1-4 of the DSR methodology presented in **Figure 1**.

3.1. Identify Problem/Opportunity From The Field

Phase 1: In today's retail's operations, a vast majority of retailers analyze in-store buying behaviours by making use of the point-of-sale (POS) system data. They gather information like (a) the number of buyers, (b) the number of purchased items by transaction, (c) the products most demanded, etc. Even though, a large amount of data can be captured from this end, retailers have no real visibility on what truly happens in stores. As an example, they have no knowledge on (i) what items are being tried in a fitting room (ii) among these items, which ones are purchased (iii) which items customers interact the most with (iv) in which fitting rooms (v) what is the utilization rates of each fitting room, etc. There is hence a huge unexploited potential for leveraging on products and consumer behaviours to help managers improve their operations and their in-store customer experience.

3.2. Defining the Objective of Our DT Store Prototype

Phase 2: To track in-store products and customers' behaviours, and more specifically in fitting rooms areas, the objective of the research is to build a DT prototype and assess its impacts on operations.

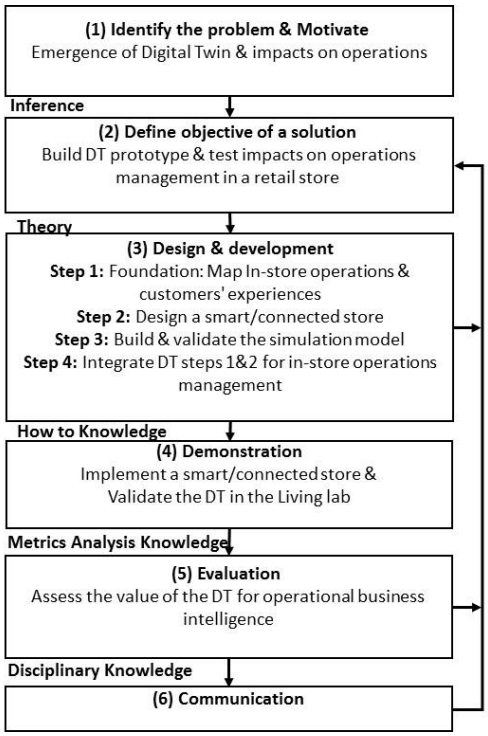


Figure 1: Research approach for building our case

3.3. Design & Development of the DT

Phase 3: The Design & development of the DT is structured in four steps. **The first step** is to map in-store daily operations to formalize customers' behaviours as illustrated in **Figure 2**. Information was gathered from a store in a fashion Canadian retail chain that includes 15 stores throughout Canada.

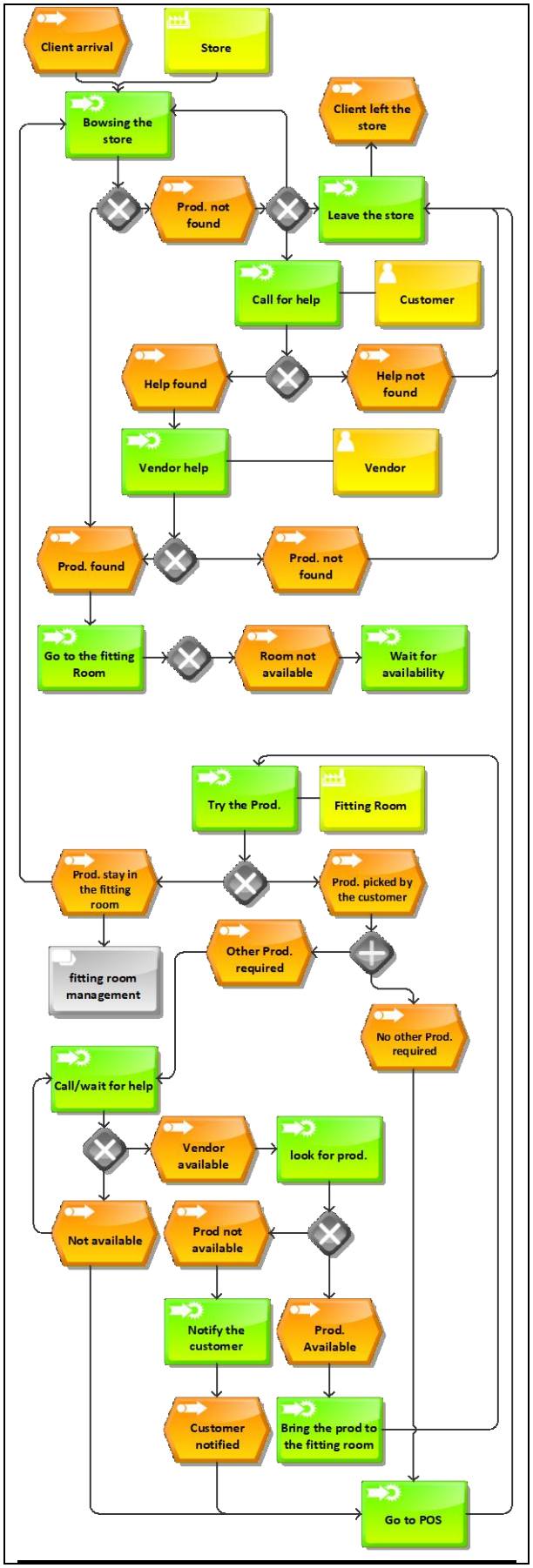


Figure 2: In-store operations map

The second step was to design our smart/connected prototype store. To track in-store customers' behaviours, we first selected suitable RFID/IoT technologies (i) to capture data related to items picked by the customer (i.e., passive UHF RFID technology) and (ii) to capture data related to the identification and tracking of the customers in the store (i.e., Bluetooth Low Energy-BLE). In order to leverage on previous research settings, our prototype was developed at the Montreal-based *IoT lab*, where various RFID and BLE technologies are available and were used for previous research in retail settings (Bendavid et al., 2018).

The third step consisted of building a comprehensive discrete-event simulation model for store daily operations and customer behaviours. The building of the model was done by taking in consideration criteria's such as uncertain customers' arrivals, service times and most interesting, uncertain customers' behaviour in fitting rooms. The simulation model was built using Arena software 15.0 by Rockwell automation. **Figure 3** displays the main view of this model.

Once the physical data capture infrastructure was designed, and the simulation model verified and validated, we therefore had to develop a first comprehensive DT prototype for store operations management by integrating phase 3's steps 2 & 3.
This integration represents **the fourth** and last step of phase 3. It is accomplished by: (i) integrating IoT captured data into a MySQL database through a Java software application (a detailed description of this part can be found in (Bendavid et al., 2018), (ii) perform statistical analysis through a python software application (iii) derive fitted distributions that are read from the Arena Model**.**
Figure 4 displays a sample of captured data analysis. One can see the distributions that are automatically fitted as soon as the python script is launched and then read by the Arena model through an Excel file at the beginning of the simulation.

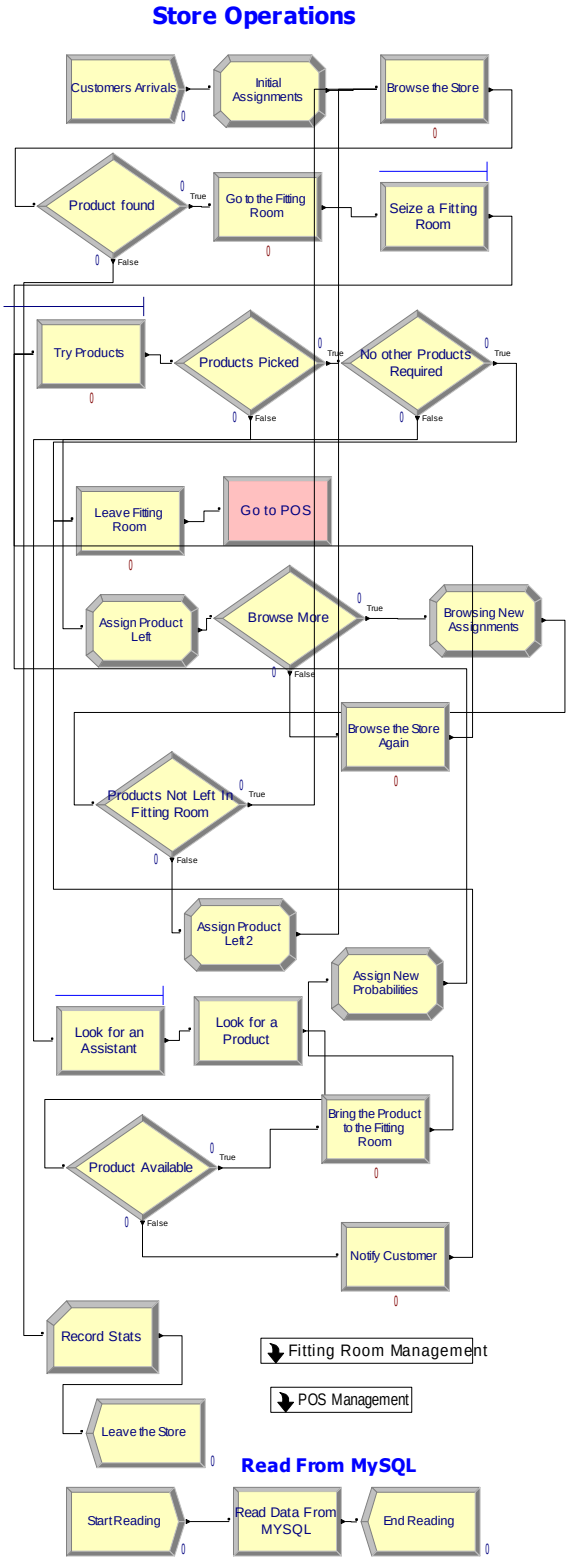


Figure 3: Arena simulation model for store daily operations

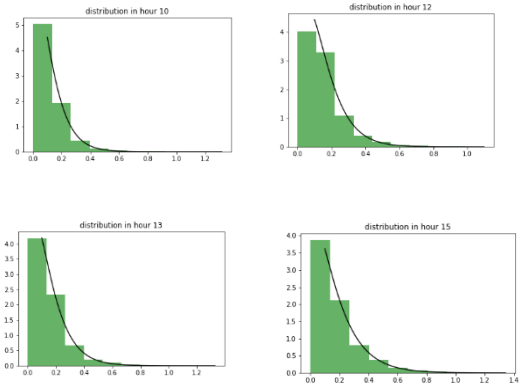


Figure 4: A sample of real time captured data analysis.

3.4. Demonstration

We are presently conducting phase 4 of this research project.

Phase 4: consisted of implementing a smart/connected store prototype at the *GreenUX lab* which is a close representation of a retail store, and validate the feasibility of the DT in this Living Lab. (i.e., the store). Beyond the choice of selecting appropriate RFID/IOT technologies to identify and capture data, specific locations were identified as "reading zones" for real-time items and customers identification.

To identify and capture customer-related data, BLE technologies (solution from *reelyActive*) were selected and implemented in three main locations (i.e., entrance, main store, fitting room). **Figure 5** presents how BLE technologies are used to locate customers within preconfigured specific zones within the store (with a radius of 3-4 meters). Hence, customers arriving at our "store" can be automatically detected using their BLE device (e.g., cell phone). As they move from zone to zone, their data is also captured by three receivers, with few seconds of latency.

Cameras already installed in the store can also be used to gather customers' movements, understand their behaviours and experiences. We plan to integrate them in a further phase of the project. Presently they allow us to capture sequences and review them afterward for analysis purposes.

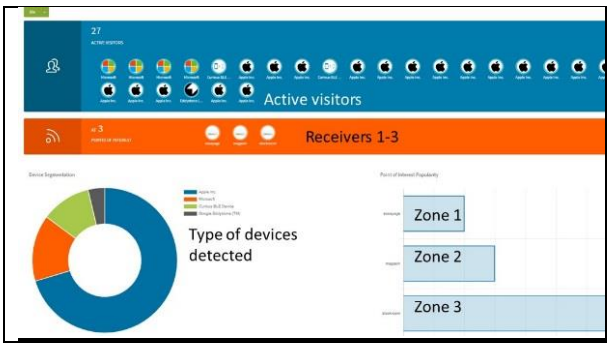


Figure 5: Using BLE to locate customers.

On the other hand, item data capture (identification and tracking) is done using fixed readers/antennas installed on the ceiling of the store. The system acts as a passive RTLS (Real-time location System) used for wide coverage. Technologies from *Impinj (xArray*) were used to demonstrate the feasibility of item tracking in three zones (i.e., back store, main store, fitting room). (**Figure 6.a)**. Items are automatically detected within one-two feet precision, and their movements can be followed in real time, contributing to the operationalization of a DT prototype.

Since the analysis of fitting rooms is important, within the scope of this study, we implemented a dedicated reader/antenna to increase the accuracy of the reads and the precision of location of data captured in this specific choke point. RFID readers detect items that are removed from the shelves and brought to other areas of the store. Once customers enter fitting rooms, RFID enabled *smart mirrors* equipped with RFID recognize items brought to and can display products' information as well as additional information, enabling customers to see other available sizes and colours (**Figure 6.b)**.

If a customer leaves the fitting room area without taking back the tried items, a store employee is informed through the system that an item has been left in the fitting room. This event triggers a request to pick up the items left in the fitting room and put them back on the shelves (**Figure 6.c)**.

Since the data is automatically captured from the daily operations in the store, the analysis of this data is continuously running, therefore generating new probability distributions and feeding the simulation model. By continuously generating updated data, this allows our DT to integrate new distributions fittings. Therefore, once the simulation scenarios are run, appropriate adjustments of operations management rules as well as policies are evaluated by the decision maker in order to improve store operations management performance.

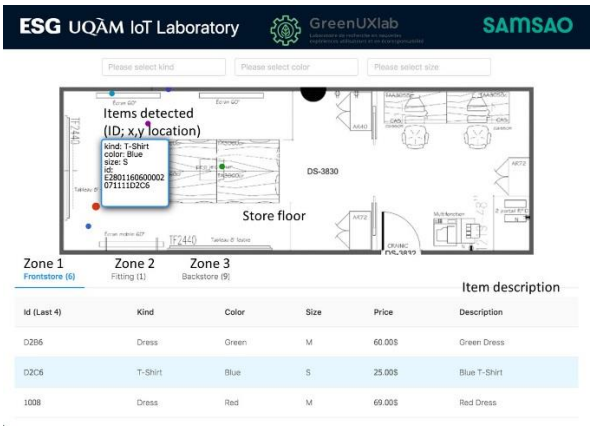


Figure 6. a. Real-time tracking of items in the store

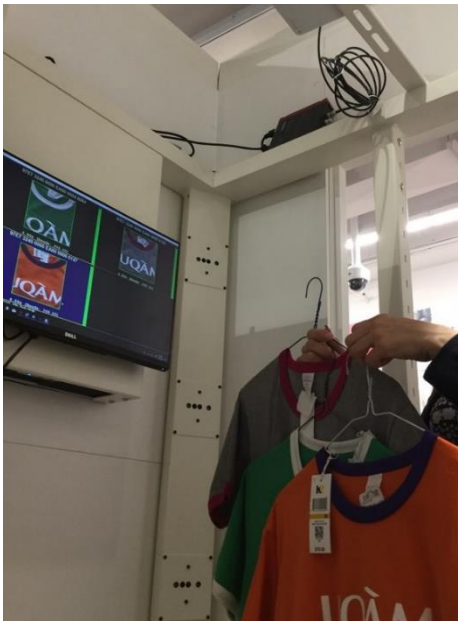


Figure 6. b. Real-time detection of items in the fitting room

Figure 6. c. Real-time visibility of fitting rooms by employees

Figure 6: Real-time visibility in the store

The next phase aiming at assessing the value of the DT for operational business intelligence is then planned with great expected outcomes for real-time decision-making.

4. EXPECTED OUTCOMES

In this research, we expect to show that the building of the proposed DT for the retail store will provide us with a huge amount of information to support real-time decision-making. **Figure 7** summarizes the envisioned work and fittingly proposes a high-level systemic view of the required components to design and build a DT.

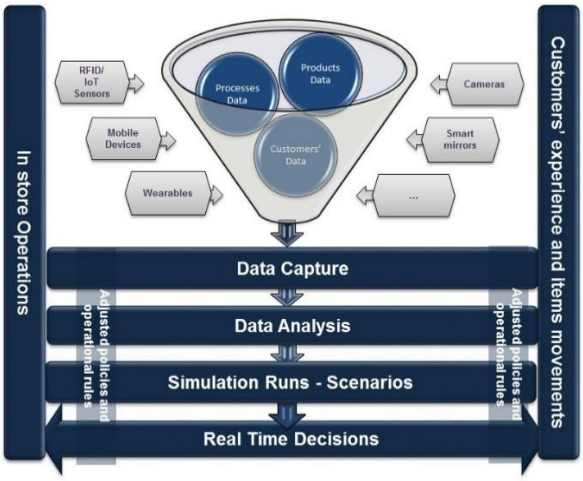


Figure 7: Integrating IoT Technologies to Simulation Modelling – a DT for Store Operations Management.

Since a retail store is a complex system, we decided to focus on the fitting room as a starting point to evaluate realistically the performance of the DT. The choice of focusing on fitting rooms areas is justified by the high potential for improvements. Indeed, as we know:

i. on average, only 1/3 of the customers that browse in the store will use a fitting room. This is explained by the fact that these fitting rooms are not properly located, or due to insufficient staff scheduling to assist customers or simply because they are always full when a customer wants to use them.
ii. Moreover, around 3/4 of customers that use a fitting room will purchase an item.

Therefore, it is obvious that one of the expected impacts of our DT utilization in this case is to make a better usage of these fitting rooms and consequently increase the number of customers that make a purchase. In our future research work, we would like to:

- Provide the real-time number of in-store customers and adjust staffing decisions consequently;
- Gather fitting room information to inform customers about rooms' availability and hence improve fitting room utilization rate;
- Increase the percentage of customers that browse the store and decide to use a fitting room;
- Increase customers' satisfaction by providing personalized support in the fitting room;
- Re-evaluate items on shelves after considering most tried/purchased vs. tried/not purchased items (i.e., items and customer behaviours);
- Identify and relocate forgotten items in fitting rooms.

5. CONCLUSION

The fast development of IoT technologies and their applications have raised a huge research interest in various sectors, among which the retail sector is positioned as an early adopter. While the real potential of such technologies and applications is being progressively revealed, this has given the rise of the DT concept and the idea that any object, process or system could be replicated by a digital representation with the objective to understand it, respond to changes and improves operations – i.e., add value (Gartner, 2019b).

Although the idea is interesting, putting it into practice by moving from the concept to the prototype of the DT is much more complex. Hence, in this research, besides presenting a digital twin prototype for in-store retail operations daily management, the main contribution is the operationalization of the main steps, and the main technologies required to build such as DT prototype. The work was conducted in the IoT lab and the GreenUX lab at ESG-UQÀM where some of the latest IoT technologies can be found.

The proposed DT in a retail environment is a first of a kind and constitutes a powerful tool to help organizations' managers to adjust their operational decisions on a regular basis. For instance, we explained how the implementation of a DT in a store can significantly improve fitting rooms' utilization, increase customer satisfaction, provide with a better staff scheduling and customer handling in fitting rooms.

The next step of this research is to run the prototype in a real in-store environment – i.e., in one selected store, to evaluate its real potential and most likely adjust some of the integrated applications.

Finally, this work highlighted some technical difficulties:

i. Firstly, from a data quality perspective, RFID and BLE data need to be extensively processed to provide accurate information and automatically feed the simulation model.
ii. Secondly, from a simulation modelling perspective, the automatization of fitting distribution still needs to be slightly improved, mainly regarding the choice of intervals to be considered in an operational day.

In our next generation of DT prototype for retail, we plan to (a) keep BLE as a preferred technology to identify customers and explore the opportunities of face recognition technologies to improve the accuracy of identification (b) use the latest passive RTLS technologies proposed by *RF Controls* (CS-445 antennas) to have a 3D visualization of items' location and movement. (c) run several simulation scenarios related to different retail stores in order to validate the robustness of our DT prototype.

REFERENCES

Bain & Company, 2018. Unlocking Opportunities in the Internet of Things. Survey study. Available at: https://www.bain.com/contentassets/5aa3a678438846289af59f62e62a3456/bain_brief_unlocking_opportunities_in_the_internet_of_things.pdf

Bendavid Y., S. Rostampour et J. Dengen (2018). Combining Internet of Things (IoT) technologies to understand product and customer in-store behaviour. 23rd Association Information & Management conference, 16-18 may 2018, Montréal, Québec, https://aim2018.sciencesconf.org

Bendavid Y., 2012. Positioning RFID Technologies in the Enterprise Information Systems Portfolio: a Case in Supply Chain Management. International Journal of Automated Identification Technology, Vol. 4, No 1, pp. 11-24.

Deirdre B., 2018. Amazon's automated grocery store will launch Monday after a year of false starts. Available at: https://www.cnbc.com/2018/01/21/amazon-go-automated-grocery-store-is-poised-to-launch.html

Gartner, 2019a. Gartner Identifies Top 10 Strategic IoT Technologies and Trends. Available online: https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019/ (accessed on 03 February 2019).

Gartner, 2019b. Gartner Survey Reveals Digital Twins Are Entering Mainstream Use. Available at: https://www.gartner.com/en/newsroom/press-releases/2019-02-20-gartner-survey-reveals-digital-twins-are-entering-mai (accessed on 03 March 2019).

Grieves M., and John V., 2017. « Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems ». In *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, edited by Franz-Josef Kahlen, Shannon Flumerfelt, et Anabela Alves, 85-113. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-38756-7_4.

Huiru W., Shen J., Feng J., Feng H., Zhou Z., and Mu W., 2018. « An agent-based modelling and simulation of consumers' purchase behaviour for wine consumption ». *IFAC-PapersOnLine*, 6th IFAC Conference on Bio-Robotics 2018, 51, nº 17: 843-48. https://doi.org/10.1016/j.ifacol.2018.08.089.

HIS M., 2017. Number of Connected IoT Devices Will Surge to 125 Billion by 2030. Available at: https://technology.ihs.com/596542/number-of-connected-iot-devices-will-surge-to-125-billionby-2030-ihs-markit-says

Kritzinger W., Karner M., Traar G., Henjes J., and Sinh W., 2018. « Digital Twin in Manufacturing: A Categorical Literature Review and Classification ». *IFAC-PapersOnLine* 51, nº 11: 1016-22. https://doi.org/10.1016/j.ifacol.2018.08.474.

Lieselot V., and Macharis C., 2011. « An Agent-Based Model of Consumer Mobility in a Retail Environment ». *Procedia - Social and Behavioral Sciences* 20: 186-96. https://doi.org/10.1016/j.sbspro.2011.08.024.

Miwa K., et Soemon T., (2008). « Simulation modelling and analysis for in-store merchandizing of retail stores with enhanced information technology ». *2008 Winter Simulation Conference*, 1702-10. https://doi.org/10.1109/WSC.2008.4736256.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24 nº 3, 45-77.

Shandong M., Robb D.J., and DeHoratius N., 2018 « Retail store operations: Literature review and research directions ». *European Journal of Operational Research* 265, nº 2 (mach 1 2018): 399-422. https://doi.org/10.1016/j.ejor.2017.07.003.

Roberti M. (2018). 100 Million Reasons Why Retailers Should Use RFID. RFID Journal. Available at: https://www.rfidjournal.com/articles/view?17974

RFID Live https://www.rfidjournal.com

Thiesse F, Al-Kassab J., and Fleisch E., 2009. « Understanding the value of integrated RFID systems: a case study from apparel retail ». *European Journal of Information Systems* 18, nº 6 (1 December 2009): 592-614. https://doi.org/10.1057/ejis.2009.33.

Utomo, D.S., Onggo B. S., and Eldridge S. 2018. « Applications of Agent-Based Modelling and Simulation in the Agri-Food Supply Chains ». *European Journal of Operational Research* 269, nº 3 (September 2018): 794-805. https://doi.org/10.1016/j.ejor.2017.10.041.

Zaino, J. (2016). Riding the Tails of Apparel Retailers. *RFID Journal*.

Zebra (2017). Reiventing Retail, 2017 Retail Vision Study. Taking retail automation and personalization to new hights. Zebra retail study, Available at: https://www.zebra.com/content/dam/zebra_new_ia/en-us/solutions-verticals/vertical-solutions/retail/vision-study/retail-vision-study-2017-en-gb.pdf

AUTHORS BIOGRAPHY

Dr. Yasmina Maïzi is an associate professor of operations management in the École des Sciences de la Gestion (ESG) - School of Management at the Université du Québec à Montréal (UQAM). She holds a M.Sc. and Ph.D. degrees in industrial engineering from Conservatoire National des Arts et Métiers (France). She has an extensive professional experience in simulation within the industry. She is an associate member of the Internet of Things (IoT) lab and CRI2GS lab at ESG-UQAM. Her research area focuses on modelling complex systems using simulation. Recently, she added a new research interest, the integration of Internet of Things (IoT) for health care operation management.

Dr. Ygal Bendavid is the director of the IoT Lab. and a professor in the Department of Management and Technology/School of Management (ESG) at the Université du Québec à Montréal (UQAM). Dr. Bendavid is a founder of RFID Academia. He holds M.Sc. and Ph.D. degrees in industrial engineering from the École Polytechnique de Montréal. As a specialist in RFID for supply chain management, he is a frequent presenter at RFID Journal LIVE! Conferences.

Dr. Janosch Ortmann is an associate professor in the École des Sciences de la Gestion (ESG) - School of Management at the Université du Québec à Montréal (UQAM). His main research interest lies in studying random particle systems and polymer models, particularly in the framework of stochastic integrable systems and the KPZ universality theory. The study of these models turns out to have interesting connections with various areas of mathematics, including algebra, queueing theory and combinatorics (via geometric analogues of the Robinson-Schensted-Knuth correspondence).

SIMULATED OPERATING CONCEPTS FOR WHOLESALE INVENTORY OPTIMIZATION AT NAVAL SUPPLY SYSTEMS COMMAND

Sean Teter[(a)], Emily Craparo[(b)], Javier Salmerón[(c)]

[(a),(b),(c)]Operations Research Department, Naval Postgraduate School, Monterey, CA, U.S.A.

[(a)]sean.m.teter.mil@mail.mil, [(b)]emcrapar@nps.edu, [(c)]jsalmero@nps.edu

ABSTRACT

Naval Supply Systems Command Weapon Systems Support (NAVSUP WSS) serves as the Navy's inventory control point, managing approximately 375,000 line items. Constrained by funding, NAVSUP WSS uses the Wholesale Inventory Optimization Model (WIOM) to maximize customer service. Since demand distributions for different parts change over time, NAVSUP WSS reruns WIOM quarterly. However, large changes to the solution create an administrative burden. To deal with this problem, referred to as churn, WIOM has a persistence parameter that can discourage change from one run to the next, but it is inherently at odds with customer service performance. This research develops the Comparative Optimized Results Simulation to explore the system's performance under different persistence settings and periodicities of running WIOM. The research finds that periodicities greater than quarterly significantly degrade customer service, and increasing the persistence parameter dramatically improves churn while only marginally degrading customer service.

Keywords: inventory management; discrete event simulation; wholesale inventory optimization model

1. INTRODUCTION

Naval Supply Systems Command, Weapon Systems Support (NAVSUP WSS) serves as the main inventory control point for the Navy. The command manages over 375,000 unique line items (NAVSUP 2018) used in the repair of ships, submarines, Navy and Marine Corps aircraft, and associated weapons systems. The effective management of this supply chain is essential in maintaining readiness of the fleet to operate and conduct combat operations around the world.

Like any organization, NAVSUP WSS has a limited set of resources with which to conduct its operations. The biggest constraint is financial. Given limited budgetary means, NAVSUP WSS strives to maximize support to the warfighter. The predominant metric used to measure customer support is fill rate. When NAVSUP WSS receives a requisition, one of two things can happen. Either the requisition is filled immediately with stock on hand, or the requisition is backordered. The fill rate metric shows the relationship between the number of requisitions filled immediately on receipt and the number of requisitions that are backordered. Fill rate is defined mathematically as follows:

Fill rate = Requisitions Filled / Requisitions Received.

For example, if 50 requisitions were received in a given period, and 43 of them were filled immediately and 7 were backordered, then a fill rate of 86% was achieved for this period. The above calculation can be applied to a specific item or to a group of items. When it is applied to a group of items, it can be done in one of two ways. First, the fill rate can be calculated as an average of all the individual item fill rates. Or, the fill rate can be calculated with the above equation without regard to what the particular item is. This is also called demand weighting, because it is equivalent to a weighted average of item fill rates, weighted according to the demands of the individual items. In this study, we use demand weighted fill rate unless specifically noted otherwise.

In the past NAVSUP WSS used commercially-developed optimization software to maximize their achieved fill rate given their budget constraints. Developed by MCA Solutions, the Service Planning and Optimization (SPO) was effective but had shortcomings. First, it was a "black box" to the users at NAVSUP WSS, who did not have access to the models and algorithms SPO used to develop its solutions. SPO did not have the ability to accept budget as a constraint. Therefore, NAVSUP WSS had to run SPO iteratively, adjusting a fill rate constraint until a satisfactory budget figure was reached. Additionally, SPO was expensive, costing around $800,000 per year in licensing fees. In order to replace SPO with a better-functioning optimization tool at reduced cost, Naval Postgraduate School faculty developed the Wholesale Inventory Optimization Model (WIOM) (Salmeron and Craparo 2017). WIOM is a mixed-integer linear program designed to maximize a function closely related to fill rate, for the wholesale inventory managed by NAVSUP WSS. Roth (2016) used simulation modeling to conclude that WIOM 3.51 was in fact superior to SPO in maximizing fill rates. NAVSUP WSS sunset SPO

and began using WIOM in April of 2017. While WIOM performs well compared to SPO, NAVSUP WSS identified further features they would like to be incorporated into WIOM. First, WIOM 3.51 did not use demand weighting. Instead, it had two settings that could be used. First, WIOM could treat each National Item Identification Number (NIIN) equally. This is not desirable because it ignores the relative importance of NIINs with high demand. Alternatively, WIOM could give preferential treatment to NIINs that were assigned to specific groups called level setting strategy indicators (LSSIs). By assigning high-demand NIINs to a certain LSSI and then assigning that LSSI a high weight, NAVSUP WSS could mitigate the demand weighting issue. Additionally, NAVSUP WSS could use a series of business rules to create low-demand cutoff points, choosing to leave very low demand NIINs out of the optimization altogether. In order to address this concern, WIOM was revised to use demand weighting, and incorporated this change into the WIOM 4.1 release.

NAVSUP WSS has an additional concern with WIOM (and SPO before it): churn. Churn is the change between solutions from one model run to the next. NAVSUP WSS runs the optimization model once every quarter. In the three months between model runs, the number of requisitions received changes the demand parameters that feed into WIOM. Subsequently, the optimization problems are quite different and considerably differing solutions are possible. Indeed, if multiple optimal (or near-optimal) solutions exist, churn may occur even in the absence of changes to the input data. Churn creates an administrative burden in contracting and can reduce senior leadership's confidence in optimization efforts. To deal with the problem, Salmeron and Craparo (2017) included a term in WIOM's objective function that calculates a churn penalty. This term contains two penalty parameters. One is indexed by NIIN, allowing the user to adjust the relative importance of each NIIN within the churn term. The other is a global persistence parameter that reflects the overall importance of the churn term. This study focuses on the global persistence parameter; for simplicity we use the term "persistence parameter" hereafter. The persistence parameter rewards a solution for maintaining legacy values from one model run to the next. The parameter is not an on/off switch; rather, it is a continuous parameter that can be set from zero to an arbitrarily large number. At zero, the persistence parameter is "off." As the parameter increases, the model more strongly prefers to retain incumbent solutions. Additionally, there is an inherent tradeoff between churn reduction and achieved fill rate. The higher the persistence parameter, the less important fill rate becomes in the objective function. This paper explores this tradeoff via discrete event simulation following work by Teter (2018).

The remainder of the paper is organized as follows: Section 2 presents a literature review of related inventory models, to include optimization and simulation methods. Section 3 describes the methodology goals, data and an introduction to the simulation metamodel. Section 4 explores the effects of periodicity and persistence settings on fill rate. Finally, our conclusions are presented in Section 5.

2. LITERATURE REVIEW

2.1. Inventory Management

Wholesale inventory management is concerned with finding strategies to meet demand requirements from customers at an acceptable service level and an acceptable cost level. Many different models have been proposed, but the two we will discuss are the order-point, order-quantity (s,Q) model and the classic inventory model.

Order-point, order-quantity models are discussed in Silver et al. (1998). In an (s,Q) system, two parameters are used to make decisions on stock replenishment. The first is the reorder point, s. As an item's stock level decreases, a reorder is triggered once the item's inventory position decreases to the level of the reorder point. Inventory position is defined as the quantity on hand plus the quantity on order minus the quantity in a backordered status (i.e., owed to customers). The second parameter is the order quantity Q. This is the quantity of material ordered every time there is a reorder. When a reorder is placed, the time it takes for this order to arrive is known as the lead time. A key feature of an (s,Q) system is that each reorder is triggered by a low inventory position, not low inventory on hand. This prevents the system from placing extra orders when there is already an order due-in that will replenish stock sufficiently. Silver et al. provide an analogy: "A good example of ordering on the basis of inventory position is the way a person takes aspirin to relieve a headache. After taking two aspirin, it is not necessary to take two more every five minutes until the headache goes away. Rather, it is understood that the relief is 'on order'— aspirin operates with a delay" (Silver et al. 1998).

WIOM uses the (s,Q) system to model NAVSUP WSS's wholesale inventory. However, NAVSUP WSS only determines reorder points. The quantity of the reorders is decided by Navy Enterprise Resource Planning (ERP), and is treated as input by NAVSUP WSS, who then strives to maximize effectiveness by deciding on appropriate reorder points.

A special case of the (s,Q) system is the classical inventory model discussed in Tersine (1994). The classical inventory model uses an (s,Q) system but with a very rigid set of assumptions. Among other things that are not relevant to our purposes, the classical inventory model assumes the following: deterministic and constant demand; constant deterministic lead time; reorders arrive as a whole lot of size Q; and backorders are not allowed, since constant demand and lead time allow backorders to be avoided with certainty. The resulting system creates a characteristic saw-tooth

pattern. This inventory model is used primarily as a means to estimate an order quantity that minimizes cost, known as the economic order quantity (Silver et al. 149-197). Since NAVSUP WSS treats the order quantity as a given input from ERP, we are not concerned with that aspect of the model. However, the model has some unique qualities that we will use when establishing initial conditions for our simulation. Specifically, a result of the model is that the average amount of inventory on hand is equal to Q/2. Furthermore, the inventory on hand at any given time is distributed uniformly from zero to Q.

Bachman et al. (2016) combine discrete event simulation with optimization for inventory models to manage items with demand that is either infrequent or highly variable. Simulation-based optimization is also proposed by Köchel and Nieländer (2005) to define optimal inventory policies in multi-echelon systems.

2.2. Discrete Event Simulation

Discrete event simulation is addressed in detail in Law (2015). Discrete event simulations are those that advance time from one discrete event to the next. These events may change the state of the system being represented, and the system cannot change during the time between events. Law presents several concepts to understand such a simulation: System state; Simulation Clock; Event List; and, Initialization Routine.

We develop a simulation using this next-event time advance principle. Events in the system are arranged in time in an event list. The simulated time moves forward from one event to the next according to the events' arrangement in time. The current event is evaluated, state changes to the system are made as necessary, and the simulation moves to the next event in time while the simulation clock is updated.

2.3. Previous WIOM Simulation Study

Roth (2016) conducted a comparative simulation study between three different optimization methods: simple calculation (a heuristic), SPO, and WIOM. Using a discrete event simulation and testing across five types of material, Roth concluded that WIOM was the best performing of these three alternatives. However, Roth's simulation relies on several strong assumptions:

- NIIN demand probability distributions are known and unchanging through time;
- NIIN demands arrive in quantities of one only;
- Demands are uncorrelated between NIINs.

In addition to these assumptions, the simulation models a lengthy warm-up period of 400,000 days to reach steady state. Due to these assumptions and warm-up period, Roth's simulation would be ineffective to try to model short-term performance of the system with frequent WIOM runs and changes in estimated demand distributions every quarter.

3. METHODOLOGY

3.1. Goals

The work creates a discrete event simulation that uses historical requisitions as input and requires no warm-up period. We call this simulation the Comparative Optimized Results Simulation (CORS). By using historical data and not requiring a warm-up period, CORS allows for multiple runs of WIOM during the test period. We conduct a series of experiments using the simulation and analyzes the output in order to gain insight into: (a) the relative tradeoff between churn and fill rate using differing settings for the persistence parameter, and (b) the effect of WIOM periodicity on fill rate.

In practice, NAVSUP WSS has historically used a set of business rules to help it overcome limitations in SPO. These business rules include mandating minimum and maximum reorder points for some NIINs, which restrict the range of solutions that SPO can use.

Additionally, WIOM accounts for churn by use of the persistence parameter and accounts for low demand items by using demand weighting. Therefore, no additional business rules will be used in this study. While exploring differing concepts of operations for NAVSUP WSS, we do not explore all possible periodicities. Running WIOM and implementing its solution is administratively burdensome, and organizationally NAVSUP WSS wants to maintain a normal battle rhythm (Ellis et al. 2017). For this reason, we assume that WIOM can only be run quarterly, semiannually, or annually.

Our work is limited to non-nuclear consumable material. Modeling repairable material is more complex and not addressed in this study.

CORS does not attempt to model all aspects of inventory management. Therefore, while the model delivers insight into performance, it only does so relatively. We are only comparing between simulations and claiming which operating condition performed better.

Using deterministic demand gives great flexibility to explore the effects of different concepts of operations that a long term steady state simulation does not. However, by using deterministic demand we are essentially restricted to one data point and a trace simulation. Thus, our conclusions are inherently limited. We can say that one concept of operations performed better than another in the simulation, but only for the given set of demands.

3.2. Data

The set of data provided includes historical requisitions. The data are a record of all demands that NAVSUP WSS received during 2013 through 2017. The next set of data provided consists of historical candidates files. These files contain information for all the NIINs that were input into SPO for each quarter. Also provided are

historical wholesale data files (Ellis 2017). These files have the majority of the data elements needed to run in WIOM, but they do not include the budget category, which is necessary to classify a NIIN as a particular type of material. After excluding repairable and nuclear items, and items with inconsistent requisition data, our dataset contains approximately 3,800 maritime NIINs and over 100,000 requisitions.

3.3. Metamodel

As input, CORS requires requisition data and WIOM outputs for each quarter of the time period being tested. To obtain the necessary WIOM outputs, we start by running WIOM for the first quarter in the time period. This run uses the candidates file for the first time period developed above, the budget figure, and the persistence parameter we are exploring. The second WIOM run for the next sequential quarter requires all the same input data plus the first WIOM solution, as it uses this information to enforce persistence. The third WIOM run requires the second WIOM solution, etc. After repeating the process for all available quarters we have a library of WIOM output. This WIOM output contains both the optimal reorder points (ROPs) and the NIIN characteristics CORS requires; namely, each NIIN's lead time (LT) and order quantity (Q). This library of 18 WIOM outputs is fed into CORS, along with the requisition data. CORS then performs its simulation and outputs system performance in terms of fill rate. Figure 1 illustrates the process.

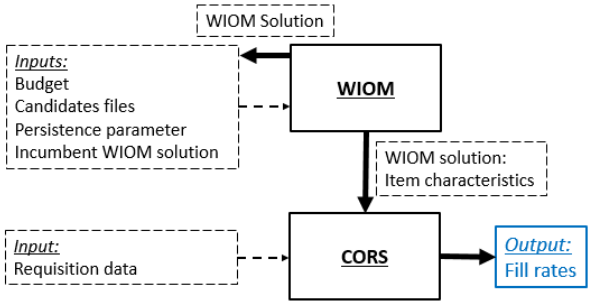


Figure 1: Metamodel Relationships

Simulating one NIIN at a time, CORS maintains an event queue with events aligned in time to trigger demand arrivals, order arrivals, and parameter changes due to new WIOM input. Each event triggers a particular logic sequence that examines the current state of the system and makes appropriate changes to the system and event queue. The simulation tracks the following system characteristics (state variables):

Order Quantity (Q)
Reorder Point (ROP)
Lead Time (LT)
Quantity On-Hand (Q_O/H)
Inventory Position (IP)
Time (t)

CORS output information that can be used to calculate fill rate in a variety of ways. First, the model outputs the overall fill rate for each NIIN for the entire simulation. Next the model outputs aggregate data for all NIINs that can be used to calculate the fill rates for a number of time frames. For each month, the total number of requisitions filled (across all NIINs) and the total number of requisitions received are both recorded. With these data, aggregate demand weighted fill rates can be calculated for any periodicity that is a multiple of months (i.e., quarterly, annually, etc.). Finally, the model outputs the average length all backordered requisitions stayed in a backorder status.

We implement the simulation logic in the R programming language (R Core Team 2016) to run CORS. A formal representation of the model, as well as additional logic to record statistics of system performance, appear in Teter (2018).

4. ANALYSIS

4.1. Operating Concepts Explored

We have two items we wish to explore: run periodicity and the persistence parameter. We only consider periodicities of quarterly, semi-annually, and annually. For the persistence parameter, we choose to use parameters that roughly correlate to none (0.0), low (0.1), medium (1.0), and high (5.0).

Additionally, we explore the possibility of a hybrid approach, where WIOM is run every quarter, but with different persistence parameters. In this hybrid idea, persistence is turned off in one model run per year in order for the solution to "reset" and adapt to any drift that has occurred in the demand distributions. The other three quarters the persistence parameter is set at the low, medium, or high level. Thus, the total number of concepts tested in our experiments is 15.

4.2. Effect of Periodicity in Fill Rate

Table 1 shows the list of settings for the 15 designs and the resulting "overall" (average over all NIINs) fill rates achieved under each design, as simulated in CORS. Note that WIOM does not directly maximize fill rate; rather, it minimizes a nonlinear penalty associated under-achieving fill rate goals. Nonetheless, overall fill rate provides a simple aggregate figure of merit by which to judge system performance. The results indicate clear differences in fill rate achievement between periodicities, with the greatest difference between Annual and Quarterly designs (close to 10%). Differences between persistence levels (within the same periodicity) are almost negligible, noting that higher levels have the desired effect of significantly reducing churn (to be shown later).

An interesting question is when the overall simulated fill rates start to diverge under different periodicities. For example, in the case of designs 4 and 9, the overall difference of about 10% does not occur until month 20 (see Figure 2). In our design comparisons, divergence takes even longer to take place. Based on what we see here, we expect at least two quarters before any impact of a WIOM implementation is experienced. This makes

intuitive sense as the average lead time across the NIINs tested is a little more than a year.

Table 1: Overall Fill Rates by Design

Design	Periodicity	Persistence	Overall fill rate
1	Annual	0.0	51.77%
2		0.1	51.88%
3		1.0	51.85%
4		5.0	51.74%
5	Semi-annual	0.0	58.34%
6		0.1	58.08%
7		1.0	58.45%
8		5.0	58.01%
9	Quarterly	0.0	61.57%
10		0.1	61.16%
11		1.0	61.43%
12		5.0	60.90%
13	Annual, then Quarterly	0.0, then 0.1	61.53%
14		0.0, then 1.0	61.46%
15		0.0, then 5.0	61.32%

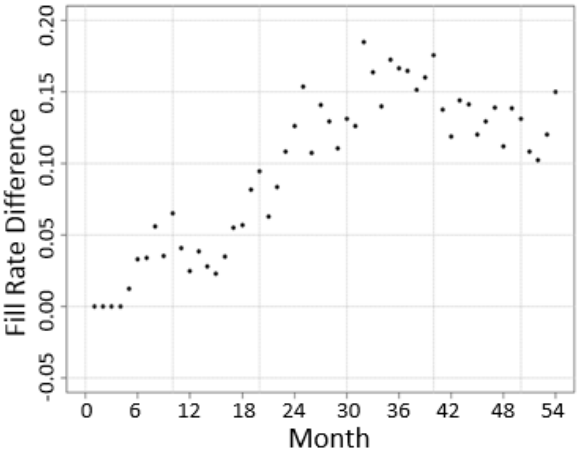


Figure 2: Monthly Differences in Simulated Overall Fill Rates between Designs 4 and 9

4.3. Effect of Persistence Level

The persistence parameter does not directly set a certain level of churn. Rather, it is a change in the weighting of the objective function for the WIOM optimization model. So, we must first analyze the effect of the persistence parameter on churn, and then analyze the effect on fill rate performance.

It is important to note here that we are comparing churn, which is calculated in WIOM, against simulated fill rate performance, which is not calculated exactly: The optimization uses a closed-form approximation amenable to calculations. The simulation calculates it more accurately, as shown by Roth (2016). The purpose here is not to compare the relative values of the two terms in WIOM's objective function. Rather, our goal is to compare churn against simulated system performance. Having shown that annual and semi-annual concepts perform poorly, we restrict the persistence analysis to quarterly periodicities only.

4.3.1. Persistence Level on Churn

The persistence parameter in WIOM enforces persistence by applying a penalty when the safety stock of a NIIN differs from the previous safety stock level. The safety stock is the expected quantity on hand when an order arrives. The penalty (*Total Churn*) can be defined by the following expression:

$$Total\ Churn := \sum_{i \in I} \frac{\left| \hat{S}_i^0 - S_i \right|}{\hat{S}_i^0 + 1}$$

where I is the set of all NIINs, $\hat{S}_i^0$ is the safety stock of NIIN i in the incumbent solution, and is S_i is the new (optimized) safety stock.

We note the churn penalty for a single item i is proportional to the relative magnitude of the change. For example, a change of solution from 9 to 10 incurs a penalty of 0.1, while a change from 9 to 19 incurs a penalty of 1.0. To compare the churn across our quarterly designs, we compute *Total Churn* for every quarter, and take the mean value across the simulation time period for each concept of operation design. The results of these calculations are shown in Table 2.

Table 2: Total Churn by Quarterly Design

Design	Persistence	*Total Churn*
9	0.0	5,673
10	0.1	600
11	1.0	154
12	5.0	50

The designs using a constant persistence parameter every quarter show a clear reduction in churn with increasing persistence parameter. The highest persistence parameter tested has, on average, less than 1% the churn present with the parameter set to 0.0.

Next, we look at hybrid concept designs, 13, 14, and 15, separating average churn rates when the persistence parameter is equal to zero and when it is not (Table 3).

Table 3: Total Churn for Hybrid Designs

Design	Persistence	Average with zero persistence	Average with positive persistence	Overall
13	0.0, then 0.1	8,816	572	2,512
14	0.0, then 1.0	9,391	155	2,328
15	0.0, then 5.0	8,925	49	2,137

We make two observations. First, in the quarters when persistence above 0.0 is used, average churn for designs 13, 14, and 15 is very similar to average churn for designs 10, 11, and 12, respectively (see Table 2). The next observation is that the large overall average churn for the hybrid designs comes from the annual runs with persistence set to 0.0. In these designs churn is very high during the annual "reset" of the WIOM solution but effectively reduced during other quarters.

The above analysis shows that the persistence parameter reduces churn. However, this definition of churn is abstract and mathematical, and there is no immediate understanding of what its values mean to the system. An alternate way to express churn that is more intuitive is to define it as the proportion of NIINs that had any change in safety stock. While WIOM does not use this definition (nor does it pursue such a goal in the objective function), we expect this measurement to decrease in concert with WIOM's definition of churn, and we wish to know if it does not. Using this alternate definition of churn as a proportion, we calculate the average across the simulation period for the different designs in Table 4. As expected, increasing the persistence parameter reduces the proportion of NIINs that have a change in safety stock. However, the reduction is less dramatic than that reflected in the churn formula. The churn formula calculated churn at persistence parameter level 5.0 as less than 1% of the churn at persistence parameter 0.0. Using this alternate definition, the fraction of items with any churn is reduced by about 75% when the persistence parameter is increased from 0.0 to 5.0.

Table 4: Churn Redefined as % of Items with Change

Design	Persistence	Average items with churn
9	0.0	39.99%
10	0.1	30.72%
11	1.0	16.12%
12	5.0	10.35%
13	0.0, then 0.1	36.73%
14	0.0, then 1.0	27.34%
15	0.0, then 5.0	23.10%

Table 5: Churn as Change in Absolute Dollar Value

Design	Persistence	Average churn ($ million)
9	0.0	6.29
10	0.1	4.68
11	1.0	2.83
12	5.0	1.95
13	0.0, then 0.1	5.49
14	0.0, then 1.0	4.55
15	0.0, then 5.0	4.08

A third way to define churn is by dollar value. For any given NIIN, we can define a change in the stock cost as the absolute value of the change in the solution times the unit cost of that NIIN. This dollar value is sometimes referred to as "execution cost." As before, this is a value of interest to the analyst but not one that WIOM pursues by design. Tables 5 shows the results, and we observe similar behavior as in the two previous definitions for churn.

4.3.2. Churn versus Fill Rate

One of the fundamental questions of this study is the trade-off between churn and fill rate performance. For this analysis we use WIOM's original calculation of churn. We start by looking at the relationship between churn value and fill rate for our seven quarterly designs. A graph of these points is presented in Figure 3. However, it is important to note that we are graphing the simulated fill rates achieved over the time period. We are not attempting to find the Pareto curve of efficient solutions, which would be applicable to the two components of the objective value calculated by WIOM. Rather, we are trying to get an idea of the trade-off of between fill rate performance and churn achieved in a production-type environment.

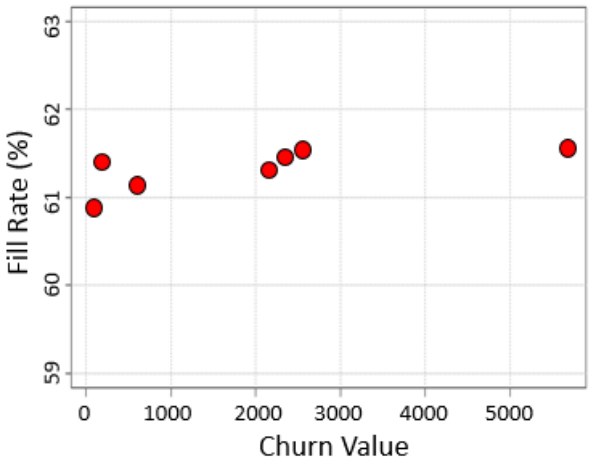


Figure 3: Fill Rate by Churn Value

It appears that there is a very slight increase (improvement) in fill rate associated with an increase (degradation) in churn, which is what we expect. But, we have few data points and the increase is very slight. Reductions (improvement) in churn are very "cheap" in terms of fill rate for these levels of persistence parameter for this set of historical demand.

4.4. Additional Persistence Exploration

Based on the results of the quarterly concepts from our original experimental design, we observe only a small trade-off relationship between churn value and simulated fill rate. However, we know that at some level a larger trade-off exists. The annual and semi-

annual designs effectively have churn-free solutions in the quarters that WIOM is not run. These designs have clear degradation in fill rate compared to the quarterly designs. Therefore, there must be some threshold of churn improvement that causes greater levels of simulated fill rate degradation. However, the persistence parameters we explored did not create churn reduction that crossed that threshold. We therefore conduct a new experiment with higher settings of the persistence parameter to find this threshold and find a steeper trade-off between churn and fill rate.

We add three new concepts of operation to our experiment. We use quarterly runs with the persistence parameter set at 10, 100, and 1000. For this analysis we exclude the hybrid designs. Our results show that increasing the persistence parameter above 5.0 only marginally decreases churn, and increasing it over 10.0 has a negligible effect. This is shown by a clear "knee" in the curve shown in Figure 4.

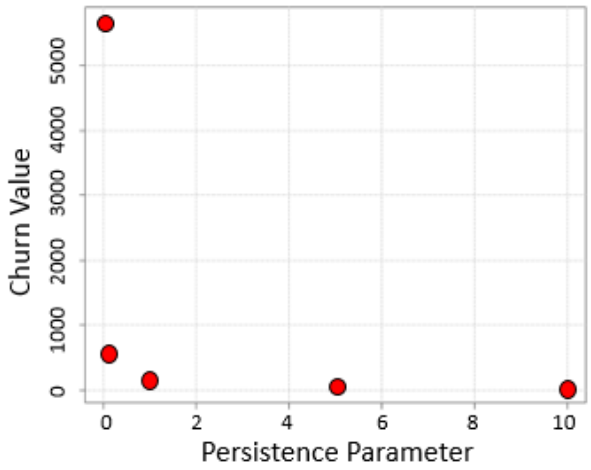


Figure 4: Extended Testing of Churn

As the increase in persistence parameter has little effect on churn, it also has little effect on fill rate performance, which remains fairly constant around 60%.

5. CONCLUSIONS

This work develops a simulation model, CORS, in order to explore the effects of different concepts of operation for wholesale inventory optimization. These concepts of operation vary in terms of the periodicity that WIOM is executed and the persistence parameter used. We explore a variety of different concepts of operations using CORS and we measure system performance for each design in terms of simulated fill rate and churn.

Through the course of this research we have gained several key insights into NAVSUP WSS's wholesale inventory system. The first insight is that it takes time for different implementation concepts to differentiate in terms of fill rate. Even very clearly different solutions take at least six months to produce different fill rates. It takes even longer for the magnitude of the difference to become clear. This insight is important because it reminds us to be cautious in judging the performance of the system in the short term.

Our next key insight into the system is that WIOM solutions have a short shelf life. The system changes sufficiently over time that there are clear degradation to fill rate performance for semi-annual designs and dramatic degradation for annual designs. While different solutions take time to diverge, it is important for the optimization model to be able to adjust to changes in the underlying demand structure quickly. We see no reason to recommend a change to the quarterly periodicity that NAVSUP WSS currently uses.

Perhaps our most important finding is that, for the historical demand considered, churn can be drastically reduced without sacrificing system performance in terms of fill rate. By implementing the use of the persistence parameter, NAVSUP WSS can gain significant improvement in churn, which reduces administrative burden in contracting and improves the ability to explain WIOM results to senior leadership. All this improvement can be gained without sacrificing fill rate performance and support to the fleet.

Our final finding is somewhat unexpected: It appears that WIOM has a limit to how far it can enforce persistence. Beyond a certain point, increasing the persistence parameter has no practical effect on churn. Even increasing the persistence parameter several orders of magnitude has virtually no effect on churn. This may be due to side constraints on WIOM reorder points preventing them to match legacy values. It could also be due to WIOM's budget constraint: If the incumbent solution is too costly for the current budget, the lowest feasible value of churn will be strictly positive.

The results found by this study became one input that led to the decision of maintaining the current periodicity by which NAVSUP WSS sets quarterly inventory levels.

It is also important to discuss what we did not find. Our first important caveat concerns the lack of reduction in fill rate with increases in the persistence parameter. In this particular case, we observed that the limit that persistence could be enforced was above the critical threshold where it would impact simulated fill rates. In this way, we could increase the persistence parameter to an arbitrarily large number and not affect fill rates. However, we do not have evidence that this is true generally. It may well be that this is simply a coincidence of this particular type of material, for these demands, and at this budget level.

The next important caveat is that our conclusions are based on only 4.5 years of data. We showed that simulated fill rates did not degrade with increases in the persistence parameter for this time period only. We also showed that the difference between a good and bad concept of operations takes time to develop. It is possible that some level of persistence does impact long-term fill rates when viewed from a longer term horizon.

ACKNOWLEDGMENTS
The authors thank Naval Supply Systems Command for the support of this research.

REFERENCES
Bachman T, Williams P, Cheman K, Curtis J, Carrol R (2016) PNG: Effective Inventory Control for Items with Highly Variable Demand, *Interfaces*, Vol. 46, No. 1, January–February 2016, pp. 18–32.

Ellis D, Cardillo J, Motter, M (2017) Information regarding NAVSUP WSS operations and WIOM implementation provided to author in person via personal communication during site visit to NAVSUP WSS, June 13–15.

Köchel P and Nieländer U (2005) Simulation-based optimisation of multi-echelon inventory systems, *International Journal of Production Economics*, Vol. 93–94, pp. 505-513.

Law A (2015) *Simulation Modeling and Analysis*, 5th ed. (McGraw-Hill Education, New York, NY)

NAVSUP (2019) NAVSUP Enterprise. Accessed April 15, 2019, https://www.navsup.navy.mil/public/navsup/enterprise/.

Roth G (2016) A simulation of alternative for wholesale inventory replenishment. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA. https://calhoun.nps.edu/handle/10945/48587.

Salmeron J, Craparo E (2017) Wholesale Inventory Optimization Model. Naval Postgraduate School, Monterey, CA.

Silver E, Pyke D, Peterson R (1998) *Inventory Management and Production Planning and Scheduling*, 3rd ed. (John Wiley and Sons, Hoboken, NJ).

Tersine R (1994) *Principles of Inventory and Materials Management*, 4th ed. (PTR Prentice Hall, Upper Saddle River, NJ).

Teter S (2018) Simulated operating concepts for WIOM implementation. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/58269

AUTHORS BIOGRAPHY

Lieutenant Commander Sean Teter (United States Navy) is currently serving as an Operations Research Logistics Analyst and Distribution Planner for U.S. Transportation Command. His work focuses on the practical application of quantitative techniques to military logistics and supply problems. Operationally, he has served at sea onboard the ballistic missile submarine USS WEST VIRGINIA and the amphibious assault ship USS BATAAN. LCDR Teter received his B.A. from Florida State University and his M.S. from the Naval Postgraduate School.

Dr. Emily Craparo is an Associate Professor of Operations Research (OR) at the Naval Postgraduate School (NPS). Her research and teaching interests focus on mathematical optimization and the design of efficient algorithms. Dr. Craparo has provided analytical and modeling support to the US Army, Navy, and Marine Corps on a variety of topics related to decision making under uncertainty and resource constraints. Her academic publications include topics such as sensor placement, optimal operation of energy systems, and capital planning under uncertainty. Prior to joining the NPS OR faculty, Dr. Craparo completed National Research Council Postdoctoral Fellowship at NPS. Dr. Craparo received her S.B., S.M., and Ph.D. from the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology (MIT).

Dr. Javier Salmerón is a Professor of Operations Research at the Naval Postgraduate School (NPS) where he teaches graduate courses in optimization and operations management. His research focuses in the area of applied optimization, where he has developed multiple tools to guide decisions in both civilian and military problems. He has published in multiple peer-reviewed journals and is the recipient of several research and teaching awards. Before his time with NPS, Dr. Salmerón worked for Spanish electric utility Iberdrola and was adjunct professor at the Statistics School of Complutense University of Madrid.

A SIMULATION TOOL FOR MASS TRANSFER INSIDE COMPRESSED AIR VESSEL FOR WATER NETWORKS PRESSURISATION

Andrea Volpi [a], Eleonora Bottani [b]

[a], [b] Department of Engineering and Architecture, University of Parma (Italy)

(a) andrea.volpi@unipr.it, [b] eleonora.bottani@unipr.it

ABSTRACT

Feeding a water distribution network with the correct pressure is a fundamental requirement for its proper operation; to this end, a simple and reliable solution commonly adopted in small and medium industrial plants is the adoption of a pressure vessel. For small systems, a membrane seals the system water from the gas compartment, anyway, as the size of the vessel increases, the adoption of sealing diaphragm or bladder is no longer feasible, and thus there is a direct contact between air and water. The high pressure of the vessel, combined with the cyclic loading and unloading phases, which replace the water inside the tank, leads to a considerable mass transfer phenomenon of air inside water. The loss of air mass cannot be monitored and detected by simply controlling system pressures; to this extent, water level measurement and reference analytical models are required. Since there is a lack in scientific literature of these models, the present study presents a model for mass transfer estimate in the systems described, starting from a real pilot plant. The main results of the model implementation in a spreadsheet, in terms of the trend of the key model parameters in time, are also reported and discussed.

Keywords: pressure vessel, compressed air vessel, mass transfer, water pressurisation systems

1. INTRODUCTION

A correct pressure is the fundamental requirement for the proper operation of water distribution networks; to this end, a simple and reliable solution is the adoption of a pressure vessel. The vessel itself is independent on the electrical power and easy to operate. For small systems, a diaphragm seals the system water from the gas compartment, avoiding any contamination, corrosion and pressure loss. The result is a fully closed system that does not suffer from corrosion or other gas-related problems.

Pressurization stations are the further development of the traditional diaphragm expansion vessels for large volume and/or high-pressure systems. The principle differs because of the use of an additional control unit, which allows transferring the expansion volume to a separate expansion vessel. Due to the highly accurate control, the pressure changes in the system are kept to minimum.

Once connected to the water system, the pump starts to raise the pressure letting the water filling in the bladder. When the pressure reaches its maximum threshold value p_{MAX}, the pump stops. Inside the tank, there is the greatest quantity of water possible. Obviously, the membrane dilates and occupies almost all the volume of the tank. When water is required by the system, it starts flowing out of the tank without using the pump but just expanding the air cushion, whose pressure decreases.

The process goes on and the membrane deflates until the pressure reaches its minimal threshold value p_{MIN}. At this stage the membrane is back to its initial size, the pump starts working again and a new cycle begins. Since the pump always grants the maximum water flow, its insertions are kept to the minimum. Moreover, since air chambers show good ability in controlling the pressure surge from a water hammer (WH) phenomenon, compressed air vessel (CAV) is often adopted in a pressurized system for limiting pressure spikes. The system is very flexible; anyway, as the volume of water increases, the adoption of sealing diaphragm or bladder is no longer feasible, and thus there is a direct contact between air and water. The system still works as described above; therefore, the high pressure of the vessel, combined with the cyclic loading and unloading phase (replacing the water inside the tank), involves a mass transfer phenomenon of air inside water.

The aim of the presented work is to present an analytic model capable of representing the behaviour of the above-mentioned system, i.e. a pressure vessel in which air compression and mass transfer phenomena happen.

2. LITERATURE REVIEW

Many authors have investigated the behaviour of pressure vessels as a remedy for WH; Ivljanin et al. (2018) have used numerical simulation to describe the homogeneous two-phase flow model and a non-equilibrium system of air mass transfer between dispersed air bubbles and continuous liquid water during WH transient with gaseous cavitation.

Besharat et al. (2016) have simulated an air chamber and studied the behaviour of air inside it, with a CAV in a pressurized system. They carried out experimental tests, 1D and 2D computational fluid dynamics (CFD) simulations for an air pocket (AP) within a CAV in the case of rapid pressurization and the occurrence of WH in a pressurized system.

Again, Besharat et al. (2017) have studied an AP, confined in CAV, under several different WH events to

better define the use of protection devices or compressed air energy storage systems. This research focused on the size of an AP within an air vessel and tried to describe how it affects important parameters of the system, i.e., pressure in the pipe, stored pressure, flow velocity, displaced volume of water and water level in the CAV.
Zhou et al. (2011) have investigated the pressure variations associated with a filling undulating pipeline containing an entrapped AP both experimentally and numerically. The influence of entrapped air on abnormal transient pressures was often ambiguous because the compressibility of the air pocket allowed the liquid flow to accelerate but also partly cushioned the system, with the balance of these tendencies being associated with the initial void fraction of the AP.
Other studies have focused on the gas absorption phenomenon under different circumstances. Okayama et al. (2018) observed and calculated with CFD method plunging pool formation and gas absorption phenomena during tapping.
Vlyssides et al. (2003) have studied the transfer rate of air to water in a sparged agitated pressure vessel. The mass transfer coefficient of air was proved to depend on the mixing energy, the airflow rate as well as the vessel's pressure and temperature.
These studies represent just some examples of the topics investigated by the current literature; as it can be noticed the air cushion in CAV is often analysed in conjunction with WH or pressure transient, while mass transfer of gas is mentioned in several processes but not in steady pressure vessel for water. Hence, there is a lack in scientific literature of a model for water vessel equipped with an air cushion as a pressurisation system for a water distribution network, which is the object of the present study.
The paper is organised as follows: Section 3 (Materials and methods) describes the working principles of a real pilot pressure vessel system (3.1) and the key features of its components (3.2); then the analytical models without mass transfer (3.3) and with mass transfer (3.4) are deeply described. Section 4 presents the obtained results and, finally, Section 5 reports summarizing conclusions.

3. MATERIALS AND METHODS

3.1. Pressure vessel principles

In line with the purpose of this study, a small-scale pilot plant of a pressurisation system for a water distribution network was built and installed in a laboratory at the Department of Engineering and Architecture of the University of Parma. This plant was useful to observe a real system and represented the basis for the development of the simulation model, to ensure its realistic behaviour.
The pressure plant is represented in Figure 1; its main components are: (1) tank, (2) pump, (3) check valve, (4) CAV, (5) vessel supports, (6) level probe, (7) and (8) pressure probes, (9) level indicator, (10) safety valve, (11) interception valve, (12) compressed air inlet, (13) controller (PLC).

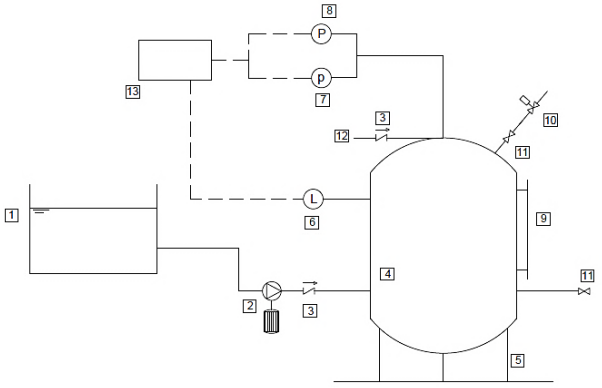


Figure 1: experimental setup.

At the beginning of the experiment, a small quantity of water is loaded in the vessel and then the CAV is pressurized by means of an external connection with compressed air distribution systems. The pressure of the air cushion inside the vessel is raised to the minimum value (p_{MIN}) required by the water distribution plant, when the minimum quantity of water is inside the vessel. The air volume obtained is referred as V_2.
Once the controller is turned on, the pump starts raising the pressure further, by letting the water filling in the vessel. When the pressure reaches its maximum threshold value (p_{MAX}), the pump stops. Inside the tank there is the greatest possible quantity of water, and the trapped air is compressed to its minimum volume V_1. At the water-air interface, mass transfer takes place and, due to the high air pressure, an amount of air passes into the water until the new maximum allowed concentration is reached.
If water is requested by the distribution network, it starts flowing out of the tank without using the pump but just expanding the air cushion, whose pressure decreases. The process goes on until the pressure reaches its minimum threshold value p_{MIN}; when this happens, the pump starts working again and a new cycle begins. It must be noticed that the air concentration of fresh water, pumped inside the vessel, is not balanced inside the CAV, due to the increased water pressure, and thus a mass transfer phenomenon of air solubilisation into water happens. Since the pump always grants the maximum water flow Q_{MAX}, its insertions are reduced to the minimum; in fact, the size of the vessel is intended to reduce to an acceptable value the switch-on cycles of the pump according to the starts per hour admitted by the electric motor.
The object of the present work is to simulate a finite number of operating cycles as described above, in order to point out the mass transfer phenomenon which takes place inside the CAV. In particular, the pressure vessel and the trend of compressed air mass (weight and composition) are studied as a function of required water flow of the distribution network over time. Intentionally, the influence of temperature on the system is neglected, since all the air transformations take place at a constant room (and water) temperature, equals to 25°C.

3.2. System description

Different elements compose the pressurization group described:

- layout elements like feeding tank, pressure vessel, valves, piping, centrifugal pump;
- monitoring system composed of pressure probes, flow meter, level gauge;
- actuators for water pump, compressed air input valve and distribution network output valve.

The pump, whose main features are presented in Table 1, is 12-volt powered and it supplies water inside the CAV; the water source is an external tank at atmospheric pressure (1 absolute atm or 1 ata). According to the datasheet of the pump, it is possible to obtain the characteristic curve, showing the delivered flow Q_{MAX} as a function of the operating voltage (volt) and discharge pressure.

Table 1: pump features.

Fuel pressure	5 bar or 8 bar (relative)
Delivery rate (5 bar, 25 C)	260 ± 5 l/h at 14 volt
Delivery rate (8 bar, 25 C)	220 ± 5 l/h at 14 volt
Pressure limiting valve	10 to 12.5 bar
Fuel compatibility	Up to E85 with shorter lifetime
Operating temperature range	-20 to +90 C
Storage temperature range	-40 to +70 C
Max vibration	3 mm at 10 to 18 Hz ≤40 m/s^2 at 18 to 60 Hz

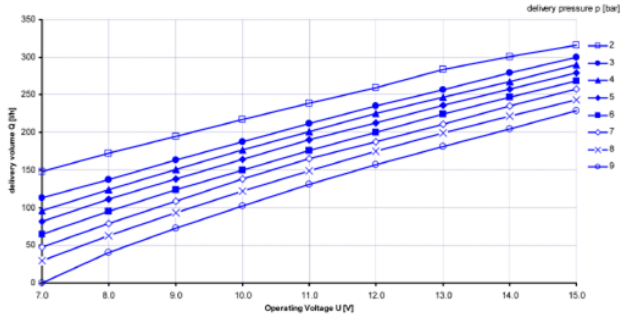


Figure 2: characteristic curve of the pump.

According to the maximum flow rate Q_{MAX} [*l/min*], it is also possible to determine the maximum number of pump switch-on per hour, indicated with *n*, according to the following equation:

$$n = \frac{30 * Q_{MAX}}{V_u} \left[\frac{switch-on}{hour}\right] \quad (1)$$

where V_u, given by V_2-V_1, represents the difference between the volume of water at p_{MIN} and the volume of water at p_{MAX}. The pressure vessel is a steel tank with an air cushion, the overall volume is $V_{PRESSURE\_VESSEL}$ (equal to 11 liters), the inner diameter is $d_{PRESSURE\_VESSEL}$ (equal to 17 cm), the cross-section is $S_{PRESSURE\_VESSEL}$ (equal to 227 cm^2), easily obtainable from the following relation:

$$S_{PRESSURE\_VESSEL} = \frac{\pi}{4} * d_{PRESSURE\_VESSEL}^{\ 2} \quad (2)$$

The height of the pressure vessel is $h_{PRESSURE\_VESSEL}$ (47 cm). The water supplied by the pump compresses the air cushion contained in the vessel until the pump stops when p_{MAX} is reached; actually, a part of the air mass in the cushion is then solubilized in the water through the air-liquid interface. Thus, the air mass inside the CAV decreases during the various system operating cycles (loading and unloading water).

The check valves allow one flow direction only.

The hoses connect different piping elements, the tank to the CAV, and the outlet port of the CAV back to the tank. This simplified configuration is used for simulation in the experimental lab equipment; in a real industrial case the outlet port is connected to the water distribution network covering the whole plant. Hoses have an overall length L=2 m and a diameter D=15 mm.

The monitoring system consists in a pressure sensor connected to air phase of the CAV, a flowmeter for the water flow on the outlet port, a level gauge measuring the water level in the CAV.

Water level is essential for the estimate of the air mass in the CAV; in fact, as the air mass decreases, the water level becomes higher with respect to the preset pressure values (respectively p_{MIN} and p_{MAX}). A combined measure of air pressure and water level is used by the controller to feedback the system during normal operating cycles (activating the pump) and gives an estimate of the decreasing trend of the air cushion (activating the air compressor or the air inlet valve to restore normal behavior).

3.3. Simulation model without mass transfer

The study of the transient operation of the pressure vessel starts with the modeling of the system and therefore with the determination of its parameters over time. In particular, the simulation model gives the current volumes of water (V_{WATER}) and air (V_{AIR}), both in liters; the air pressure (p_{AIR}), in absolute and relative atmospheres; the water flow at the outlet port and the delivery rate of the pump.

In a simplified case, in which the phenomenon of the air solubility is neglected, the following formula describes the air transformation in the CAV:

$$p_{AIR_{(t)}} * V_{AIR_{(t)}} = p_{AIR_{(t+1)}} * V_{AIR_{(t+1)}} \quad (3)$$

where *t* is a generic simulation step and *t+1* is the next step, simulation interval Δt between *t* and *t+1* is set to 1 second. As boundary condition, at the beginning of the simulation (*t=1*) p_{AIR} is set to atmospheric pressure (1 ata) and V_{WATER} to 3 liters.

For each simulation step, the model first calculates the $V_{AIR_{(t+1)}}$ [*l*], taking into account the volume of water $V_{PUMP_{(t+1)}}$ pumped inside the CAV during the simulation step; $V_{PUMP_{(t+1)}}$ is defined as the product of the supply flow rate of the pump Q_{PUMP} [*l/min*] multiplied by the time interval Δt.

In the same way the water volume $V_{USER_{(t+1)}}[l]$ at the outlet (i.e. the amount delivered to the distribution network) is computed as the product of the required flow Q_{USER} [*l/min*] and the considered time interval Δt.
The pump operates in on-off mode. It is activated when the air pressure inside the pressure vessel is lower than the minimum set point pressure (p_{MIN}=3 ata), while it stops whenever the air pressure reaches the maximum target value (p_{MAX}=4 ata). In particular, as shown in Figure 2 and Table 1, the delivery rate is not constant; rather, it varies according to the following equation (reported also in Table 2):

$$Q_{PUMP} = 0.08 * p_{OUTLET}^2 - 0.73 * p_{OUTLET} + 5.39 \quad (4)$$

The above relation is derived through an interpolation of Figure 2 and p_{OUTLET} is expressed in relative atmospheres (and it equals p_{AIR} inside the CAV). Equation 4 expresses the water flow entering the pressure vessel as a function of air pressure of the CAV, which is itself a function of time.

Table 2: pump delivery rate.

p_{OUTLET} (p_{AIR} of CAV)	Q_{PUMP} [l/h]	Q_{PUMP} [l/min]
2 atm	255	4.25
3 atm	235	3.92
4 atm	225	3.75

Another on-off system, using a valve as actuator, regulates the flow of water for the users Q_{USER} as a function of the time. When the valve is activated, it opens the flow section and the water flows from the CAV to the user (actually it goes back to the tank in our lab system). Given section 1 placed on the water surface of the pressure vessel, and section 2 at the end of the outlet pipeline, the following relation is used to estimate the water flow:

$$Q_{USER} = \sqrt{\frac{g*(z_1 - z_2) + \frac{p_1 - p_2}{\rho}}{R}} \quad (5)$$

Where:

- ρ is the density of water at 25°C (1,000 kg/m^3)
- g is the gravitational acceleration (9.81 m/s^2)
- R is an overall friction coefficient [m^{-4}] and can be derived applying Equation 6:

$$R = \frac{8*\lambda*L_{PIPE}}{\pi^2*D_{PIPE}^5} \quad (6)$$

R depends on the geometrical characteristics of the piping (hose and valves, length and diameter) and on a dimensionless friction coefficient λ, derived from the Reynolds number (*Re*) and set at 0.023. From this relationship and from the characteristics of the system it is possible to compute, for any time *t*, the water level z_1 [*cm*] inside the pressure vessel; this can be obtained as the ratio between the volume of water V_{WATER} and the cross section of the pressure vessel:

$$z_1 = \frac{V_{WATER}}{S_{PRESSURE\ VESSEL}} \quad (7)$$

For section 2, z_2=0, being the outlet pipe height set to the ground reference; p_1=p_{AIR} inside the pressure vessel and p_2=1 ata (atmospheric pressure).

3.4. Simulation model with mass transfer

In the real scenario, the solubility phenomenon of air cushion into the water takes place; thus, there is a variation over time of the air mass m_{AIR} [g] inside the vessel. In a generic time, the air mass consists mainly in the sum of the oxygen (O_2) and nitrogen (N_2) contributions, i.e.:

$$m_{AIR_{(t)}} = m_{O_{2_{(t)}}} + m_{N_{2_{(t)}}} \quad (8)$$

In atmosphere, m_{N_2} equals 79% of the air mass while m_{O_2} is approximately 21%, neglecting the mass percentages of the other elements, such as noble gases and CO_2.
Any variation of m_{AIR} has effects on the air pressure, according to the following relation, which is valid under the (realistic) hypothesis of an isothermal transformations:

$$\frac{p_{AIR(t)}*V_{AIR(t)}}{m_{AIR(t)}} = \frac{p_{AIR(t+1)}*V_{AIR(t+1)}}{m_{AIR(t+1)}} \quad (9)$$

It is possible to quantify the contributions of oxygen m_{O_2} and nitrogen m_{N_2} using their respective mass fractions w:

$$m_{O_2} = m_{AIR} * w_{O2} \quad (10)$$
$$m_{N_2} = m_{AIR} * w_{N2} \quad (11)$$

where w_{O2} and w_{N2} are the mass fractions of the two gases (i.e. w_{O2}=21% and w_{N2}=79% as previously mentioned).
According to Section 3.3, for any simulation step the air pressure is computed according to the available volume V_{AIR}. The model thus estimates the increase in the concentration of gases in the water as a function of the pressure of the vessel.
In every time interval Δt, a small part of m_{AIR} moves from the air cushion and dissolves in water, leading to the dissolution of oxygen and nitrogen:

$$m_{O_{2_{DISSOLVED_{(t+1)}}}} = C_{O_{2_{DISSOLVED_{(t+1)}}}} * V_{WATER_{(t)}} \quad (12)$$
$$m_{N_{2_{DISSOLVED_{(t+1)}}}} = C_{N_{2_{DISSOLVED_{(t+1)}}}} * V_{WATER_{(t)}} \quad (13)$$

where $C_{O2_{DISSOLVED_{(t+1)}}}$ and $C_{N2_{DISSOLVED_{(t+1)}}}$ are the concentrations, in [*mg/l*], of oxygen and nitrogen respectively in the simulation step *t+1*. In turn, these

concentrations depend on the simulation step and on the pressure p_{AIR}. Moreover, their values are governed by two variables, i.e. the Henry's law solubility constants K_H and the solubilization speed of oxygen and nitrogen in water. The Henry's constants depend on media, temperature and pressure; for the considered system they account for:

$K_{H,O2}$ at 25°C 756.7 [atm l mol^{-1}]
$K_{H,N2}$ at 25°C 1600 [atm l mol^{-1}]

At time *t+1*, the masses of O_2 and N_2 in the air cushion thus account for:

$$m_{O2_{(t+1)}} = m_{O2_{(t)}} - m_{O2\ DISSOVED_{(t+1)}} \quad (14)$$
$$m_{N2_{(t+1)}} = m_{N2_{(t)}} - m_{N2\ DISSOLVED_{(t+1)}} \quad (15)$$

The air mass therefore varies over time according to equations (14)-(15); concerning the water phase, since the volume V_{WATER} is known it is possible to derive the concentrations of O_2 and N_2 [*g/l*] according to the following equations:

$$C_{O2_{(t+1)}} = \frac{C_{O2_{(t)}} * V_{WATER} + C_{O2_{(1)}} * V_{PUMP}}{V_{WATER} + V_{PUMP}} + \frac{C_{O2\,DISSOLVED_{(t+1)}}}{1000} \quad (16)$$
$$C_{N2_{(t+1)}} = \frac{C_{N2_{(t)}} * V_{WATER} + C_{N2_{(1)}} * V_{PUMP}}{V_{WATER} + V_{PUMP}} + \frac{C_{N2\,DISSOLVED_{(t+1)}}}{1000} \quad (17)$$

The first part of the sum in the equations above represents the weighted average of the concentrations of O_2 and N_2 in the V_{WATER} stored in the pressure vessel at time *t* plus the volume added by the feeding pump in Δt. In fact, the pump introduces in the vessel a volume of water V_{PUMP} for which the concentration of oxygen and nitrogen is not the same as that of the liquid already contained. Rather, pumped water has a concentration of O_2 and N_2 given by the balance resulting in the tank at atmospheric conditions $C_{O2_{(1)}}$ and $C_{N2_{(1)}}$, described below.
The second term in equations (16)-(17) is the contribution involved by solubilization. To calculate the dissolved concentrations, it is first necessary to estimate the maximum allowed concentration for oxygen and nitrogen, according to the Henry's law of solubility:

$$C_{O2,MAX_{(t+1)}} = \frac{p_{(t+1)} * w_{O2}}{K_{H,O2}} * 31.99 \left[\frac{g}{mol}\right] \quad (18)$$
$$C_{N2,MAX_{(t+1)}} = \frac{p_{(t+1)} * w_{N2}}{K_{H,N2}} * 28.01 \left[\frac{g}{mol}\right] \quad (19)$$

The first term of the products is the mass fractions of O_2 and N_2, the pressures in absolute atmospheres, the molecular weights of O_2 and N_2 and, at denominator, the Henry constants for O_2 and N_2.
At the beginning of the simulation (*t*=1), the maximum concentrations of O_2 and N_2 are calculated at atmospheric conditions (1 ata):

$$C_{O2,MAX_{(1)}} = C_{O2_{(1)}} = \frac{1 * w_{O2}}{K_{H,O2}} * 31.99 \left[\frac{g}{mol}\right] \quad (20)$$
$$C_{N2,MAX_{(1)}} = C_{N2_{(1)}} = \frac{1 * w_{N2}}{K_{H,N2}} * 28.01 \left[\frac{g}{mol}\right] \quad (21)$$

Concerning the increase of concentration of gases in water, the speed of gas solubilization is proportional to the difference of concentrations and to the surface/volume ratio of the reactor where the phenomenon takes place:

$$n_{GAS} = k_L \frac{A}{V}(C^* - C_L) = k_L a (C^* - C_L) \quad (22)$$

where:

- k_L is the transfer coefficient of the liquid film [$m\ s^{-1}$];
- *A* and *V* are respectively equal to the gas-liquid exchange area and the volume of the reactor considered;
- *a* is the interface area (*A/V*) [$m^2\ m^{-3}$];
- $(C^*\text{-}C_L)$ is the concentration gradient [$mg\ l^{-1}$], also called "diving force";
- C^* is the maximum concentration (saturation solubility) of the considered gas in the liquid [$mg\ l^{-1}$];
- C_L is the current concentration, at time *t*, of the gas in the liquid [$mg\ l^{-1}$].

Each component of equation (22) has been evaluated for the specific case under examination. The product $k_L a$ is characteristic for every ventilation system with an air-liquid interface. The interfacial area *a* is given by the ratio between the surface area (*A*, which is $S_{PRESSURE\_VESSEL}$) and the volume (V_{WATER}); while the surface area is constant, the volume varies according to the quantities of water inside the CAV. An analysis of literature shows that k_L has to be experimentally determined for each specific system; to be more precise, the whole amount $k_L a$ should be investigated rather than the two terms (k_L and *a*) separately. Hence, to run the simulator, the following values were set for the case under examination, according to the literature (http://astratto.info/lacqua-ha-un-elevato-potere-solvente-verso-i-solidi-ionici-e-i.html):

$k_L\ O_2 = 2.72 * 10^{-6}$ m/s
$k_L\ N_2 = 2.72 * 10^{-6}$ m/s.

The execution of an experimental campaign, as a further development of the present work, could allow for a more precise estimate of the k_L coefficients for the system investigated.
Going back to the $C_{DISSOLVED}$ term in equations (16)-(17), the concentration of O_2 and N_2 dissolved in water at step *t+1* is determined as follows:

$$C_{O2_{DISSOLVED_{(t+1)}}} = k_{L,O2} a_t * \Delta t * (C_{O2,MAX} - C_{O2})_t \quad (23)$$
$$C_{N2_{DISSOLVED_{(t+1)}}} = k_{L,N2} a_t * \Delta t * (C_{N2,MAX} - C_{N2})_t \quad (24)$$

Equations (23)-(24) state that there is an increase in the liquid concentration for a specific gas only if the current concentration is lower than the maximum allowed concentration.

3.5. Model implementation and simulation strategy

The set of equations detailed in the previous subsections, linked together to reproduce the whole model developed, have been implemented in Microsoft Excel™ to build a simulation tool.
The objective of the tool is to describe the system under different working conditions. In particular, the model was run with a twofold aim: validating the model and estimating the mass transfer phenomena inside the vessel.
Regarding the first objective, the simulation outcomes depicted in Figures 3, 4 and 5 represent the trend of air pressure p_{AIR}, water level z_1, pump and water demand by the user (Q_{PUMP} and Q_{USER}) as a function of the time. These variables describe the macroscopic behaviour of the system, as pointed out in Section 3.3.
The second aim is the estimate of the mass transfer phenomena, described in Section 3.4. Figures 6 and 7 stress the variations of O_2 and N_2 concentrations in water (C_{O2} and C_{N2}), as well as their maximum allowed concentrations depending on p_{AIR} ($C_{O2,max}$ and $C_{N2,max}$). Finally, Figure 8 plots the O_2 and N_2 mass reduction in the CAV versus time, over a simulation time of 1 day and 216 loading cycles.
The results of the simulations are detailed in the section that follows.

4. RESULTS

Figure 3 depicts the trend in time of vessel pressure p_{AIR}, water level, pump operation (*on-off*) and users' water demand (Q_{USER} from 0 to 100%) during the initial operating activity of the CAV.

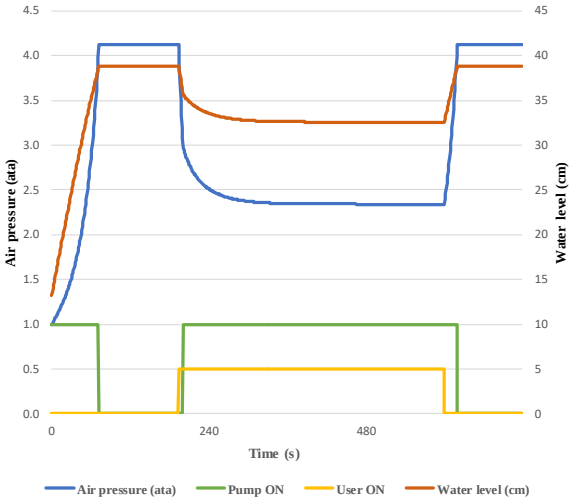


Figure 3: plot of vessel pressure, water level, pump operation and users' water demand versus time (simulation duration: 12 minutes).

The initial pressurizing phase is characterised by the water pump on, filling the vessel and reducing V_{AIR} until p_{MAX} is reached. Then a steady state with stable pressure is reached and maintained. When the distribution network requests a water flow set to 50% of Q_{MAX}, the vessel reacts by deflating the air cushion; hence, pressure drops and the water pump switches on to restore the pressurisation. A stable working point is found at about 2.3 ata; at this pressure, the water demand is balanced by the delivery rate of the pump. This state ends when the outlet valve is closed; at this point, pressure rises to p_{MAX} and, after some time, the pump is switched off as well.

Figure 4 reports the same plot over a longer time interval (1 h), where 5 loading cycles have been simulated.

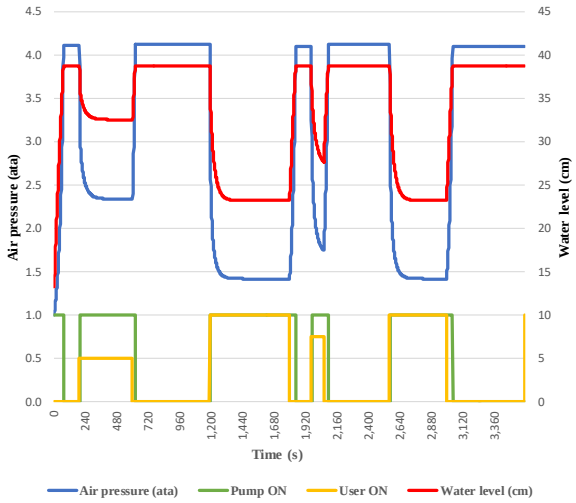


Figure 4: plot of vessel pressure, water level, pump operation and users' water demand versus time (simulation duration: 1 h).

Figure 5 focuses on the water flows of the pump and the distribution network, the latter simulated by means of a valve. The simulated scenario and conditions are the same as Figure 3; after the transient effect due to the output valve opening, the system finds a steady working condition where the two flows are perfectly balanced.

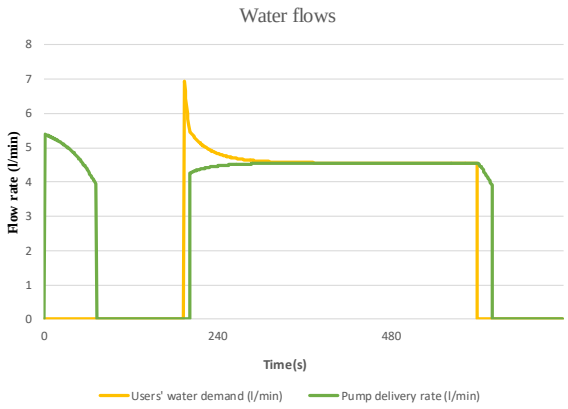


Figure 5: plot of users' water demand and pump delivery rate versus time (simulation duration: 12 minutes).

The above figures present the trend of the relevant variables which is, as a matter of facts, qualitatively confirmed by the physics underlying the problem. The next plots aim instead to investigate the hidden side of

the CAV, i.e. the mass transfer phenomenon, focusing on the behaviour of pressurised oxygen and nitrogen in the air cushion.

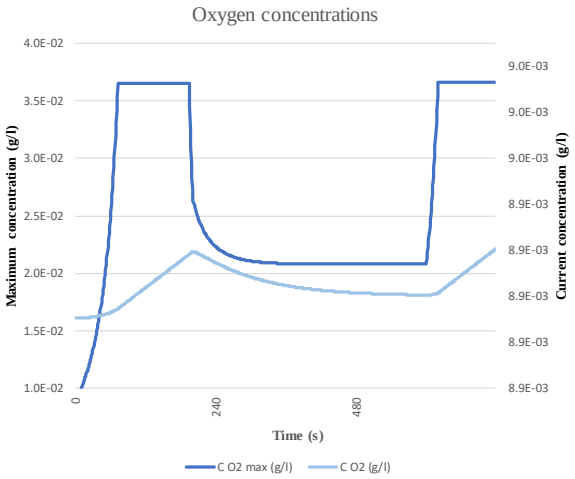


Figure 6: plot of oxygen maximum concentration in water $C_{02,MAX}$ and current concentration C_{02} versus time (initial phase, 12 minutes)

Figure 6 reports the trend of the oxygen concentrations in the first 12 minutes of system working. $C_{02,MAX}$ increases with pressure (according to the Henry's law) and the current concentration C_{02} increases as well, although it is always lower than $C_{02,MAX}$. This is consistent with the fact that the driving force $C_{O2,MAX} - C_{O2}$ is responsible for the mass transfer (see equation 23). The same behaviour can be observed in Figure 7, where the nitrogen concentrations $C_{N2,MAX}$ and C_{N2} have been added; the simulation duration in this figure has been extended to 1 h.

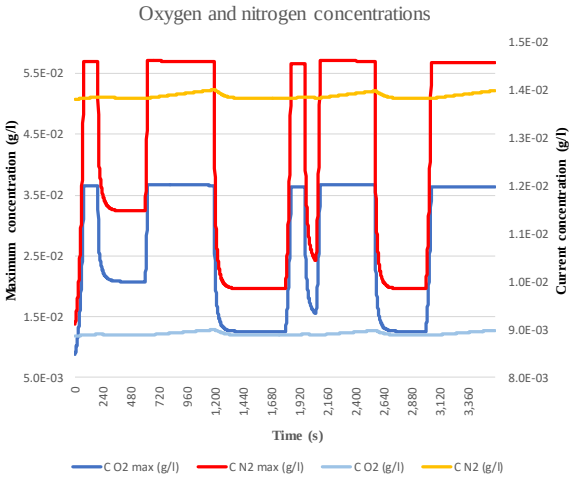


Figure 7: plot of oxygen and nitrogen maximum and current concentrations versus time (simulation duration: 1 h).

In order to observe the reduction of the trapped air in the air cushion, the simulator has been tested on multiple loading/unloading cycles. More in details, 216 cycles have been simulated; each cycle is composed of a loading phase of 200 seconds (Q_{USER}=0) followed by a unloading phase of 200 seconds (Q_{USER}=100% as per equation 5). The overall simulated time is 86,400 seconds corresponding to 1 working day. During each cycle, a small quantity of gas dissolves in water, as pointed out in Figure 8. Oxygen mass, taken as an example, decreases from 2.18 g to 2.10 g, corresponding to a 3.6% reduction.

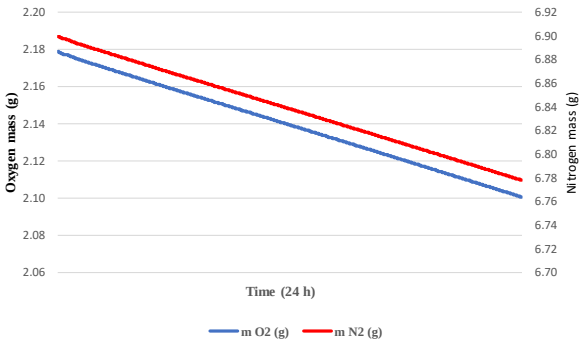


Figure 8: plot of oxygen and nitrogen mass reduction in the CAV versus time (simulation duration: 1 day, 216 load/unload cycles)

This phenomenon cannot be observed and detected by simply controlling the pressures values (p_{MIN} and p_{MAX}) because the system automatically reacts to the air mass loss by replacing it with water. This means that, although the air mass decreases in time, the pressure vessel still works correctly between p_{MIN} and p_{MAX}. The user can detect the air loss by observing an increase in the water level corresponding to p_{MAX} and some changes in the on-off cycles of the pump. More precisely, their frequency increases while the duration of the on-phase decreases.

5. CONCLUSIONS

The model developed in this paper, built in Microsoft Excel™, considers both the compression of the air cushion according to the volume of water contained into the vessel and the mass transfer phenomenon, whose rate depends on temperature and pressure. Nitrogen and oxygen contributions have been separately considered, due to their different behaviour.

Today's systems installed and running are based on negative feedback systems, whose closed loops separately control the water feeding pump (according to the pressure measured in the vessel) and the compressed air inlet valve (according to the measured water level at a given pressure). This system works fine although it is blindfolded: it just reacts to the current measured variables and does not care about their trend and history. On the contrary, thanks to the adoption of the proposed predictive model, a microcontroller or a PLC managing the system can also be able to detect possible failures before they happen and compromise the functionality of the system itself. In fact, predictive modelling can predict outcomes, such as the value of given variables, according to the state of the system. Most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred. For example, predictive models can be used

to compare the current value of pressure, or water level in the vessel, with predicted ones, according to the previous state of the system and the other measured variables. In case the match is not positive, within a certain tolerance, something has occurred and the failure might be approaching.

The results obtained in this paper allow to conclude that the model can act as an enabler of pro-active system monitoring, as it allowed for many system variables to be predicted in time. Nonetheless, some steps are still needed. In particular, results need to be validated by comparing them with those obtained (for instance) using a CFD software. Moreover, as soon as the lab equipment described is up and running, the model can be validated with experimental data acquired from the real pressure vessel. The same considerations hold true for the mass transfer coefficients used in this paper, which were taken from literature but not directly measured; they could be confirmed by in-field experiments.

REFERENCES

Besharat, M., Tarinejad, R., Taghi Aalami, M., Ramos, H. M. (2016). "Study of a Compressed Air Vessel for Controlling the Pressure Surge in Water Networks: CFD and Experimental Analysis", Water Resource Manage, 30, 2687–2702;

Besharat, M., Viseu, M. T., Ramos, H. M. (2017). "Experimental Study of Air Vessel Behavior for Energy Storage or System Protection in Water Hammer Events", Water, 9(1), 63;

Ivljanin, B., Stevanovic, V. D., Gajic, A. (2018). "Water hammer with non-equilibrium gas release", International Journal of Pressure Vessels and Piping, 165, 229–240;

Okayama, A., Nakamura, O., Higuchi, Y. (2018). "Analysis of Plunging Pool Formation and Gas Absorption Phenomenon during Tapping", ISIJ International, 58, 4;

Vlyssides, A., Barampouti, E. M., Mai, S. (2003). "Kinetics of Air Absorption by Water in Sparged Agitated Pressure Vessels", Industrial & Engineering Chemistry Research, 42(24), 6323 – 6235.

Zhou, L., Liu, D., Karney, B., Zhang, Q. (2011). "Influence of Entrapped Air Pockets on Hydraulic Transients in Water Pipelines", Journal of Hydraulic Engineering, 137(12), 1686-1692;

AUTHORS BIOGRAPHY

Eleonora BOTTANI is Associate professor in Mechanical Industrial Plants at the Department of Engineering and Architecture of the University of Parma since September 2014. She graduated (with distinction) in Industrial Engineering and Management in 2002 and got her Ph.D. in Industrial Engineering in 2006, both at the University of Parma. Her primary research activities concern logistics and supply chain management issues; secondary topics encompass industrial plants and food plants. She is author (or co-author) of more than 170 scientific papers (H-index=19), referee for more than 60 international journals, editorial board member of five scientific journals, Associate Editor for one of those journals, and editor-in-chief of a scientific journal.

Andrea VOLPI graduated in July 2003 in Mechanical Engineering at the University of Parma. Since January 2006, he has been working as Ph.D. student in Industrial Engineering Department at the same University. His major activities have been devoted to managing the research projects carried out in RFID Lab, a forefront laboratory in the same department. Since November 2015, he works as Associate Professor at University of Parma, and his research activities are mainly concerned with logistics, supply chain and operation management; thus his skills and competences are mainly related to RFID and logistics topics which are expressed in many papers produced. He acts as a referee for some international scientific journals, such as International Journal of RF Technology: Research and Applications, International Journal of Logistics: Research and Applications.

AN INTEGRATED SIMULATION-BASED CONSTRUCTION CREW ALLOCATION AND TRADE-OFF WITH ENERGY AND CARBON FOOTPRINT

Hadia Awad[(a)], Mustafa Gül [(b)], Osama Mohsen[(c)], Simaan AbouRizk[(d)]

[(a),(b),(c),(d)] Department of Civil and Environmental Engineering, University of Alberta

[(a)]haawad@ualberta.ca, [(b)]mustafa.gul@ualberta.ca, [(c)]omohsen@ualberta.ca, [(d)]abourizk@ualberta.ca

ABSTRACT

On-site construction in winter consumes a considerable amount of energy and emits a significant volume of greenhouse gases, especially in cold regions. It has been reported that on-site winter heating accounts for 34% of carbon emissions of the framing phase for panelized house construction. In this paper, in order to quantify and analyze carbon emissions from on-site construction, the on-site panelized construction process is simulated in a combined discrete and continuous event simulation model based on which the possibility of reducing activity durations are investigated for the aim of reducing emissions. The integrated simulation methodology is demonstrated using case studies in Edmonton, Canada. Carbon emission which includes propane consumption for winter heating and diesel consumption for on-site mobile equipment and vehicles is calculated. Historical temperature data is analyzed to simulate weather behavior. Results show that on-site heating is the largest contributor to carbon emissions in panelized construction.

Keywords: panelized construction, on-site winter heating, carbon footprint, continuous-event simulation, discrete-event simulation

1. INTRODUCTION

Offsite construction, including modular and prefabricated, is becoming a more accepted and adopted approach in the construction industry to achieve better quality, less environmental impact and reduced time (Hong Xian et al. 2014). According to a study by Kawecki in 2010, the deployment of this method was increased by 48% between 1992 and 2002. Panelized construction accounts for the reduction of project duration by 63%, cost by 16%, and waste by 76% (The Panelized Process 2007). There is a belief that panelized construction method has many benefits compared with other construction methods; and some of which are: improved quality of end-products (i.e. wall assemblies, roofs, etc.) which can save energy, increase process velocity, and decrease the impact of bad weather on the overall construction process (The Panelized Process 2007). It also has a contribution to sustainability by the reduction of energy consumption and decreasing GHG emissions during the process of construction (Li et al. 2014). In addition, it is more cost efficient in terms of the need for on-site crew. Therefore, with increasing the cost of materials and skilled crew and in order to accelerate the construction phases, the demand of using panelized method is rapidly increasing (Friedman and Cammalleri 1992). A study by Friedman (1992) compared the construction expenses using panelized and conventional method for one single-family and concluded that there is no significant difference between them in terms of costs. More studies in this area indicate that saving costs of up to 6% would happen in some panelized methods (Ginter 1991). It reduces the CO_2 emissions comparing with on-site construction method by decreasing trips, the usage of equipment and also winter heating (Quality, Speed and Cost 2014).

The consumption of energy during the construction process is one of the most important impacts of a building on the environment (Palaniappan 2009). The construction phase of a building demands energy, material and other resources which imposes various forms of loads on the environment (Chinin et al. 2011). A study focused on pre-panelized construction method for its potential advantage of decreasing carbon footprint and energy consumption, found that savings of 30% in carbon emissions were achieved (Kawecki 2010). The pre-panelized method involves the assembly and fabrication of panels in plant, transportation of those panels to the construction site, and finally the erection of them on site. Thus, CO_2 emissions could be generated from the extract of materials, transporting them to the plant, fabrication of panels, moving them to the site and the construction process on site (Li et al. 2014). The distance between the plant and the construction site is the largest contributor of CO_2 emissions during the construction process (Kawecki 2010).

Not only climate change is today's environmental concern, but it will also impact the future generations. Different studies have given indications of global warming and a drastic ice melting in Polar Regions. "Canada's total GHG emissions in 2017 were 716 megatonnes of carbon dioxide equivalent (Mt CO_2 eq). The decrease in emissions since 2005 was primarily driven by reduced emissions from the electricity generation sector." (Greenhouse gas emissions 2019).

In Canada, over 80 % of total national greenhouse gas emissions are associated with the production or

consumption of fossil fuels for energy purposes." (Greenhouse gas emissions 2019). Findings proved the significant role of the gas in climate change and therefore the increasing level of warmth on the earth can gravely influence our future (The Carbon Dioxide Greenhouse Effect 2011). After examining different global energy scenarios in detail, the International Energy Agency (IEA) has indicated that global primary energy use is likely to increase by 36% between 2008 and 2035 (IEA 2010a; Dodoo 2011). These findings may heighten current concerns about energy security. Furthermore, fossil fuels are very likely to account for a significant share of future primary energy use, unless effective measures are implemented to promote sustainable energy systems in the global community (IEA 2011a).

Based on a study conducted by Suzuki et al. in 1995, it was found that the least energy consumption in construction (3 GJ/m2) is for wood structures. Another study by Gonzalez and Garcia in 2006 about CO_2 emission indicated that green materials selection and architecture design can significantly reduce the GHG impact on our environment. Miner et al. (2008) compared Energy consumption and CO_2 emissions of wood-framed buildings with non-wood-based buildings and found that wood-framed buildings needed 15-16% less energy for cooling and heating purposes in comparison with concrete-based buildings. In addition, this study found that greenhouse gas emission associated with wood-framed buildings were 20-50% lower compared to concrete-based buildings or steel-based building systems (Miner 2008). Mah (2007) also states that wood waste accounts for 60% of all waste in Canada.

Systems simulation has proven its effectiveness in analyzing various manufacturing operations (Wales and AbouRizk 1996). A study made by AbouRizk et al. in 2011 states that the dynamic and complex characteristics of construction operations make it a challenge to properly estimate project duration, resource utilization, and job conditions, since they highly depend on external factors such as weather conditions, holidays, resource availability, unscheduled breakdowns, etc. In order to create a simulation model of a housing project it is required from the simulation expert to be aware of the uniqueness of each project and knowledgeable of (a) the logic and sequence of the operation; (b) simulation algorithm and techniques; and (c) software tools and applications (AbouRizk et al. 2011). Simulation is defined by AbouRizk (2011) as *"the science of modeling a construction production system and experimenting with the resulting model on a computer"*. The history of simulation software refers back to 1955-1960; namely "the period of search", and mainly took five main stages through the next 30 years to evolve, as described by Nance, 1995. Halpin (1973) was the first one to introduce the concept of simulation to construction processes and operations. That concept came after Teicholz (1963) who adapted link-node methodology to earth moving operations, and Gaarslev (1969) who compared the results of queuing theory and simulation when studying simple two and three cycle construction systems. Esfahani (2013) compared the simulation engine developed by Hajjar and AbouRizk (1996), which can get the distributions and run the model for several times, with other methods such as MS Excel spreadsheet which can also estimate the amount of CO_2 emissions but only for one year and concluded that using Simphony.NET is more accurate.

Simulation of process and construction operations has been extensively used in the past decades. For example, Mohsen et al. (2008) used Simphony.NET simulation engine to investigate the onsite assembly of the modules used to build five dormitory buildings. Altaf et al. (2015) developed an online simulation-based and RFID production control system in a panelized construction factory. Also, Ismail et al. (2017) adopted a simulation technique to support construction project planning. RazaviAlavi et al. (2017) developed a simulation model to optimize construction site layout planning. Moreover, Golabchi et al. (2018) proposed an integrated approach to design and evaluate construction safety and labor productivity using simulation modeling and visualization. Mohsen et al. (2018) utilized discrete-event simulation to model the floor operations at a cabinet manufacturing facility.

2. METHODOLOGY

The scope of this study focuses on using different simulation techniques (discrete and continuous event simulation methods) for the purpose of quantifying and investigating the possibility of minimizing the CO_2 emissions associated with the transportation of materials from the plant to the construction site, on-site construction equipment, and on-site winter heating taking place during the panelized construction of single family houses, using the General Purpose Template GPT in a simulation engine developed by Hajjar and AbouRizk (1996) at the University of Alberta.

The actual input data of this study is collected from 200 panelized housing projects performed recently by Landmark Group of Builders in Edmonton, Alberta between 2011 and 2013. The construction process involved six main stages: (1) date-to-field - framing start, (2) framing start – siding start, (3) siding start – drywall boarding start, (4) drywall boarding start-stage1 finishing start, (5) stage1 finishing start – carpet finish, and (6) carpet finish – possession date. The obtained data (also considered as historical data) involves the precise start and finish dates and durations of each of the mentioned stages. For the detailed activity durations, transportation needs, and resources required to perform those activities, the researchers considered experts' knowledge to obtain the most likely (mode), or minimum and maximum (uniform) durations, number of vehicle trips and their types, and special resources. Actual construction operations as performed in the field are stochastic by nature (AbouRizk et al. 2011). Consequently, the researchers made advantage of the available historical data to (1)

validate the simulation model, and (2) investigate the gaps between the expert and actual data, which result from the weather conditions, unavailability of different resources, poor documentation and abstraction of the project details, and other unknown reasons. As previously mentioned, there are many sources of CO_2 emissions during the construction phase. Figure 1 summarizes the research methodology followed to conduct this study. The previously mentioned simulation engine is used to simulate the panelized single family house construction process, temperature variation throughout the year, as well as the associated CO_2 emissions with such process.

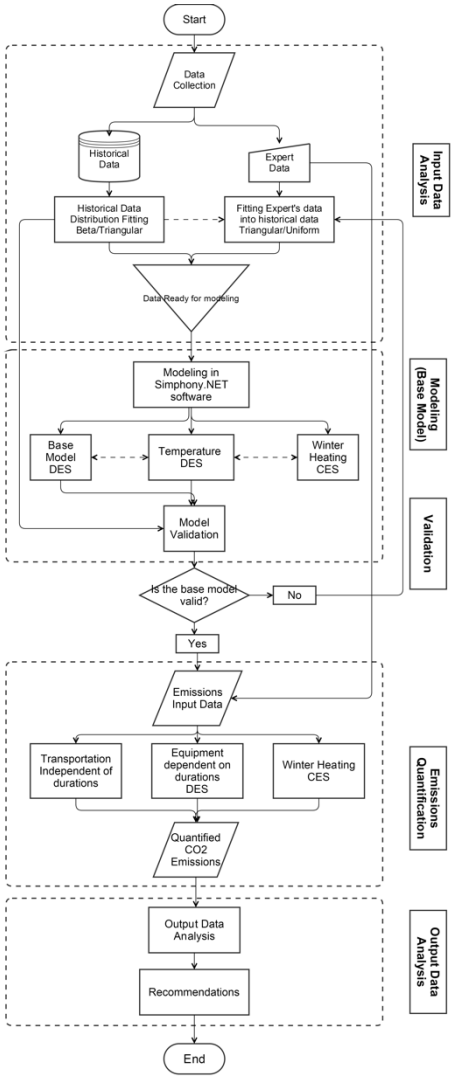


Figure 1: Research methodology

3. DATA COLLECTION AND ANALYSIS

The collection and analysis of input data are considered as a major task in simulation, where one of the first steps in that task is to hypothesize a distributional form for the input data (Banks 1996). A study conducted by AbouRizk in 1992 states two types of data experienced by the random process simulationist; (a) collected observations from historical data (as demonstrated in Table 1); and (b) judgement of a person knowledgeable about the process, where there are no available observations. A study by AbouRizk in 1990 also focused on choosing appropriate distribution as an input for the model. It states that if the data records are unavailable, the modeller must rely on intellectual advice of an expert for modeling. Experts usually specify a uniform distribution or triangular distribution for a data. Choosing a proper distribution is a very important part of the modeling (AbouRizk and Halpin 1990).

Similarly, in the present study, there are two types of data collected to serve this project; historical data mentioned previously, and experts' knowledge data collected from relevant researchers and industrial partners. The historical data is generalized to six main stages of house construction, but neither includes detailed information about resources nor sub-activity durations. This data is accurate and represents real durations for each of the six stages mentioned before. It also includes start and finish dates of each stage. The research team took a good advantage of those dates to investigate and compare between the time it takes to build a house in both summer and winter; in other words snow versus no-snow seasons. In Alberta, it is assumed that snow season starts on the first of October and ends by the end of March, and followed by no-snow season commencing from the first of April till the end of September. Data was split and analyzed on that basis. The historical data was used in that project in three main ways; (a) to validate the output data from the base simulation model, since it mimics the actual construction process; (b) to help fit experts' data input into proper probability distributions; and (c) to fill the gaps that result from the divergence between real construction process data reflecting its stochastic nature, and theoretical data obtained from experts, which is most likely deterministic. In his book "Discrete-Event System Simulation" (1996), Jerry Banks stated that the validation process should be achieved in an iterative process of calibrating the simulated model and comparing it against the actual system behavior, and also manipulating the gaps between the two, to improve the simulated model. This point will be explained later in the following paragraph. As mentioned earlier, the dynamic and complex characteristics of construction operations make it a challenge to properly estimate project duration, resource utilization, and job conditions (AbouRizk 2011). According to AbouRizk (1992), the beta distribution has proven its advantage in modeling the activity durations for most of the simulation applications in construction. Contrarily, the Johnson and Pearson systems do not recommend the beta distribution (AbouRizk 1992). In this study two software applications were used to fit the historical data into proper distributions; (a) @Risk software, and (b) simulation engine (Hajjar and AbouRizk 1996). The researchers compared the results and found great similarities between both software platforms in terms of fitted distributions.

Most of the data was fitted into either beta of triangular distributions. Reviewing the historical data, it was observed that the panelized construction process takes an average of 150 working days to construct a house in winter. Consequently, it was hard for the research group to collect field data in the limited time of the project, and the only possible way to obtain detailed activity durations was the use of other relevant studies, or experts' knowledge input. This type of data was given as one deterministic value, which is regarded as the mode. In order to give the model a stochastic behavior, each activity duration was inherited a low and ultimate value with the same ratio of the historical data distribution it lies within. Consequently, the parameters of a triangular distribution can be determined (low, ultimate, and mode).

Table 1: Fitted distributions of historical data using @Risk and simulation engine (Hajjar and AbouRizk 1996)

	From	To	Oct1-Mar 31	Apr 1-Sep 30
1	Date-to-field	Framing Start	Beta (8.9852,3.9139,1.897,28.366)	Triangular (5,33.4,58)
2	Framing Start	Siding Start	Beta (1.06,3.6,8,55)	Beta (0.973,3.291, 8.5,50.5)
3	Siding Start	Drywall Boarding Start	Beta (2.905,5.658,10,40)	Beta (1.1157,3.048 ,12,72)
4	Drywall Boarding Start	Finishing Stage1 Start	Beta (3.24,5.0697,12,35)	Triangular (15,15,33)
5	Finishing Stage1 Start	Carpet Finish	Beta (1.537,8.416,20,45)	Triangular (14,26,29)
6	Carpet Finish	Possession Stage	Triangular (18,22,99)	Triangular (18,26,35)

4. MODEL

By adopting historical data, experts' knowledge data, and similar previous studies, a simulation model was created to mimic the real panelized construction process for the purpose of quantifying the CO_2 emissions associated with construction. The modeling process in this study consists of two major components; (a) base model which represents the construction process in detail; and (b) quantification of emissions resulting from the operations in the base model.

4.1. Base Model

The General Purpose Template GPT in the simulation engine was selected to simulate the construction process through a DES model. The time unit was selected to be in actual dates, so that the research team can investigate the duration and CO_2 emissions with respect to different project start dates. From the historical data, it was observed that each activity-duration varies according to the time of the year it was commenced. Figure A-3 in the appendix demonstrates the panelized construction process. After running the model for 30 run counts, the results have shown that the project duration varies between 133 and 156 days according to the date of the year the project started. On the other side, the propane gas tank consumption/refill rate was modelled as a CES, where the outside temperature controls whether or not the heater is turned on. The refill rate of the propane tank is controlled by the consumption rate of propane gas. A five-ton truck refills the propane tank when the amount of propane gas drops to 10 Liters. Figure A-6, A-7, and A-8 demonstrate the combined DES and CES models of the propane tank rate of consumption/refill and the refilling truck cycle.

4.2. Emissions Quantification

The purpose of this study is to investigate the carbon dioxide emissions resulting from the panelized construction operation, focusing on three main emission types: (1) crew and material transportation emissions, (2) on-site equipment emissions, and (3) winter heating emissions. The first type relies on how many vehicle trips are performed back and forth between the construction site and the manufacturing plant to transport either the crew or materials. This type of emission is easy to quantify since it only depends on the amount of labour and materials to transport, and distance to travel. It was found from relevant studies that the average distance travelled by the different types of vehicles is equal to 40 Kilometers. The second type is split into two sub-categories; duration-dependent, and duration-independent. The duration-dependent equipment emissions are those affected by the task duration, such as the compressor and generator. In this case, the shorter the activity duration, the lower the carbon dioxide emissions will be. Unlikely, the duration-independent equipment performs its tasks based on the amount of work to be done, regardless of how fast or intensive the laborers are, such as the crane and the excavator. The equations for emission quantification of those two types of emissions were obtained from a study by Li et al. (2014) at the University of Alberta. The third emission component involves winter heating. On-site winter heating is mandatory in cold regions such as Alberta, whenever the outside temperature reaches below −5˚C. The reason behind winter heating is to keep the construction crew, as well as the building materials and equipment safe and warm during severe weather conditions in winter. Once the heater is installed, it becomes available for heating whenever heating is needed (temperature is below −5˚C). The consumption rate of the propane-gas-filled heater is 100.1 liters per day, where its capacity is 300 liters (Mah 2007). A five-ton truck is responsible for refilling the heater tank when it reaches 10 liters. Each Liter of propane produces 1.51 Kilograms of CO_2 (Mah 2007). CO_2 emissions resulting from winter heating are encountered from two main sources; (1) propane gas consumed for heating, and (2) truck trips for heater tank refilling. It is obvious that winter heating is controlled by one factor, which is outside temperature. A recent study by Li (2014) defines the daily minimum temperature by the polynomial Equation below, which has been concluded by the analysis of historical

temperature data from the city of Edmonton online database.

$$y=9*10^{-11}x^5-5*10^{-08}x^4+10^{-6}x^3+0.0022x^2-0.101x-10.393 \quad (1)$$

This polynomial equation was manipulated by a discrete-event simulation DES model (Figure A-4) in the simulation engine to mimic actual weather data (Figure A-1 and A-2), and be used to study the behavior of the CO_2 emissions resulting from on-site winter heating, and thus, quantify those emissions over the course of the project. Using the daily temperature variation generated, the consumption rate of propane gas can be quantified by a continuous-event simulation CES model, and thus the number of truck trips can also be determined, making it trivial to compute the amount of CO_2 emissions of winter heating.

From experts' judgement input, a 1000-Gallon truck full of propane takes approximately 30 minutes to unload. This means that the rate of fill is equal to (1000 Gal*24 hours/day*60 minutes/hour)/30 minutes = 4,800 Gal/day. 1 US gallon is equivalent to 3.78541 liters. The rate of fill in terms of Liters per day will be (4,800 Gal*3.78541 liters/Gal) per day = 18,169.968 Liters/day (300 liters of propane are equivalent to 79 Gallons). Figure 2 demonstrates the continuous model that represents the in-flow rate of filling propane gas into the tank, and the out-flow rate of the propane gas consumed in winter heating. Table 2 and Table 3 summarize the emissions resulting from material transportation based on a 40-Km travel distance and on-site equipment, respectively.The consumption rate of propane is directly dependent on the outside temperature. The emission rate resulting from winter heating is 62.7 Kilograms of CO_2/million BTU/hour (Mah 2007). It is observed from Figure 2 that propane gas consumption happens only if the outside temperature is below −5˚C.

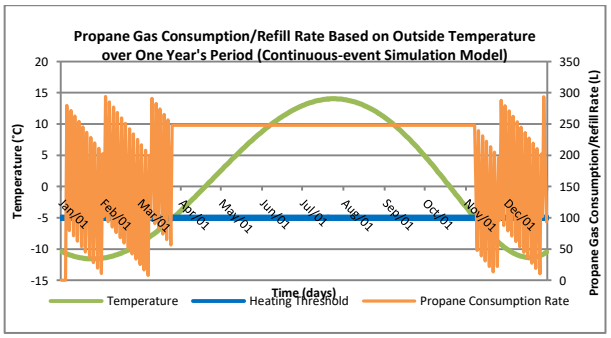


Figure 2: Weather and propane consumption data generated by the simulation engine

Based on the historical data analysis, activity durations in summer differ from those in winter. Accordingly, the results from the simulation model show a similar divergent behavior of project durations depending on whether the project commenced in summer, or in winter, or in between. The results show that the project duration varies between 133 days (starting in January) and 158 days (starting in December). In order to reduce the carbon footprint resulting from the construction process, the duration of the process itself should be reduced. However, before any embellishments have been considered, the base model was validated using the available historical data as described in the following section.

Table 1: Emissions resulting from material transportation (duration-independent) based on a travel distance of 40 Km

Vehicle Type	Emission Kg/Km	Total Emission Kg/ 1 Vehicle trip (40 Km)
van	0.23	9.2
0.5t	0.34	13.6
1.0t	0.7	28
3.0t	0.82	32.8
5t	1.16	46.4
concrete pump	1.16	46.4
concrete mixer	1.16	46.4

Table 2: Emissions resulting from on-site equipment (duration-dependent) per unit time

Equipment	kg/hr	Kg/day (8hr/day)
spreader	40	320
generator	2.68	21.44
excavator/backhoe	40	320
crane	16	128
compressor	2.68	21.44
Bobcat	28.63	229.04

5. BASE MODEL VALIDATION

Although simulation is a beneficial way for solving problems, the users are always concerned whether or not the outcome is correct. Thus, decision makers validate models in order to determine their accuracy (Sargent 2007). Schlesinger et al. (1979) defined model validation as "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model". According to the study by Balci and Sargent in 1982a, 1982b and 1984b, because of the availability of historical data, the best approach for validation is creating the simulation model deploying a sample from distributions of historical data. Validation by comparing simulated data to historical data is one powerful technique according to Sargent (2003). Consequently, before any output data analysis has been assessed, the base model has been validated through creating a simple validation model (consisting of the six main stages previously mentioned) and fitting historical data into task durations, and comparing the base model to the output results from the historical data being simulated (Figure 3).

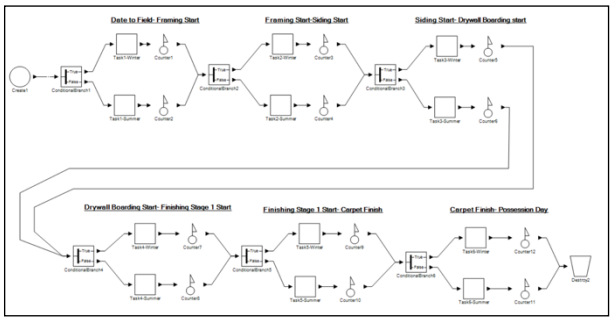


Figure 3. Validation through the simulation of fitted historical data

Fortunately, the project duration of the validation model was almost typical to the simulation model at the same time of the year. It took 140 to 157 days to finish the project according to the validation model. By comparing those results with the results from the base model simulation, it was observed that both models are very similar, thus, the base model can be considered as a reliable model.

6. RESULTS, DISCUSSION AND CONCLUSION

A basic simulation model is generated and integrated, which can compute three types of emissions accompanied with panelized single family house construction operation, including transportation, on-site mobile equipment, and winter heating emissions. This model is an abstraction of a real life problem, and in search of more accurate results, and thus further details and factors should be put into consideration in the future. Other emission types may be added to this model. In the modeling phase, comparing historical data to experts' judgement data, the research team found some time gaps, where no work has been done, or some sub-activities were not documented, or other factors impacted the work progress. It is of a great importance to quantify the impacts of uncertainty variables which affect the project schedule, and consequently the project cost (Wales and AbouRizk 1996).

6.1. Results

The output data was obtained from the simulation engine by running the simulation model several times starting at each month of the year to investigate different respective project emission scenarios. This section discusses the results and behavior of each of the three types of emission. As mentioned earlier, transportation emissions are fixed regardless of the project duration or weather conditions. Those emissions only depend on the travelling distance between the manufacturing plant and construction site, and amount of material to be hauled (i.e. number of trips). The traveling distance is, unfortunately, hard to change. It was observed from the results that the first stage of the construction process accounts for 49% of the total transportation emissions and decreases gradually as the project progresses, as shown in Figure 4. Emissions resulting from on-site equipment depend primarily on its hourly rate of emission and activity durations, specifically the generators and compressors. It was noticed from this study that emissions resulting from equipment are directly proportional to activity durations. It was also observed that the crane and the excavator are responsible for an outstanding amount of emissions compared with other equipment (Figure 5).

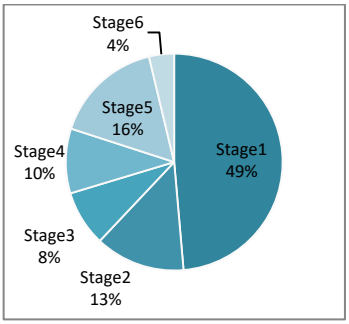


Figure 4. Transportation emissions at different stages of the project (duration-independent)

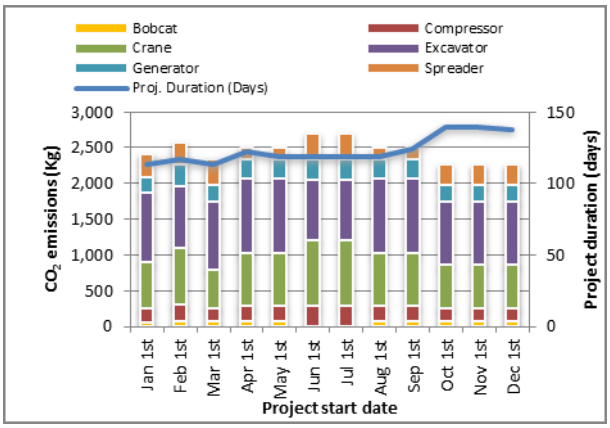


Figure 5. On-site equipment emissions based on project start date (duration-dependent)

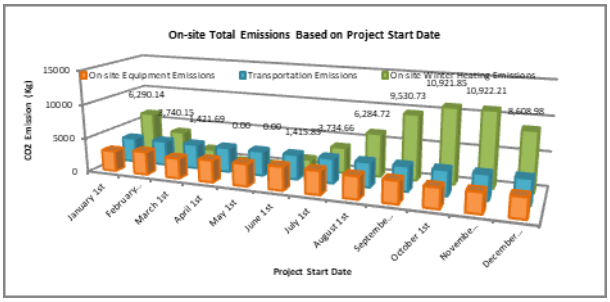


Figure 6. Different types of CO_2 emissions associated with different project start dates

From the simulation engine statistics collection analysis, winter heating was found to be biggest component of CO_2 emission throughout the construction process. It relies primarily on the time of the year at which the project started, and also on the overall project duration. A project starting between October and November accounts for an overall CO_2 emission of 11,000 kilograms, but if the same project started between April and May, the associated CO_2 emission will be reduced to zero kilograms (Figure 6).

6.2. Embellishments and Recommendations

The goal of a successful project is to perform the job in minimum duration (for cost saving) with maximum resource utilization (for maximum productivity) (AbouRizk 2011). From the ecological point of view, for the sake of reducing CO_2 emissions resulting from the construction process according to this study, it is advisable to either mitigate (if possible) on-site construction activities taking place between September and December, or reduce project durations by increasing the crew size at some specific activities. Some activities were observed to have significantly longer durations than others such as framing, siding, drywall boarding and taping, and painting. Those activities were triggered in this part of the study to reduce their respective durations. It is worth mentioning that this study is an abstraction of a real life problem, and that this assumption might not be feasible if other conditions (such as space confinement, resource availability, job quality, etc.) were put in consideration. It is assumed that doubling the crew size (labor-intensive) would reduce the corresponding activity duration by the half. It is also assumed that the transportation needs would remain the same, since the crew was originally assigned a large van for transportation.

As a result of the mentioned embellishment, the project duration has been reduced to a minimum of 113 days (starting in March) and a maximum of 140 days (starting in October). This embellishment accounts for the reduction of the construction process by 19 days, which is equivalent to 13% of the overall original project duration. Regarding the corresponding CO_2 emissions, 21.5% of the on-site equipment emissions can be reduced, 9.6% of winter heating emissions were also cut, and an overall reduction of 10% could be achieved due to this embellishment, as shown in Table 3. Regarding transportation emissions, the project duration does not imply any positive effect on it, as it was previously defined as duration-independent. A study made by Mah in 2007 at the University of Alberta states that switching vehicles from diesel to propane can reduce up to 30% of the CO_2 emissions produced during construction. Future studies may be conducted to investigate the impact of changing the fuel type of different vehicle types on their carbon footprint.

Table 3. Output analysis of the embellishment results

Type of Emission	Base Model	Embellishment	Reduction (Days)	Reduction (%)
On-site Equipment Emissions (Kg)	3131.9	2457.3	674.6	0.22
Transportation Emissions (Kg)	3602.4	3602.4	0	0
On-site Winter Heating Emissions (Kg)	5239.2	4734.8	504.4	0.1
Total Emissions (Kg)	11973.6	10794.5	1179.1	0.10
Proj. Duration (Days)	142.4	123.6	18.8	0.13

As a future recommendation, further studies should be conducted to investigate the possibility of the reduction of the carbon footprint resulting from the panelized construction method of single family houses in cold regions. The research team also recommends the collection of more accurate input data that can explain the significant gaps between actual and theoretical data. Triggering those gaps can give a better image of the possible solutions for reducing the whole project duration, thus reducing its carbon footprint. It is also advisable to conduct a future study which focuses on the optimization between winter heating cost (which accounts for maximum CO_2 emission) and the cost of reducing construction activities within winter months (September through December).

APPENDIX

A. Simulation Model

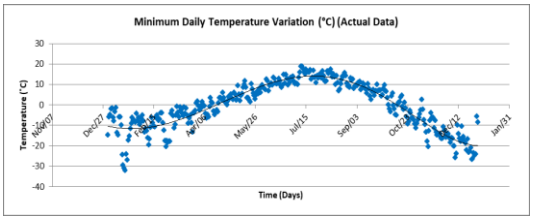


Figure A−1: Historical temperature data obtained from the city of Edmonton online database

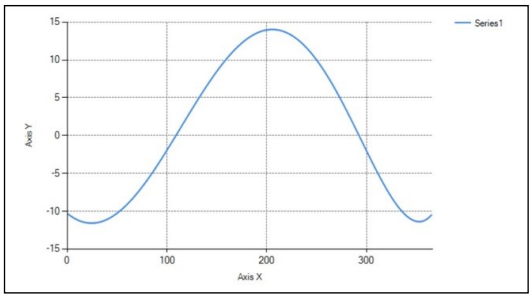


Figure A−2: Minimum daily temperature generated by the simulation engine using a 5th degree polynomial equation

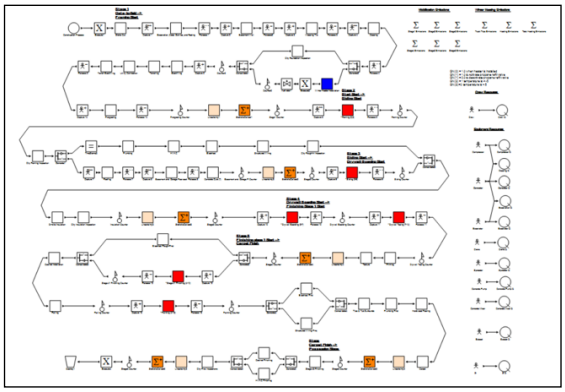


Figure A−3: DES model of the panelized construction process, duration in days

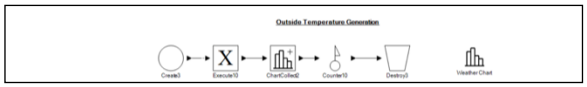


Figure A–4: DES model generating temperature variation over one year's period

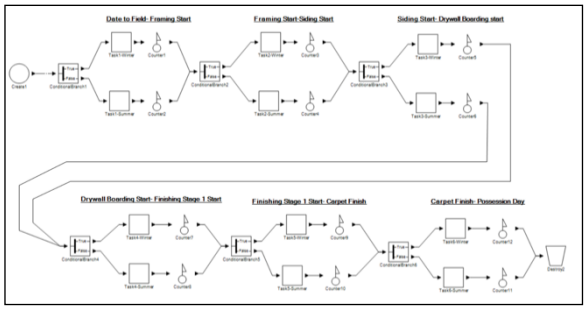


Figure A–5: Base model validation through simulation of historical data (summer and winter) through a discrete-event simulation model

B. Simulation Engine Charts

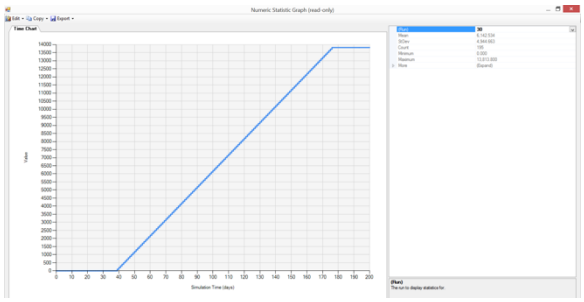


Figure B–1: Propane stock (liters) if construction started in October 1st

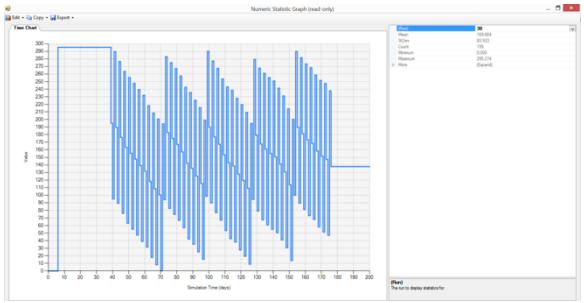


Figure B–2: Propane Tank consumption/refill rate if construction started in October 1st

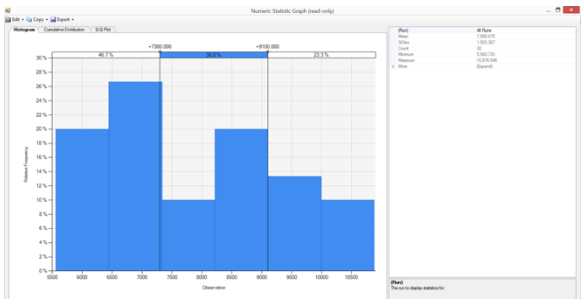


Figure B–3: Propane Emissions if project started on October 1st

C. CO_2 Emissions Model (Combined DES and CES)

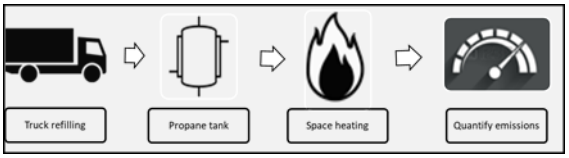


Figure C–1: Conceptual model of the CES simulation model

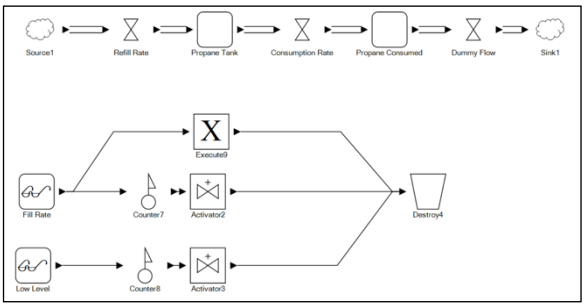


Figure C–2: CES model of the consumption/refill rate of propane tank

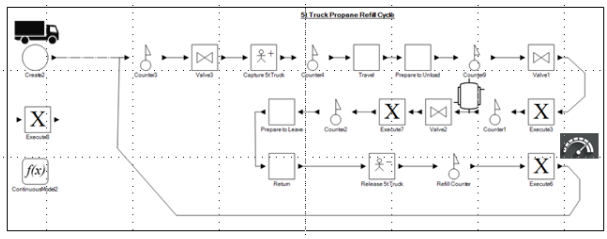
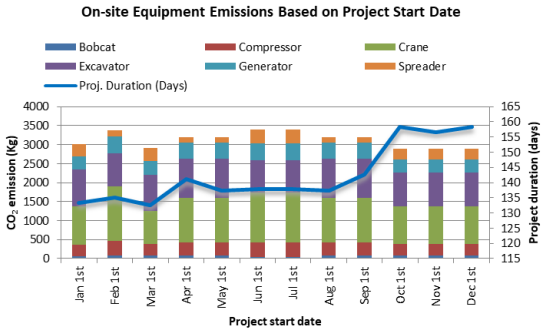
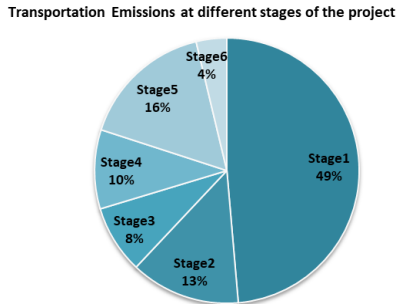
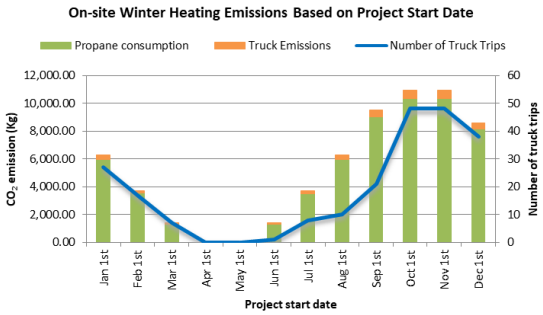


Figure C–3: DES model of the refilling truck cycle

D. Base Model Results

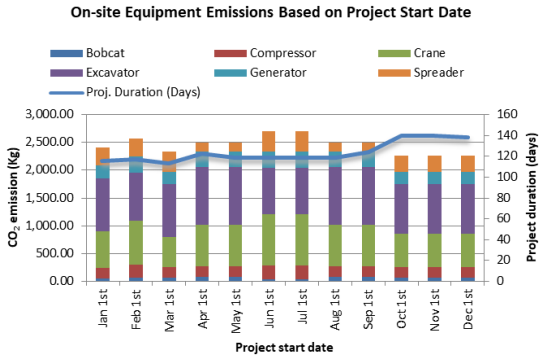


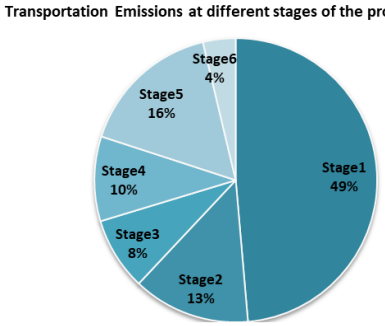


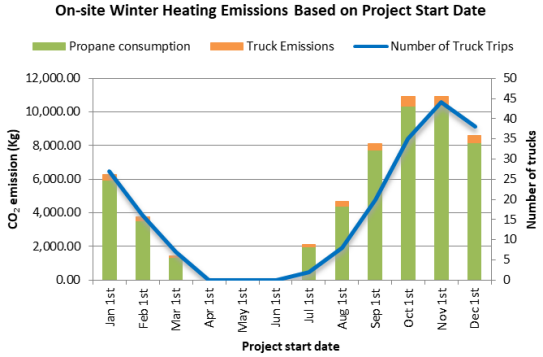


E. Embellished Model Results

(13% reduction in project duration, 21.5% reduction in equipment emissions, and 10% in winter heating emissions)







REFERENCES

AbouRizk S., Halpin D., Mohamed Y., Hermann U., (2011)," Research in Modeling and Simulation for Improving Construction Engineering Operations." Journal of Construction Engineering and Management, 843-852.

AbouRizk S., Halpin D., (1990), "Probabilistic Simulation Studies For Repetitive Construction Processes." Journal of Construction Engineering and Management, 575-594.

AbouRizk S., Halpin D., (1992), "Statistical Properties of Construction Duration Data." Journal of Construction Engineering and Management, 525-544.

Altaf M., Liu H., Al-Hussein M., (2015). “Online Simulation Modeling of Prefabricated Wall Panel Production Using RFID System.” Proceedings of the 2015 Winter Simulation Conference, pp. 3379-3390. December 6-9, Huntington Beach, CA, USA. https://doi.org/10.1109/WSC.2015.7408499

Balci O., Sargent R.G. (1982a). ”Validation of multivariate response simulation models by using Hotellig's two-sample T2 test”. Simulation 39(6):185-192.

Balci O., Sargent R.G. (1982b). “Some examples of simulation model validation using hypothesis testing”. Proceedings of the 1982 Winter Simulation Conference 620-629.

Balci O., Sargent R.G. (1984a). “A bibliography on the credibility assessment and validation of simulation and mathematical models”. Simuletter 15(3): 15-27.

Banks, Jerry, and John S. Carson, (1996), "Discrete-event System Simulation". 2nd ed. Upper Saddle River, N.J.: Prentice Hall.

Chini, A., Shrivastava, S. (2011). “Estimating Energy Consumption During Construction Of Buildings: A Contractor’s Perspective.”. Proceedings of the World Sustainable Building Conference, Helsinki, Finland, Oct. 18-21, pp. 3815-3824.

Dodoo A. (2011). Life Cycle Primary Energy Use and Carbon Emission of Residential Buildings. Doctoral Thesis. Östersund, Sweden","Ecotechnology and Environmental Science, Mid Sweden University.

Environment and Climate Change Canada (2019) Canadian Environmental Sustainability Indicators: Greenhouse gas emissions. Retrieved on June 30, 2019. Available at: www.canada.ca/en/environment-climate-change/services/environmentalindicators/greenhouse-gas-emissions.html

Esfahani, N.M., (2013). “Carbon Footprint Assessment of the Pre-panelized Construction Process”. Master dissertation, University of Alberta.

Friedman, A. (1992)." Prefabrication vs Conventional Construction in Single-Family Wood-Frame Housing". Building Research and Information, 2:4, 226-28.

Friedman A., Cammalleri V. (1992). "Prefabricated Wall Systems and the North American Home Building Industry". School of Architecture, McGill University.

Ginter Inc. (1991)."Comparative Evaluation of Factory Built House Construction Methods vis-avis the Traditional Construction Method on Site". Report

prepared for La Société d'habitation du Québec and the Canada Mortgage and Housing Corporation, Montreal.

Golabchi, A., Han, S., AbouRizk, S. (2018). “A simulation and visualization-based framework of labor efficiency and safety analysis for prevention through design and planning”, Automation in Construction, Volume 96, Pages 310-323. https://doi.org/10.1016/j.autcon.2018.10.001

Gonzalez, M.J. and Garcia, N.J. (2006). "Assessment of the decrease of CO2 emissions in the construction field through the selection of materials". Practical case study of three houses of low environmental impact. Building and Environment 4: 902-909.

Hajjar D., AbouRizk A. (1996), “Building a Special Purposes Simulation Tool for Earth Moving Operations”. Proceedings of the 1996 Winter Simulation Conference, 1313-1320.

Li Hong, Naseri Esfehani M., Gul M., Yu H., Mah D., Al-Hussein M., (2014), “Carbon Footprint of Panelized Construction: An Empirical and Comparative Study “. Construction Research Congress, 2014, 494-503.

Ismail, A., Srewil, Y. and Scherer, R.J. (2017). “Integrated and collaborative process-based simulation framework for construction project planning”, Int. J. Simulation and Process Modelling, Vol. 12, No. 1, pp.42–53. https://doi.org/10.1504/IJSPM.2017.082789

Kawecki L.R., (2010), “Environmental Performance of Modular Fabrication: Calculating the Carbon Footprint of Energy Used in the Construction of a Modular Home “. Doctoral dissertation, Arizona State University.

Mohsen, O., Abdollahnejad, S., Sajadfar, N., Mohamed, Y., AbouRizk, S. (2018). “Modeling and Simulation of Cabinet Manufacturing Processes: evaluation and recommended controls.” Proceedings of the 17th International Conference on Modelling and Applied Simulation (MAS), September 17-19, Budapest, Hungary.

Mohsen O.M., Knytl P.J., Abdulaal B., Olearczyk J., and Al-Hussein, M. (2008). “Simulation of modular building construction” Proceedings of the 2008 Winter Simulation Conference, Austin, TX, pp. 2471-2478. https://doi.org/10.1109/WSC.2008.4736356

Mah, D. (2007). “Analysis of Material Waste from the Framing Stage in Residential Construction Based on Landmark Homes Field Investigation”. Field Report, University of Alberta, Edmonton, AB, Canada.

Nance, R. E. (1995), "Simulation Pogromming Languages: An Abridged History“, Proceedings of the 1995 Winter Simulation Conference, X. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, eds., Arlington, VA, Dec. 13-16,pp. 1307-1313.

Naseri Esfahani M., (2013). “Carbon Footprint Assessment of the Pre-panelized Construction Process”. Master dissertation, University of Alberta.

Palaniappan S., (2009). ”Environmental performance of on-site construction processes in post-tensioned slab foundation construction”. A study of production home building in the greater Phoenix area. Doctoral dissertation, Arizona State University.

“Quality, Speed and Cost”, (2014), Retrieved April, 17, 2015 from http://ihhwc2014.ca/wp-content/uploads/2014/10/S18-Modular-and-Container-Housing-by-Kathleen-Maynard.pdf

RazaviAlavi S and AbouRizk S. (2017). "Site layout and construction plan optimization using an integrated genetic algorithm simulation framework." Journal of Computing in Civil Engineering, 31(4). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000653

Sargent R.G. (2003)." Verification and validation of simulation models" Proceedings of the 2003 Winter Simulation Conference. Chick S., Sinchez P.J., Ferrin D., Morrice D.J.,eds.

Schlesinger et al. (1979). “Terminology for model credibility”. Simulation, 32(3):103-104.

Suzuki M., Oka, T., and Okada, K. (1995). "The estimation of energy consumption and CO2 emission due to housing construction in Japan". Energy and Buildings 22: 165-169.

“The Carbon Dioxide Greenhouse Effect.” (2011). In the Discovery of Global Warming. Retrieved April, 5, 2015 from http://www.aip.org/history/climate/co2.htm

“The Greenhouse Effect.” (2012). In Canada’s Action on Climate Change. Retrieved April, 5, 2015 fromhttp://www.climatechange.gc.ca/default.asp?lang=En&n=F2DB1FBE-1, http://www.climatechange.gc.ca/default.asp?lang=En&n=1A0305D5-1, http://www.climatechange.gc.ca/default.asp?lang=En&n=21654B36-1

“The Panelized Process.” (2007) Retrieved April, 5, 2015 from http://www.canadiantimber.ca/process.html#

Upton B., Miner R., Spinney M., and Heath L. S. (2008). “The greenhouse gas and energy impacts of using wood instead of alternatives in residential construction in the United States”. Biomass and Bioenergy, 32(1), 1-10.

Wales, R., & AbouRizk, S. (1996). “An integrated simulation model for construction”. Simulation Practice and Theory, 401-420.

INTERNET OF THINGS AND INDUSTRIAL AUTOMATION APPROACH APPLICATION FOR BEEKEEPING PROCESSES AUTOMATION

Anatolijs Zabasta, Nadezda Kunicina, Kaspars Kondratjevs, Leonids Ribickis

The Institute of Electrical Engineering and Electronics, EEF
The Riga Technical University, Riga, Latvia

Anatolijs.Zabasta@rtu.lv, Nadezda.Kunicina@rtu.lv, Kaspars.Kondratjevs@rtu.lv,
Leonids.Ribickis@rtu.lv

ABSTRACT
In this paper we discuss the Internet of Things (IoT) and industrial systems automation approach application for development a prototype of an autonomous beekeeping system. In this research we focus on one of the Arrowhead Framework core systems named the Event Handler system. MQTT service broker being a part of autonomous beekeeping system demonstrates the services provided by the Event Handler System. The proposed services broker applies a visual data flow programming approach. A Node-RED is used to prove viability and advantages of offered architecture implemented in a local automation cloud of autonomous beekeeping System of Systems. The autonomous beekeeping system' prototype provides useful data on beehives and bee apiary status (internal and ambient temperature, humidity, weight of the hives, etc.) to the system users, so they can evaluate the hive status and take further action.

Keywords: Internet of Things, Arrowhead, UML, Wireless Sensor Networks, autonomous beekeeping.

1. INTRODUCTION

Beekeepers, who to engage intensive honey production, have to reallocate bee apiaries several times pursuing better conditions for harvesting of honey products by the end of a summer season. Therefore, they need such monitoring system that ensure monitoring of the hives conditions remotely, e.g. whether the weight of a beehive approaches to maximum, thus the harvesting must be arrange, otherwise, honey production will be interrupted. In winter time, a beekeeper wants to know, whether the inside temperature is critical, if the family is missing feed, to detect, and prevent dangerous deviations in time. In our research we offer a solution that performs bee apiary control without interfering with its processes, which helps to optimize frequency of the apiary inspection. The prototype helps to analyze monitoring data in correlation with video, meteodata, weight changes in time as well as interpretation of nest temperature, humidity and linking to local ambient conditions.

This research is implemented in a frame of the project "Autonomous Beekeeping", which is started at the beginning of 2018. The project is funded by the European Agricultural Fund for Rural Development Program 2014-2020 Cooperation: Support new products, methods, processes and technologies.

The paper is structured as follows. The Chapter 2 provides a review of the literature devoted to application of SOA for industrial systems automation and innovative approaches for automation of beekeeping maintenance processes. The model and composition of a prototype of the autonomous beekeeping system is described in the Chapter 3. Discussion about main discoveries and critical issues is done in the Chapter 4. The review of main contribution of the research and possible directions for future works are discussed in the Chapter 5.

2. REVIEW OF THE LITERATURE

Several researches have been devoted to monitoring of bee colonies. One of them is a project "The Application of Information Technologies in Precision Apiculture" (ITAPIC 2016), which proposed implementation of precision agriculture technologies and methods in the beekeeping. Zacepins, Stalidzans, and Meitalovs (2012) defined the notion of precision beekeeping as an apiary management strategy based on monitoring of individual bee colonies to minimize the resource consumption and maximize the productivity of bees. Altun (2012), Zacepins et al. (2015) depicted a few examples of approaches and solutions: temperature and humidity controlling system, which is powered by solar energy. IoT system prototype for honey bee colony temperature monitoring is described by Zacepins, Kviesis, Pecka, and Osadcuks (2017).

Striving to apply IoT and industrial automation approaches to beekeeping monitoring processes automation we developed a system, which is based on Service Oriented Architecture (SOA) (Alessandrelli, Petraccay, and Pagano 2013; Karnouskos S. et al. 2010), in order to provide highly compatibility with existing and emerging solutions. Gross (2005) developed a model with the purpose of flexible composition and reuse of software artifacts. The method uses UML (Norris 2004) as primary model-based notation for analysis and design

activities. The model applies a notation System or a System-of-Systems (Maier M. W. 1998) as a component that can interact with others through interfaces and can be decomposed in other Systems or components.

Blomstedtet et al (2014) presented the first steps of realizing the Arrowhead vision for interoperable services, systems and systems-of-systems aiming to support the documentation of their structural services. Each service, system and system-of-systems within the Arrowhead Framework must be documented and described in such way that it can be implemented, tested and deployed in an interoperable way. The research (Varga 2016, Delsing 2017) goes beyond Blomstedtet et al (2014), therefore, interoperability in-between almost any service provided by heterogeneous systems are here addressed by the core services that are necessary to meet the requirements and enable a collaborative automation cloud. Both papers present an overview of the framework together with its core elements and provides guidelines for the design and deployment of interoperable, Arrowhead-compliant cooperative systems.

The local cloud concept developed by the Arrowhead Framework (Delsing ed. 2016) addresses many challenges related to IoT-based automation, and is unique in its support for integration of applications between secure localized clouds. The viability of Arrowhead Framework has been demonstrated by implementing of this approach in diverging fields of industrial automation. The case study of small enterprise (Lindström 2018) concerns a multi-usable cloud service platform for big data collection and analytics.

A case study of the service broker, implemented for control of utilities systems in urban environment, is presented in (Zabasta et al. 2018). Arrowhead core Event Handler service s created and tested as a MQTT enabled service broker, for wiring together divergent hardware devices and nodes, and APIs for online services.

A macro-programming model capable to oversee the network of distributed sensors as a whole rather than to program individual pieces is suggested by Newton, Arvind, and Welsh (2005). Giang (2017) and Blackstock, M. and Lea, R. (2014) offered an advanced version of dataflow model to express application logic of IoT devices suitable for large scale IoT, called as "Adaptive Distributed Dataflow".

In our research we studied Data Flow Systems, which provide a data flow-programming paradigm for IoT applications. The WoTKit Processor (Blackstock, and Lea 2012) provides the basic requirements for lightweight toolkits: integration, visualization and processing components and a RESTful API. NoFlo (NoFlo 2019) is a flow based programming environment for JavaScript. It is targeted to ease the development of web applications. Like Node-RED, NoFlo provides a visual editor for a data flow runtime leveraging Node.js. Unlike FRED, NoFlo is not a cloud service; developers install and deploy NoFlo themselves, on its own, or embedded in an application. Glue.Things Composer (Kleinfeld, R. et al. 2014) is designed to connect to other components of the GlueThings platform: Device Manager, and Deployment Manager. A mashup editor is built on Node-RED. Blockly (Blockly 2019) is a client-side JavaScript library for creating visual block programming languages and editors. It is a project of Google and is open-source under the Apache 2.0 License.

Since the IoT has been recognized as a promising solution since in very divergent areas such as public utility systems automation (Kunickis, Dandens, Bariss 2015), efficient transportation systems (Alps I. 2016) centralized healthcare (Sultanovs, Skorobogatjko, and Romanovs 2016), efficient waste management, smart grids, digital tourism, etc., we tested it applicability in an agriculture area, namely in the beekeeping.

In this research we focus on one of the Arrowhead Framework core systems named the Event Handler system. MQTT service broker being a part of autonomous beekeeping system demonstrates the services provided by the Event Handler System.

The proposed smart services broker applies a visual data flow programming approach (Blackstock, and Lea, 2014). A Node-RED (Lewis 2016) was selected among other tools to prove viability and advantages of offered architecture implemented in a local automation cloud of autonomous beekeeping System of Systems.

3. DEVELOPMENT OF THE MODEL OF THE SYSTEM

3.1. Arrowhead Framework Approach for the Autonomous beekeeping

For creation of Autonomous beekeeping system (AB system), we applied the Arrowhead Framework approach that supports the development of generic SOA systems. The Arrowhead Framework (AF) acts as an enabler for systems from different areas: industrial automation, energy production, home automation, smart grids, etc. to facilitate their interaction with each other and exchange information. This multi-area approach can enable considerable savings in terms of efficiency, interoperability and maintenance cost. AF approach promotes the different application systems in an easy and flexible way being able to collaborate successfully due to support provided by the common core services.

The Arrowhead Framework (AF) includes a set of Core Services, which support the interaction between Application Services (e.g. sensor readings, controlling of actuating devices, energy consumption, temperature measurement services, etc.). The Core Services handle the support functionality within the AF to enable application services to exchange information (see Fig.1). The AF is built upon the local cloud concept, where local automation tasks should be encapsulated and protected from outside interference.

Each local cloud must contain at least the three mandatory core systems: Discovery, Authorization and Orchestration, which allow to establish connections between Arrowhead application services (Delsing ed. 2016).

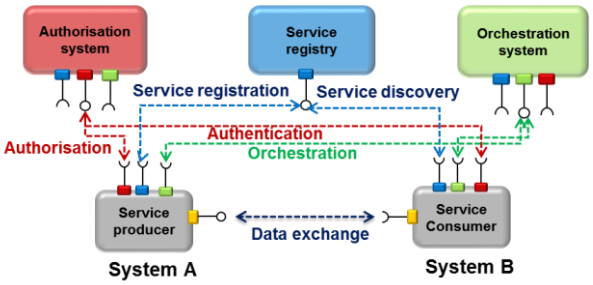


Figure 1: Data exchange between a service producer system and a service consumer system (Delsing 2015).

3.2. Event Handler System and its services

The core system, named as Event Handler System (Albano, Ferreira, and Sousa 2016), belongs to the automation supporting core systems. It provides functionality for the handling of events that occur in an Arrowhead network. The Event Handler System (EHS) receives events from Event Producers and dispatches them to registered Event Consumer. A notation "event" could imply an application service, for example, data provided by water flow sensor, or another event, such as a log of the service failure in case, when EHS is used as a part of the quality of service evaluation system.

There are scenarios, which require a stronger level of decoupling, in terms of space, time and synchronization, when Arrowhead Core Services are used to ensure exchange of services between the systems.

Space decoupling means that a publisher and a subscriber do not need to be aware of each other's location or identities. Time decoupling means that a publisher and a subscriber do not need to be online and actively collaborating in the interaction at the same time. Synchronization decoupling allows asynchronous notification of subscribers by using event services callbacks.

For such scenarios Event Handler System acts as an intermediary between the event producer and the event consumer, providing asynchronous and one-to-many or many-to-many communication model (Eugster et al. 2003). Filtering rules to incoming events could be applied, based on the predefined criteria, for example, on the subscription to the particular services provided by the EHS. In such case, only events, which are interested for Event Consumer, will be sent.

The Event Handler System provides three services – the Registry Service, Publish Service and Notify Service. The SOA approach is implementing due to operations performed in the context of the of the Publish/Subscribe paradigm (see Fig. 2).

The EHS Registry Service is provided in order to store and keep track of the service consumers and service producers in the System of System. If a consumer desires to receive services, or a producer wants to publish services, both need to be registered through this service. At the registration time the producer should advertise the kind of services it produces. The consumer has to specify the filtering rules regarding incoming events by defining a set of conditions to be applied to all incoming events, to be routed to the subscriber.

The EHS Publish Service and Notify Service are used to deliver data regarding the events. A producer system accesses the Publish service of EHS to post the events it produces. The EHS defines, which consumers must receive the event, and then launches the Notify service to provide the incoming event to a particular subscriber of the service.

One of the functions implemented by the EHS is the storage of information regarding events (services) for future access. The storage of information could be implemented locally at the data base of the Event Handler System, or through a Historian Service, which is one of the Arrowhead cores services used to store data (Pereira et al. 2014). To perform this function, EHS proceeds events and routs them on a storage area together with information about a subscriber, which received this event and its meta data. The EHS Get Historical Data service applies filtering rules to permanently stored events (in a database, log file or through the Historian) and retrieves data regarding stored events.

When talking about Quality of Service (QoS) there are always two different parts of delivering an event: a publisher to EHS and a subscriber. It is worth to look at them separately, since there are subtle differences. The QoS level for a publisher to EHS is depending on the QoS level the publisher sets for the particular message. When the EHS transfers an event to a subscriber, it uses the QoS of the subscription made by the subscriber earlier. That means, the QoS guarantees can get downgraded for a particular client, if it subscribed with a lower QoS.

In work (HiveMQ 2015), three Quality of Service levels in MQTT are depicted. Since in this research we developed a MQTT service broker, it looks reasonable to interpret QoS notation defined in (HiveMQ 2015) aiming to depict QoS of the Event Handler System.

QoS 0 – at most once. The minimal level is zero and it guarantees a best effort delivery. A message won't be acknowledged by the receiver or stored and redelivered by the sender. This is often called "fire and forget" and provides the same guarantee as the underlying TCP protocol.

QoS 1 – at least once. When using QoS level 1, it is guaranteed that a message will be delivered at least once to the receiver. But the message can also be delivered more than once. The sender will store the message until it gets an acknowledgement from the receiver.

QoS 2. The highest QoS is 2, it guarantees that each message is received only once by the counterpart. It is the safest and also the slowest quality of service level. The guarantee is provided by two flows there and back between sender and receiver.

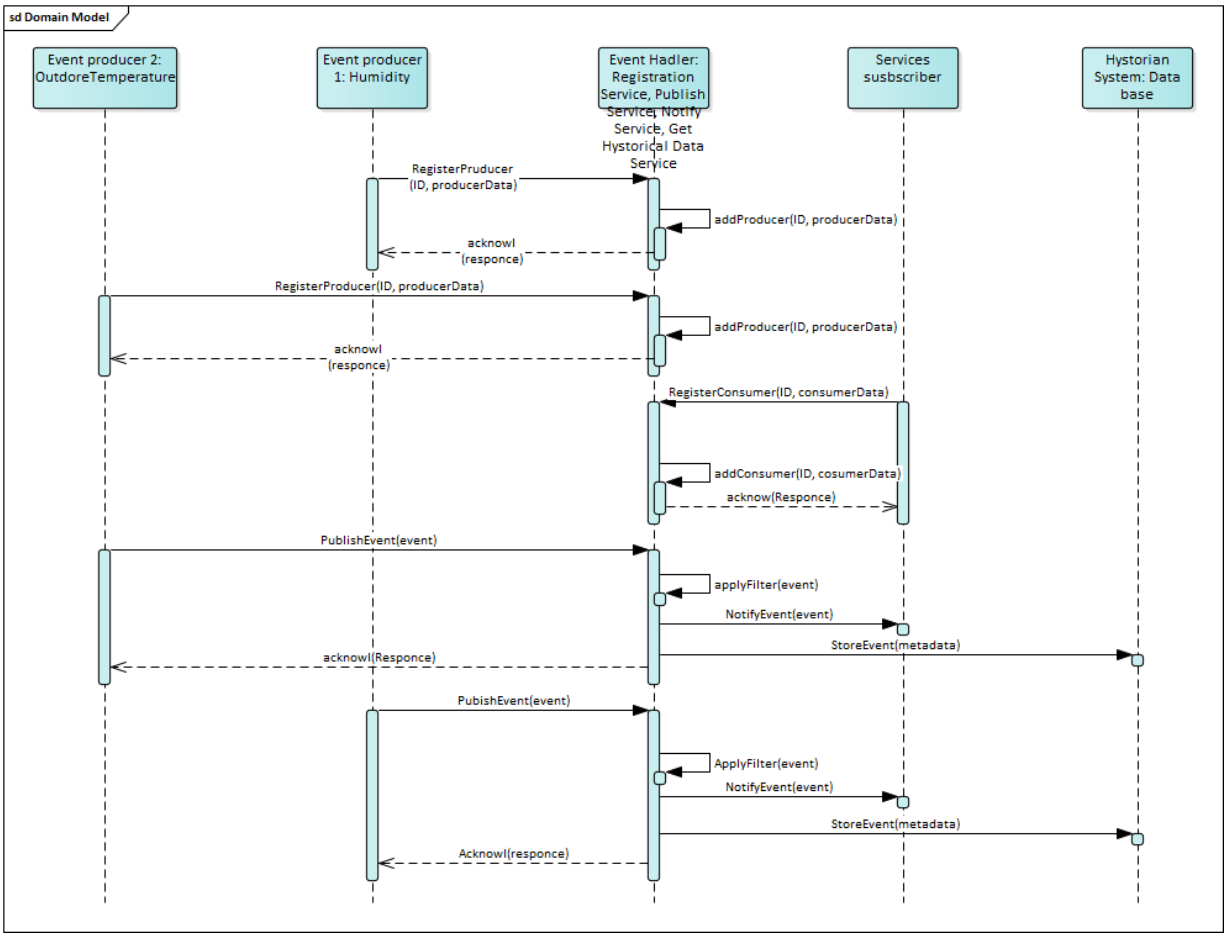


Figure 2: Event Handler System interaction with the systems.

The Event Handler System can handle a variety of communication protocols applied by IoT embedded devices, such as MQTT (Message Queuing Telemetry Transport) (OASIS Standard 2014; ISO 2016; Shelby, Hartke, and Bormann 2014), Constrained Application Protocol (CoAP) and REST (Representational state transfer), and can decode commonly used semantics, e.g. SenML (Sensor Markup Language), XML, JSON (JavaScript Object Notation), and plain text. EHS should be able to convert between protocols in a message exchange between different users. For example, a device can push data to the Event Handler System CoAP or MQTT, while clients can either use MQTT, or poll data using HTTP.

In our research we developed Event Handler System as a service broker, which enables SOA based services and data flow between divergent type of embedded devices and nodes, for example: humidity, wind strength, weight, outdoors and indoor temperature sensors, etc. The implementation of EHS core system as a MQTT service broker is depicted in the Chapter 3.5.

3.3. Autonomous beekeeping System functionality and composition

The first prototype of the autonomous beekeeping (AB) system was tested at five of bee hives allocated at the Riga Botanic Garden during the summer 2018. The weight of bee hives, humidity as well as the temperature sensors, which measure the temperature inside of the beehives and outside temperature, were tested (see Fig.3).

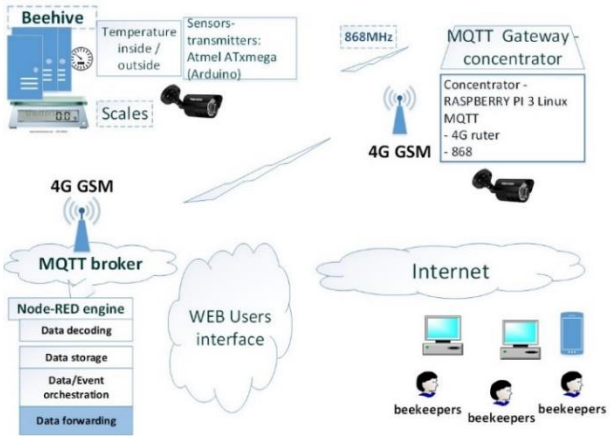


Figure 3: A bock-scheme of autonomous beekeeping system.

A communication system of AB System-of-Systems comprises sensors-transmitters and gateway-concentrators as the elements of wireless sensor network (WSN). The sensors-transmitter consists of a microcontroller, readout interface, power supply and ISM band radio module operating at 868MHz.

The data from the sensor-transmitter are transmitted to the concentrator from sensors using ISM range signal 868 MHz (for example, weight sensor 1C3003AE0030 transmitted data via radio gateway-concentrator).
The transmission range of sensors-transmitters is up to several hundred meters depending on the installation environment. The measurement data are encapsulated in telegrams to be transmitted using Manchester coding and GFSK modulation to one or to multiple gateways that provide an area of coverage.
The gateway node consists or a radio module, GSM GPRS module. Additional or 802.11b/g/n Wi-Fi module that integrates a microcontroller (ESP8266) is available. The Wi-Fi module supports both station: Wi-Fi client mode and access point modes.
A 4G reliable and secure LTE router with I/O, GNSS and RS232/RS485 has been used for communication between gateway-concentrator and MQTT broker at the back-end. Router delivers mission-critical cellular communication and GPS location capabilities.

3.4. The system UML model

In order to create a model of the autonomous beekeeping system we use a methodology represented by OMG's (Object Management Group) solution for system abstraction, modelling, development, and reuse—Model Driven Architecture (MDA) (OMG 2005). The key component of a system modelling, which underlies the principles of MDA, is the Unified Modelling Language (UML) a widely accepted standard for modelling and designing different types of systems. In order to create the UML model of AB system we apply Enterprise Architect Professional modelling tool.

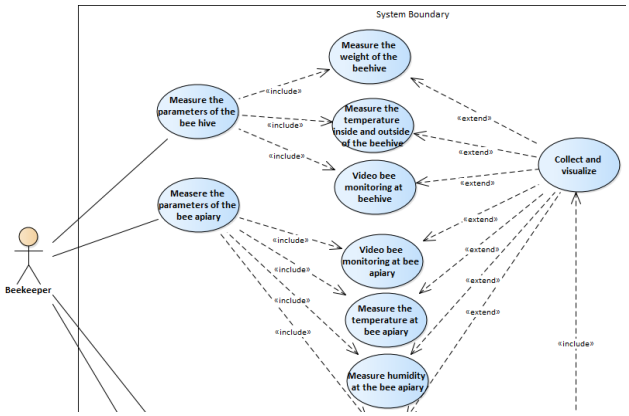


Figure 4: Use case diagram of AB system

Fifteen primary Use cases have been defined aiming to describe AB model (see Fig.4). Among them are Collect and visualize parameters, Measure beehive weight, Measure beehive's internal temperature, Measure the beehive's internal moisture, Monitor energy consumption, etc. We use the actor symbol to represent the agent that activates the use case, probably beekeeper or the other AB system' client, which is willing to analyze the behavior of a bee colony.

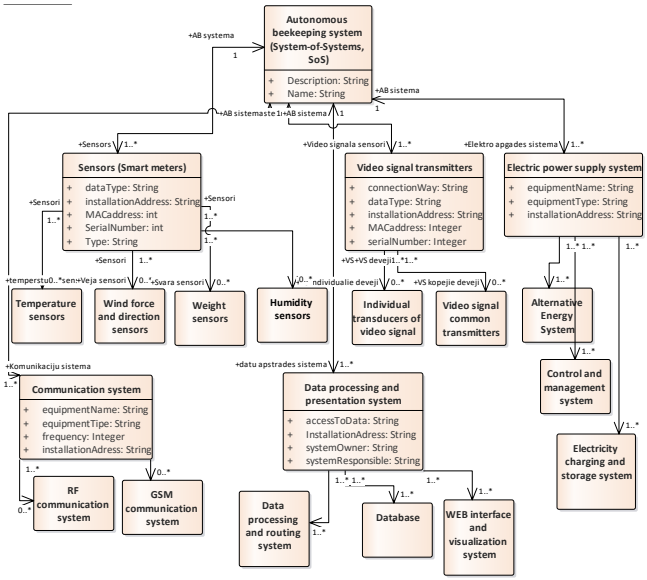


Figure 5: Domain Objects diagram of autonomous beekeeping system

A Domain Objects diagram at Fig. 5 shows nineteen classes, where four of them represent different kinds of sensors with its services components, four classes represent power supply and energy storage, four classes data storage processing and visualization services, two classes data communication services, and etc. The parameters of attributes and operations in classes have been omitted in the interest of figure clarity.
Data processing and presentation system will be explored and described in more details as a MQTT service broker in the chapter 3.5.

3.5. Event Handler System implementation

We implemented Event Handler System as a MQTT service broker that applies a software integration platform Node-RED to interconnect heterogeneous systems in IoT way. The service broker applies Message Queuing Telemetry Transport (MQTT) protocol, which is a Client Server subscribe messaging transport protocol. It is lightweight, open, simple, and designed so as to be easy to implement. These characteristics make it suitable for constrained environments such as for communication in Machine-to-Machine (M2M) and Internet of Things (IoT) contexts.
The beehive monitoring system deployed a Node-RED (Lewis 2016) as a gateway concentrator for wiring together hardware devices (temperature, weight, humidity sensors, etc.), APIs and online services. At the heart of Node-RED is a visual editor allowing complex data flows to be wired together with a little coding skills. Node-RED main functionality is to decode and to route MQTT smart metering data to further service orchestration or use them in external services as monitoring system or external clients (see Fig. 3 and 6).

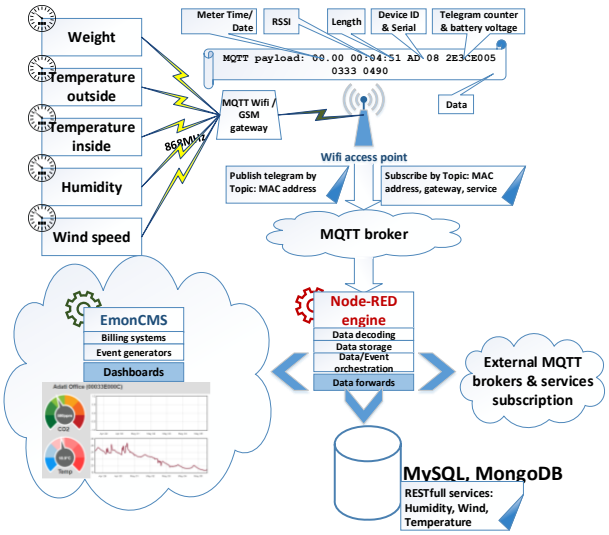


Figure 6: The Scheme of MQTT Service Broker

The services broker depicted in this paper ensures QoS level "0" that is sufficient for the nature of services consumed by the clients of Autonomous Beekeeping system.

Fifure: 7. Outdoor weight scale and sensors installation.

The temperature sensor-transmitters transfer data to the server-broker every 15 minutes, the weight sensor transmits data to the server every 5 minutes. The data collected by the MQTT broker are displayed in graphical form. The users interpret data in graphics or as "dashboard", based on their own experience and understanding of ongoing processes in the hive.

For monitoring of the temperature two sensors are allocated inside of a beehive: one in the center and the second one in the side part of a beehive The third sensor is allocated at the outside wall of the beehive to get measurements of ambient air temperature. The weight of each hive is measured by a specially designed weight platform (Fig. 7). Measurement data are sent by a sensor-transmitter to the gateway-concentrator without additional processing.

Beekeepers and the other project partners are interested to compare behavior of bee colonies and particularly, at the transitions of the seasons, when bees start awaking before to leave hives after the winter time. A multi-chart view provides an opportunity of comparison of bee colonies (see Fig. 8).

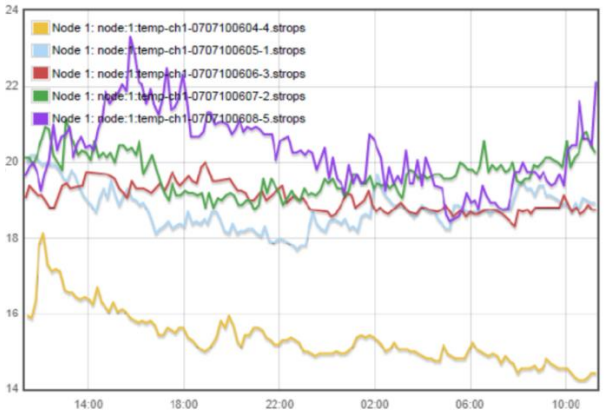


Figure: 8. Multi-chart view of temperature measurement inside of 5 beehives

Node-RED ensure decoding and to routing MQTT smart metering data (Fig. 9) to further service orchestration or/ and for use in external services as customer billing or monitoring systems.

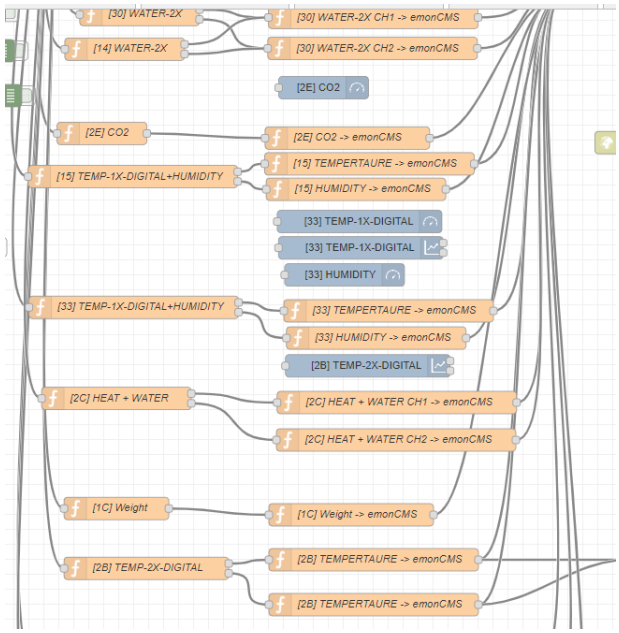


Figure: 9. The nodeRED workflow for wireing of different devices and data processing.

An Inject node (is not shown due to the lack of space) is a testing element to generate a simulated payload for debugging purposes. Below is an example of a sample payload for a sensor with weight data:

MQTT payload: `00:04:40 42 0B 1C3003AE0030 0337 002580`, where 1C3003AE0030 is a sensor ID, RSSI LENSN03, 37 (battery level) and 002580 is data read out.

The task is to decode the payload and forward it to a data storage and visualization service using the IoT approach. The first step is to define a MQTT source, which is a gateway.

Gateways post received metering data to topic based on their MAC address that also serves as a configuration service by subscribing to MAC configuration. The node "SERIAL+VALUE" creates an array of elements from the initial payload (telegram); each array element is processed separately depending on the needed of output or post-processing. A new object is created by dividing name/value pairs, where the crucial elements are: serial number – that identifies the data type and coding format and the unique device id, which makes the metering devices distinct. The second value is the data block needed for decoding procedures. Double output creates two new MQTT messages containing separated data from each sensor, with measurement type topic and devices unique serial number.

For further processing (e.g. monitoring, for serving of external clients) emoncms (Open Energy Monitor) code base is used (node Wight->emonCMS). From the example above a type+device_id object name is generated. After data delivery to emoncms, these data objects are discovered as inputs. Inputs are generic variable that can be processed using emoncms processing engine.

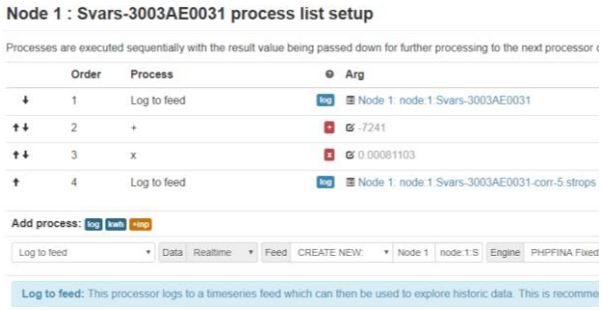


Figure: 9. An algorithm of the block for Query Emoncms API

Before saving information for analysis and representation, pre-calculation of sensors measurement offset values should be done in order to calculate the actual sensor readout value. At Fig.9 an example of "log to feed" value pre-calculation for a *weight* (*Svars* in Latvian) sensor is depicted. Users can customize data transformation, delta offsets, calibration, scaling etc. Typically, the broker does not save MQTT data. This can be done by defining a flow to a data storage, like MySQL DB. In this application, data storage is defined in emoncms with the "Log to feed" processor.

RESTful services can be implemented on demand from the MySQL DB via EmonCMS API by encapsulating the API request into a restful call or by direct definition of a request in form of a RESTful HTTP request. An example of returned value from a request for the latest data of the weight measurement sensor. A request from 03.05.2019 to 05.05.2019 at an interval of 5 minutes from the feed ID=798. GET: http://broker.ventspils.lv:9990/emoncms/feed/data.json?id=798&start=1556443800000&end=1557049500000&interval=900&skipmissing=1&limitinterval=1

```
1556443800000,44938,[1556444700000,44914.666666667],[1556445600000,44987.666666667,….
```

4. DISCUSSION OF THE RESULTS

In our research we applied the Arrowhead Framework approach to develop and document a small SoS using industrial automation and IoT methods. This real SoS operates in a local automation cloud, which comprises core systems and application systems services (temperature, humidity etc.). A core system, namely the Event Handler System, provides several core services: Registry, Publish, Notify and Get Historical Data services.

The use case of autonomous beekeeping system demonstrates, how a MQTT service broker implements the functions and services of EHS. Among the others it provides opportunity to different stakeholders to "subscribe" to the services taking into account a define QsS.

In the future research we plan to add a new SoS to the existing automation cloud, therefore a new automated beekeeping system, which belongs to a different stakeholder, to be incorporated.

We are going to apply advantages of the AF in order to enlarge the automation cloud without significant reconfiguration, which would require much time and efforts. Reconfiguration process should facilitate a correct evolution of the SoSs by updating documentation, systems interfaces, while minimizing the required changes. This could be done thank to:

- Formal description of the SoS structure in UML – SysML.
- Possibility to map the elements of the documentation to engineering procedures.
- Feasibility and benefits of AF approach, which are demonstrated at real-life use cases (see chapter 2).

A concern about security issue should be investigated in the future research, since several stakeholder systems will operate in a single automation cloud.

The research provided in AB project is aware of achievements of the FP7 Framework project ITAPIC, which was devoted to "precision beekeeping" topic. Unlike ITAPIC, the AB project:

- Applies new, more efficient wireless technology and IoT solutions:
- It focusses on development of an integrated / multifunctional (mass, temperature, humidity, meteorological data measurement, video data) system, aiming at application in today's practical beekeeping.
- The system allows remote monitoring not only at individual bee colonies, but also at apiary level.
- We test the AB system at real production conditions, under the supervision of highly qualified and professional beekeepers.

- Recommendations for beekeepers how to interpret equipment data measurements will be elaborated.

The restrictions of the research derive from requirements of the European Agricultural Fund for Rural Development Program. According to the Program, the cost of the prototype must be affordable for potential clients in Latvia, which are represented by farmers, who maintain usually 50-350 beehives. The AB system prototype is developed as a compromise between cost and quality of the system.

5. CONCLUSIONS AND NEXT STEPS

In our research we developed Event Handler System (EHS) as a service broker, which enables SOA based services and data flow between divergent type of embedded devices and nodes, such as outdoor and indoor temperature, humidity, weigh monitoring and humidity sensors. We applied the previous experience, when the EHS provided services have been used for monitoring and control of Smart City systems and utility networks (water supply, district heating, etc.) (Zabasta 2018). In our research, the service broker demonstrates how Arrowhead Event Handler system can be implemented for processes automation at rather different field such as intensive beekeeping.

By now the Event Handler system' working prototype is located and maintained at the "Ventspils Digital Center" (VDC) servers cloud, which belongs to Ventspils City Council as an institution responsible for development and maintenance of ICT infrastructure in the Ventspils city, Latvia. It is planned to migrate ICT infrastructure of AB system from VDC to one of public hosting providers in order to ensure sustainability of AB system beyond the project.

Development of autonomous beekeeping system' prototype is still in progress. The research by the end of the project should be focused on the opportunity to apply alternative energy sources, such as wind and photovoltaic elements, as well on security issues.

ACKNOWLEDGEMENT

The article is based upon the project "Autonomous Beekeeping" funded by the European Agricultural Fund for Rural Development Program, 2014-2020, Cooperation: Support new products, methods, processes and technologies.

REFERENCES

Albano M., Ferreira L. L., and Sousa J., 2016. Extending publish/subscribe mechanisms to SOA applications. 2016 IEEE World Conference on Factory Communication Systems (WFCS), P. 1-4.

Alps I., Gorobecs M., Beinarovica A., Levcenkovs A., 2016. Immune Algorithm and Intelligent Devices for Schedule Overlap Prevention in Electric Transport. No: 2016 57th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Latvia, Riga, 13.-14. October, 2016. Riga: IEEE, 2016, pp. 1-7.

Altun A. A. 2012. Remote Control of the Temperature-Humidity and Climate in the Beehives with Solar Powered Thermoelectric System, J. Control Eng. Appl. Informatics, vol. 14, no. 1, pp. 93–99, 2012.

Alessandrelli D., Petraccay M., and Pagano P., 2013. T-Res: Enabling reconfigurable in-network processing in IoT-based WSNs. In Proceedings - IEEE International Conference on Distributed Computing in Sensor Systems, DCoSS 2013, pages 337-344.

Blackstock M., and Lea R., 2012. WoTKit: a lightweight toolkit for the web of things, in Proceedings of the Third International Workshop on the Web of Things. ACM, 2012, p. 3.

Blackstock, M. and Lea, R. 2014. Toward a Distributed Data Flow Platform for the Web of Things (Distributed Node-RED). Proceedings of the 5th International Workshop on Web of Things (New York, NY, USA, 2014), pp. 34–39.

Blockly, 2019. Google Blockly Homepage: https://developers.google.com/blockly/. [Accessed: March 2019].

Blomstedt, F., Ferreira, L., Klisics, M., Chrysoulas, C., de Soria, I., Zabašta, A., Moris, B., Eliasson, E. Johansson M., Varga P., 2014. The Arrowhead Approach for SOA Application Development and Documentation. In: Proceeding 40th Annual Conference of the IEEE Industrial Electronics Society (IECON 2014), United States of America, Dallas, 29 Oct-1 Nov., 2014. Dallas: The Institute of Electrical and Electronics Engineers (IEEE), 2014, pp.2637-2637.

Delsing J., ed., 2016. IoT based Automation - made possible by Arrowhead Framework. CRC Press, Taylor & Francis Group.

Delsing J., 2017. Local Cloud Internet of Things Automation: Technology and Business Model Features of Distributed Internet of Things Automation Solutions 2017 In: IEEE Industrial Electronics Magazine, Vol. 11, no 4, p. 8-21 Article in journal (Refereed).

Eugster P. T., et al, 2003. Felber P., Guerraoui R., Kermarrec A-M., 2003. The Many Faces of Publish/Subscribe, ACM Computing Surveys 35(2), 2003, pp. 114-131

Giang N., Lea R., Blackstock M., Leung V.C. M., 2016. On Building Smart City IoT Applications: a Coordination-based Perspective, in Proceeding SmartCities '16, December 12-16, 2016, Trento, Italy pages 1-6.

Gross H-G., 2005. Component-Based Software Testing with UML, Springer-Verlag Berlin Heidelberg, ISBN 3-540-26733-6.

HiveMQ, 2015. MQTT Essentials Part 6: Quality of Service 0, 1 & 2, Written by The HiveMQ Team, published: February 16, 2015, [accessed in April 2019].

ISO 2016. MQTT 3.1.1 Specifications. International Organization for Standardization, ISO/IEC 20922. https://www.iso.org/standard/69466.html. [Accessed: April 2019].

ITAPIC, 2016. The Application of Information Technologies in Precision Apiculture (ITApic) ERA-NET ICT-Agri Project, http://www.itapic.eu/index.php [accessed in April 2019].

Karnouskos S. et al., 2010. Karnouskos S., Colombo A.W., Jammes F., Delsing J., Bangemann T., 2010 Towards an architecture for service-oriented process monitoring and control, in: 36th Annual Conference of the IEEE Industrial Electronics Society (IECON-2010), Phoenix.

Kleinfeld, R. et al. 2014. Glue.Things: A Mashup Platform for Wiring the Internet of Things with the Internet of Services. Proceedings of the 5th International Workshop on Web of Things New York, NY, USA, 2014, pp.16–21.

Kunickis M., Dandens A., Bariss U., 2015. Justification of the Utility of Introducing Smart Meters in Latvia. Latvian Journal of Physics and Technical Sciences. Volume 52, Issue 6, 1 December 2015, Pages 13-21.

Lewis K., 2016. Node-RED visual programming for the Internet of Things (IoT) is now a JS Foundation Project. https://www.ibm.com/blogs/internet-of-things/open-source-iot/ [Accessed in March 2019].

Lindström J., Hermanson A., Blomstedt F., and Kyösti P., 2018. A Multi-Usable Cloud Service Platform: A Case Study on Improved Development Pace and Efficiency. In: Applied Sciences, 2018, 8(2), 316, Vol. 8, no 2, https://doi.org/10.3390/app8020316.

Norris D., 2004. Communicating Complex Architectures with UML and the Rational ADS, In Proceedings of the IBM Rational Software Development User Conference.

Maier M. W., 1998. Architecting Principles for Systems-of-Systems, In Systems Engineering, volume 1, issue 4: pp. 267-284.

Newton R., Arvind, and Welsh M., 2005. Building up to macroprogramming: An intermediate language for sensor networks. In: 4th International Symposium on Information Processing in Sensor Networks, IPSN 2005, volume 2005, pages 37-44.

NoFlo, 2019. Flow-Based Programming for JavaScript NoFlo: http://noflojs.org/. [Accessed: March 2019].

OASIS Standard, 2014. MQTT Version 3.1.1, http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.pdf. [Accessed: March 2019].

OMG, 2005. OMG Model Driven Architecture, [Online]. Available from: http://www.omg.org/mda [Accessed: February, 2019].

Open Energy Monitor, Open source monitoring for understanding energy, https://openenergymonitor.org/ [Accessed in March 2018].

Pereira et al, 2014. Pereira, Punal P., Jens Eliasson J., and Delsing J. An authentication and access control framework for CoAP-based Internet of Things. In Industrial Electronics Society, IECON 2014-40th Annual Conference of the IEEE, pp. 5293-5299. IEEE, 2014.

Shelby Z, Hartke K., and Bormann C., 2014. The constrained application protocol (CoAP), RFC 7252 (Proposed Standard), Jun. 2014.

Sultanovs E., Skorobogatjko A., and Romanovs A., 2016. Centralized Healthcare Cyber-Physical System's Architecture Development. In: Proceedings of the 2016 57th International Scientific Conference on Power and Electrical Engineering of Riga Technical University, Latvia, Riga, 13-14 October, 2016. Riga: RTU Press, 2016, pp.153-158.

Varga P., Blomstedt F., Ferreira L. L., Eliasson J., Johansson M., Delsing J, and de Soria I. M., 2016. Making system of systems interoperable— The core components of the Arrowhead Framework, J. Network Computer Appl., vol. 81, pp. 85–95, Mar. 2016.

Zabasta A., Kuņicina N., Kondratjevs K., Patlins A., Ribickis L., Delsing J., 2018. MQTT Service Broker for Enabling the Interoperability of Smart City Systems". International Conference on Energy and Sustainability in Small Developing Economies - ES2DE18. Funchal, Spain, 9 - 11 July 2018, pp.1-6.

Zacepins A., V. Brusbardis V., Meitalovs J., and Stalidzans E., 2015. Challenges in the development of Precision Beekeeping, Biosyst. Eng., vol. 130, pp. 60–71, Feb. 2015.

Zacepins A., Kviesis A., Pecka A., and Osadcuks V., 2017. Development of Internet of Things concept for Precision Beekeeping, Conference: 2017 18th International Carpathian Control Conference (ICCC), P.1-5.

Zacepins A., Stalidzans E., and Meitalovs J., 2012. Application of information technologies in precision apiculture, in Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012), p.1-6.

AUTHORS BIOGRAPHY

Anatolijs Zabasta, Dr. Sc. Ing., Senior researcher and Projects manager has been working in RTU since 2010. The fields of scientific interests are electrical engineering, embedded systems, critical infrastructure. He was a researcher and project manager in EU projects: ERASMUS+ CBHE, TEMPUS, FP7 ARTEMIS, INTERREG, COST Actions and Latvian State Research Program's projects. A. Zabasta is an author of 75 scientific publications. Contacts: +371-29232872, anatolijs.zabasta@rtu.lv.

Nadezhda Kunicina, Dr. Sc. Ing. Professor, has been working in RTU from 2005. The fields of scientific interests are electrical engineering, embedded systems,

sustainable transport systems, and energy effectiveness in industrial electronics and electric transport. She was a senior researcher and a scientific project manager of FP7 ARTEMIS, INTERREG, COST Actions, ERASMUS+ CBHE, TEMPUS and Latvian State Research Program's projects. N. Kunicina is an author of 139 scientific publications. Contacts: +371-67089052, nadezda.kunicina@rtu.lv.

Kaspars Kondratjevs, PhD student, Researcher has been working in RTU since 2013. The fields of scientific interests are electrical engineering, embedded systems, critical infrastructure, wireless communication systems. He was a researcher in EU projects: FP7 ARTEMIS, INTERREG, COST Actions and in Latvian State Research Program's projects. K. Kondratjevs is an author of 26 scientific publications. Contacts: +371-26123500, Kaspars.kondratjevs@gmail.com.

Leonids Ribickis, Dr. Habil. Sc. Ing., prof., Director of Institute of Industrial Electronics and Electrical Engineering, Rector of RTU, Academician of the Latvian Academy of Sciences. His academic expertise - training of the specialists in the field of electro physics, energy effective lighting, power electronics, equipment design, industrial as well as micro- and nanoelectronics, usage of semiconductors and energy saving. Member of the Board of the World Energy Council Latvian National Committee, Head of the Latvian sub-department of the IEEE. Contacts: +371-67089300, leonids.ribickis@rtu.lv

SIMULATION-BASED EVALUATION OF AUTOMATED TRADING STRATEGIES: A MANIFESTO FOR MODERN METHODS

Dave Cliff
Department of Computer Science, University of Bristol, Bristol BS8 1UB, U.K
cstdc@bristol.ac.uk

ABSTRACT
In many investment banks and major fund-management companies, automated "robot" trading systems now do work that 20 years ago would have required large numbers of human traders to perform: the rise of robot traders is a major success-story for artificial intelligence (AI) research. Although the technical details of currently profitable automated trading systems are closely guarded commercial secrets, the rise of robot trading can be traced back to a sequence of key AI research papers. Each of these key papers relied on minimal abstract simulation models of real financial markets: the simulators provide test-beds for trials in which the performance of different trading strategies could be evaluated and compared. Recent studies have revisited these seminal results, using more realistic simulations of contemporary financial markets, and have cast major doubts on core conclusions drawn in the original publications. Therefore, it seems reasonable to argue that present-day simulation methods are exposing significant problems in past research on automated trading. This position paper presents no new empirical results but instead presents a review of key past papers and an argument, a manifesto, for establishing a shared market-simulator test-bed that adequately reflects current real-world financial markets, for use in future evaluation and comparison of trading strategies.

Keywords: financial markets, automated trading, simulation methodology.

1. INTRODUCTION

In the past fifteen years most major financial markets around the world have undergone seismic shifts in the extent to which highly-paid (and, presumably, highly intelligent) human traders have been replaced by automated trading systems known variously as *algorithmic trading systems* or more simply as *robot traders*. In many investment banks and major fund-management companies, automated trading systems now do work that previously would have required tens or hundreds of human traders to perform: in this sense, the rise of robot traders is a major success-story for artificial intelligence (AI) research.

Although the technical details of profitable automated trading systems are closely guarded commercial secrets, the rise of robot trading can be traced back to a sequence of key publications by academic and industrial AI researchers, commencing in 1993 and continuing through to 2008. Each of these key papers relied (to varying extents) on simulation models of real financial markets: the market simulators provide test-beds for trials in which the performance of different trading strategies could be evaluated, and emphasis was given to establishing which strategy out-performed (or "dominated") other strategies previously described in the published literature. All of these key studies involved minimal, abstract simulation models of real financial markets. Results from this sequence of papers are widely cited and have until very recently been essentially unquestioned. However, recent studies have revisited these results, extending the nature of the trials that the various strategies are subjected to, and exploring their responses in more realistic simulations of contemporary financial markets: this has cast major doubts on core conclusions drawn in the original publications.

The recent studies involve highly compute-intensive brute-force exhaustive simulation approaches, methods that arguably would have been prohibitively expensive to attempt when the original research was undertaken in the 1990's. Thus, it seems that modern-day simulation methods are exposing significant problems in past research. This position paper presents no new empirical results but instead presents an argument, a manifesto, for establishing a modernized methodology for evaluating trading agents. Specifically, it is proposed here: (1) that the simple abstract models of markets that were used previously should be replaced by a simulation and modelling approach that more accurately reflects the micro and macro structure of present-day financial markets and the traders that interact within them, markets in which *co-adaptive dynamics* are a major factor; (2) that researchers pay more attention to the combinatorics of rigorous evaluation – if tens or hundreds of millions of simulation trials are required to rigorously establish a result, we should not shy away from that; and (3) that open-source software methods should be fully exploited to ensure that the international community of researchers working on automated trading share common simulation tools, thereby easing replication and extension of earlier results. In the final section of this paper, I describe a major update to an existing open-source financial-market simulator which is now offered as a freely available resource for the research community. The updated simulator captures many aspects of current financial markets that have been absent in previous simulation-based studies, and is offered as a free resource to the community in the hope that it becomes a trusted common test-bed for future simulation-based evaluation of automated trading strategies.

In Section 2 we explain the background to this work: there is quite a lot to cover. Section 2.1 briefly reviews the rise of automated trading in the global financial markets. Section 2.2 then introduces the concepts and terminology from the economics of market

microstructure that are relevant to the discussion here. After that, Section 2.3 reviews a sequence of key papers in the field and then Section 2.4 discusses recent papers which, using contemporary simulation approaches, overturn conclusions drawn in the earlier papers. Sections 3 and 4 then respectively discuss microstructural and macrostructural issues in simulating financial markets for evaluating automated trading systems. This leads into Section 5's closing description of a modernized simulator (i.e. one that better reflects the structure of current financial markets) created by extending an established public-domain open-source exchange simulator, and its intended use-cases.

2. TRADERS, MARKETS, AND KEY PAPERS

The 2002 Nobel Prize in Economics was awarded to Vernon Smith, in recognition of Smith's work in establishing and thereafter growing the field of *Experimental Economics* (abbreviated hereafter to "EE"). Smith showed that the microeconomic behaviour of human traders interacting within the rules of some specified market, known technically as an *auction mechanism*, could be studied empirically, under controlled and repeatable laboratory conditions, rather than in the noisy messy confusing circumstances of real-world markets. The minimal laboratory studies could act as useful proxies for studying real-world markets of any type, but one particular auction mechanism has received the majority of attention: the *Continuous Double Auction* (CDA), in which any buyer can announce a bid-price at any time and any seller can announce an offer-price at any time, and in which at any time any trader in the market can accept an offer or bid from a counterparty, and thereby engage in a transaction. The CDA is the basis of major financial markets worldwide.

Each trader in one of Smith's experimental CDA markets would be assigned a private valuation, a secret *limit price*: for a buyer this was the price above which he or she should not pay when purchasing an item; for a seller this was the price below which he or she should not sell an item. These limit-price assignments model the client orders executed by sales traders in real financial markets; we'll refer to them just as *assignments* in the rest of this paper. Traders in EE experiments from Smith's onwards are often motivated by payment of some form of real-world reward that is proportional to the amount of *profit* that they accrue from their transactions: the profit is the difference between the limit price specified when a unit is assigned to a trader, and the actual transaction price for that unit.

The limit prices in the assignments defined the market's supply and demand schedules, which are commonly illustrated in economics texts as supply and demand curves on a 2D graph with quantity on the horizontal axis and price on the vertical axis: where the two curves intersect is the market's theoretical competitive equilibrium point – a pair of (price, quantity) coordinates. A fundamental observation from microeconomics (the study of markets and prices) is that competition among buyers pushes prices up, and competition among sellers pushers prices down, and these two opposing influences on prices balance out at the competitive equilibrium point; a market in which transaction prices rapidly and stably settles to the theoretical equilibrium price is often viewed by economists as *efficient* (for a specific definition of efficiency) whereas a market in which transactions consistently occur at off-equilibrium prices is usually thought of as inefficient: for instance, if transaction prices are consistently above the theoretical equilibrium price then it's likely that buyers are being ripped off. By varying the prices in the assignments to the traders, the nature of the market's supply and demand curves could be altered, and the effects of those variations on the speed and stability of the market's convergence toward an equilibrium could be measured.

Smith's initial set of experiments were run in the late 1950's, and the results and associated discussion were presented in his first paper on EE, published in the highly prestigious *Journal of Political Economy* (JPE) in 1962. It seems plausible to speculate that when his JPE paper was published, Smith seemingly had no idea that it would mark the start of a line of research that would eventually result in him being appointed as a Nobel laureate. And it seems even less likely that he would have foreseen the extent to which the experimental methods laid out in that 1962 paper would subsequently come to dominate the methodology of researchers working to build adaptive autonomous trading agents by combining tools and techniques from AI, ML, agent-based modelling (ABM), and agent-based computational economics (ACE). Although not a goal stated at the outset, this strand of AI/ML/ABM/ACE research converged toward a common aim: specifying an artificial agent, an autonomous adaptive trading strategy, that could automatically tune its behavior to different market environments, and that could reliably beat all other known automated trading strategies, thereby taking the crown of being the current best trading strategy known in the public domain, i.e., the "dominant strategy". Over the past 20 years the dominant strategy crown has passed from one algorithm to another. Vytelingum's (2006) *AA* strategy, was widely believed to be the dominant strategy, but recent results using contemporary large-scale computational simulation techniques indicate that it does not perform so well as was previously believed from its initial success in small numbers of trials.

Given that humans who are reliably good at trading are generally thought of as being "intelligent" in some reasonable sense of the word, the aim to develop ever more sophisticated artificial trading systems is clearly within the scope of AI research, although some very important early ideas came from the economics literature: a comprehensive review of relevant early research was given in Cliff (1997). Below in Section 2.1 we first briefly introduce eight key publications leading to the development of AA; then describe key aspects of EE market models in Section 2.2; and then discuss each of the eight key publications in more detail in Section 2.3. After that, Section 2.4 summarizes the results of Vach

(2015) and Cliff (2019), which cast doubts on the hitherto apparently resolved issue of which trading agent is best.

2.1. A Brief History of Trading Agents

If our story starts with Smith's 1962 JPE paper, then the next major step came 30 years later, with a surprising result published in the JPE by Gode and Sunder (1993): this popularized a minimally simple automated trading algorithm now commonly referred to as *ZIC*. A few years later two closely related research papers were published independently and at roughly the same time, each written without knowledge of the other: the first was a Hewlett-Packard Labs technical report by Cliff (1997) describing the adaptive AI/ML trading-agent strategy known as the *ZIP* algorithm; the second summarized the PhD thesis work of Gjerstad, in a paper co-authored with his PhD advisor (Gjerstad & Dickhaut 1998), describing an adaptive trading algorithm now widely known simply as *GD*. After graduating his PhD, Gjerstad worked at IBM's TJ Watson Labs where he helped set up an EE laboratory that his IBM colleagues used in a study that generated world-wide media coverage when the results were published by Das *et al.* at the prestigious *International Joint Conference on AI* (IJCAI) in 2001. This paper presented results from studies exploring the behavior of human traders interacting with GD and ZIP robot traders, in a CDA with a Limit Order Book (LOB: explained in more detail in Section 2.2, below), and demonstrated that both GD and ZIP reliably outperformed human traders. Neither GD nor ZIP had been designed to work with the LOB, so the IBM team modified both strategies for their study. A follow-on 2001 paper by Tesauro & Das (two co-authors of the IBM IJCAI paper) described a more extensively *Modified GD* (MGD) strategy, and later Tesauro & Bredin (2002) described the *GD eXtended* (GDX) strategy. Both MGD and GDX were each claimed to be the strongest-known public-domain trading strategies at the times of their publication.

Subsequently, Vytelingum's 2006 thesis introduced the *Adaptive Aggressive* (AA) strategy which, in a major journal paper (Vytelingum *et al.*, 2007), and in later conference papers (De Luca & Cliff 2012a, 2012b), was shown to be dominant over ZIP, GDX, and human traders. Thus far then, AA held the title.

However Vach (2015) presented results from experiments with the *OpEx* market simulator (De Luca, 2015), in which AA, GDX, and ZIP were set to compete against one another, and in which the dominance of AA was questioned: Vach's results indicate that whether AA dominates or not can be dependent on the ratio of AA:GDX:ZIP in the experiment: for some ratios, Vach found AA to dominate; for other ratios, it was GDX. Vach studied only a relatively small sample from the space of possible ratios, but his results prompted Cliff (2019) to exhaustively run through a wide range of differing ratios of four trading strategies (AA, ZIC, ZIP, and the minimally simple SHVR strategy described in Section 2.2), doing a brute-force search for situations in which AA is outperformed by the other strategies. The combinatorics of such a search are quite explosive: Cliff reported on results from over 3.4 million individual simulations of market sessions, and his findings indicated that Vach's observation was correct: AA's dominance does indeed depends on how many other AA traders are in the market; and, in aggregate, AA was routinely outperformed by ZIP and by SHVR. Subsequent research by Snashall (2019) employed the same exhaustive testing method, using a supercomputer to run more than one million market simulations, to exhaustively test AA against IBM's GDX strategy: this again revealed that AA does not always dominate GDX: see Snashall & Cliff (2019) for further discussion.

2.2. On Simulation Models of Markets

Vernon Smith's early experiments were laboratory models of so called *open-outcry trading pits*, a common sight in any real financial exchange before the arrival of electronic trader-terminals in the 1970s. In a trading pit, human traders huddle together and shout out their bids and offers, and also announce their willingness to accept a counterparty's most recent shout. It was a chaotic scene, now largely consigned to the history books. In the closing quarter of the 20th Century, traders moved *en masse* to interacting with each other instead via electronic means: traders "shouted" their offer or bids or acceptances by typing orders on keyboards and then sending those orders to a central server that would display an aggregate summary of all orders currently "shouted" (i.e., quoted) onto the market. That aggregate summary is very often in the form of a *Limit Order Book* or LOB: the LOB summarizes all bids and offers currently live in the market. At its simplest, the LOB is a table of numbers, divided into the *bid side* and the *ask side* (also known as the *offer side*). Both sides of the LOB show the best price at the top, with less good prices arranged below in numeric order of price: for the bid side this means the highest-priced bid at the top with the remaining bid prices displayed in descending order below; and for the ask side the lowest-priced offer is at the top, with the remaining offers arranged in ascending order below. The arithmetic mean of the best bid and best ask prices is known as the *mid-price*, and their difference is the *spread.* For each side of the LOB, at each price on the LOB, the total quantity available is also shown, but with no indication of who the relevant orders came from: in this sense the LOB serves not only to aggregate all currently live orders, but also to anonymize them.

Traders in LOB-based markets can usually cancel existing orders to delete them from the LOB. In a common simple implementation of a LOB, traders can accept the current best bid or best offer by issuing a quote that *crosses the spread*: i.e., by issuing an order that, if added to the LOB, would result in the best bid being at a higher price than the best ask. Rather than be added to the LOB, if a bid order crosses the spread then it is matched with the best offer on the ask side (known as *lifting the ask*), whereas an ask that crosses the spread is matched with the best bid (*hitting the bid*); and in either case a transaction then occurs between the trader that had posted the best price on the relevant side of the LOB, and

the trader that crossed the spread. The price of the resulting transaction is whatever price was hit or lifted from the top of the LOB.

Smith's earliest experiments pre-dated the arrival of electronic trading in real financial markets, and so they can be thought of as laboratory models of open-outcry trading pits: they were simulations of real markets, but were initially not *computer-based* simulations. Even though the much later work by Gode & Sunder, Cliff, Gjerstad & Dickhaut, and Vytelingum all came long after the introduction of electronic LOBs in real markets, these academic studies all stuck with Smith's original methodology, of modelling open-outcry markets (often by essentially operating a LOB with the depth fixed at 1, so the *only* information available to traders is the current best, or most recent, bid and ask prices).

Nevertheless, the studies by IBM researchers (Das *et al.*, 2001; Tesauro & Das, 2001; Tesauro & Bredin 2012), and also the replication and confirmation of AA results by De Luca & Cliff (2011a, 2011b) and by Stotter *et al.* (2013), all used LOB-based market simulators. The IBM simulator *Magenta* seems to have been proprietary to IBM; developed at TJ Watson Labs and not available for third-party use, but De Luca made an open-source release of his *OpEx* simulator (De Luca, 2015) which was subsequently used by Vach (2015) in the studies that prompted our work reported here. Also of relevance here is the *ExPo* simulator described by Stotter *et al.* (2013, 2014): in the work by De Luca, by Vach, and by Stotter *et al.*, Vytelingum's original AA needed modifications to make it work in a LOB-based market environment.

In this paper neither OpEx nor ExPo will be discussed further, but instead we will concentrate on BSE (BSE, 2012; Cliff, 2018) which is another open-source EE market simulator, initially developed as a teaching aid but subsequently used as a platform for research (see, e.g. Le Calvez & Cliff, 2018, Snashall & Cliff 2019). BSE has the advantage of being relatively lightweight (a single Python script of c.2500 lines) and hence readily deployable over large numbers of virtual machines in the cloud. BSE maintains a dynamically updated LOB and also publishes a *tape*, a time-ordered record of all orders that have been executed. It comes with built-in versions of ZIC and ZIP, and also some additionally minimally-simple non-adaptive trading strategies that can be used for benchmarking against other more complex strategies added by the user. One of these, the *Shaver* strategy (referred to in BSE by the abbreviation SHVR) simply reads the best prices on the LOB and, if it is able to do so without risking a loss-making deal then it issues an order that improves the current best bid or best ask by 0.01 units of currency (i.e., one penny/cent), which is BSE's *tick size*, i.e. the minimum change in price that the system allows. Another of the BSE built-in trading strategies is even simpler than SHVR: the *Giveaway* strategy (abbreviated to GVWY) attempts to make no profit at all from trading, and simply posts a bid or offer price that is equal to the limit price assigned to it for that unit. As we will further discuss later in Section 4, when evaluated using conventional market simulation methods like those used in the papers reviewed below, GVWY can prove to be a profitable strategy: this counterintuitive result is one indication that evaluation via conventional means has some significant limits.

2.3. Eight Key Papers with a Common Methodology

2.3.1. Smith 1962

Although precedents can be pointed to, Smith's 1962 JPE paper is widely regarded as the seminal study in EE. In it he reported on experiments in which groups of human subjects were randomly assigned to be either *buyers* or *sellers*. Buyers were given a supply of artificial money, and sellers were given one or more identical items, of no intrinsic value, to sell. As discussed above, each trader in the market was assigned a private valuation, a secret *limit price*, above which they should not pay when buying and below which they should not accept when selling.

After the allocation of assignments to all subjects, they then interacted via an open-outcry CDA while Smith and his assistants made notes on the sequence of events that unfolded during the experiment: typically, buyers would gradually increase their bid-prices, and sellers would gradually lower their offer-prices (also known as ask-prices) until transactions started to occur. Eventually, usually within a few minutes, the experimental market reached a position in which no more trades could take place, which marked the end of a *trading period* or "trading day" in the experiment; any one experiment typically ran for n=5-10 periods, with all the traders being resupplied with money and items-for-sale at the start of each trading period. The sequence of n contiguous trading periods (or an equivalently long single-period experiment with continuous replenishment, as discussed in Section 4.3) is referred to here as one *market session*.

Smith could induce specific supply and demand curves in these experimental markets by appropriate choices of the various limit-prices he assigned to the traders. The market's theoretical *equilibrium price* (denoted hereafter by P_0) is given by the point where the supply curve and the demand curve intersect. Smith found that, in these laboratory CDA markets populated with only remarkably small groups of human traders, transaction prices could reliably and rapidly converge on the theoretical P_0 value despite the fact that each human trader was acting purely out of self-interest and knew only the limit price that he or she had been assigned.

Smith's analysis of his results focused on a statistic that he referred to as α, the root mean square (RMS) deviation of actual transaction prices from the P_0 value over the course of an experiment. In his early experiments, P_0 was fixed for the duration of any one experiment; in later work Smith explored the ability of the market to respond to "price shocks" where, in an experiment of N trading days, on a specific day $S<N$ the allocation of limit prices would be changed, altering P_0 from the value that had been in place over trading periods $1, 2, \ldots, S,$ to a different value of P_0 that would then remain constant for the rest of the experiment, i.e. in

trading periods *S+1, S+2, ..., N*. For brevity, in the rest of this paper Smith's initial style of experiments will be referred to as *S'62* experiments.

2.3.2. ZIC: Gode & Sunder (1993)

Gode & Sunder's 1993 JPE paper used the S'62 methodology, albeit with the CDA markets being electronic (a move Smith himself had made in his experiments many years earlier), so each trader was sat at a personal terminal, a computer screen and keyboard, from which they received all information about the market and via which they announced their orders, their bids or offers, to the rest of the traders in the experiment. Gode & Sunder (G&S hereafter) first conducted a set of experiments in which all the traders were human, to establish baseline statistics. Then, all the human traders were replaced with automated trading systems, absolute-zero minimally-simple algo traders which G&S referred to as *Zero Intelligence* (ZI) traders. G&S studied markets populated with two type of ZI trader: ZI-Unconstrained (ZIU), which simply generated random prices for their bids or offers, regardless of whether those prices would lead to profitable transactions or to losses; and ZI-Constrained (ZIC), which also generated random order prices but were constrained by their private limit prices to never announce prices that would lead them to loss-making deals. G&S used fixed supply and demand schedules in each experiment, i.e. there were no price-shocks in their experiments.

Not surprisingly, the market dynamics of ZIU traders were nothing more than noise. But the surprising result in G&S's paper was the revelation that a commonly used metric of market price dynamics known as *allocative efficiency* (AE, hereafter) was essentially indistinguishable between the human markets and the ZIC markets. Because AE had previously been seen as a marker of the degree to which the traders in a market were behaving intelligently, the fact that ZIC traders scored AE values largely the same as humans was a shock. Gode & Sunder proposed that a different metric should instead be used as a marker of the intelligence of traders in the market. This metric was *profit dispersion* (PD, hereafter) which measures the difference between the profit each trader accrued in an experiment, compared to the profit that would be expected for that trader if every transaction in the market had taken place at the market's theoretical equilibrium price P_0: humans typically showed very low values of PD (which is assumed to be good) while ZIC traders did not. On this basis, G&S argued that PD should be used in preference to AE.

Other researchers were quick to cite G&S's ZIC result, and often used it to support the claim that, given the ZIC traders have no intelligence, then for transaction prices to converge toward the theoretical equilibrium price and/or for a group of traders to score highly on AE, somehow the "intelligence" required to do this must reside within the rules of the CDA market system rather than within the heads of the traders. Strangely, G&S's 1993 paper provides no concrete causal mechanistic explanation of how their striking ZIC results arise; they describe their methods, and the results observed, but the internal mechanisms that give rise to those results are left as something of a mystery, as if the CDA market was an impenetrable black-box.

A causal mechanistic analysis of markets populated by ZIC traders was subsequently developed in (Cliff 1997), which considered the probability mass functions (PMFs) of prices generated by ZIC buyers and sellers, and the joint PMF of transaction prices in ZIP markets, which is given by the intersection of the bid-price and offer-price PMFs: the shape of the transaction-price PMF is determined by the nature of the supply and demand curves in the market, and (Cliff, 1997) demonstrated that the supply and demand curves in a ZIC market experiment could be arranged so that the expected value of the transaction prices (computable as an integral over the PMF) is identical to the theoretical equilibrium price given by the intersection point of the supply and demand curves. This was why the five ZIC experiments reported in G&S's 1993 paper showed transaction prices that were centered on the theoretical equilibrium price in each case: the supply and demand curves were arranged in such a way that this was the expected outcome. Cliff (1997) showed that with different arrangements of supply and demand curves, such as situations where one or both curves were flat (as had been used in Smith's 1962 JPE paper), the expected price of transactions in ZIP markets could differ considerably from the theoretical equilibrium price, and so transaction prices in those ZIC markets would fail to exhibit human-like convergence toward the theoretical equilibrium value. In these differently-designed experiments, ZIC traders would be revealed for exactly what they are: simple stochastic processes that only coincidentally exhibit human-like market dynamics when the experimenters happen to have chosen to impose the right kind of supply and demand curves. The (Cliff 1997) analysis showed that the level of intelligence in the ZIC traders was insufficient to recreate human-like market dynamics more broadly, and so a more intelligent automated trading strategy was required.

Independently, roughly a decade later, and via a wholly different line of attack Gjerstad & Shachat (2007) also demolished the argument that G&S's ZIC results indicate that the efficiency or intelligence in the market system lies solely within the CDA mechanism. Nevertheless, G&Ss 1993 results continue to be widely and uncritically cited by various authors.

2.3.3. ZIP: Cliff (1997)

Taking direct inspiration both from Smith's work and from the ZI paper by G&S, Cliff (1997) developed a ZI trading strategy that used simple machine-learning techniques to continuously adapt the randomly-generated prices quoted by the traders: this strategy, known as ZI-Plus (ZIP) was demonstrated to show human-like market dynamics in experiments with flat supply and/or demand curves: (Cliff 1997) also showed theoretical analyses and empirical results which demonstrated that transaction prices in markets populated only by ZIC traders would

not converge to the theoretical equilibrium price when the supply and/or demand curves are flat (or, in the language of microeconomics, "perfectly elastic"). EE studies in which the supply and/or demand curve was flat had previously been reported by Smith and others, but G&S had not explored the response of their ZIC traders to this style of market. The work in (Cliff, 1997) involved no human traders: all the focus was on markets populated entirely by autonomous agents, by ZIP traders. In total there are results from fewer than 1,000 simulated market sessions reported on in (Cliff 1997). In all other regards (Cliff 1997) continued the S'62 tradition: key metrics were Smith's α, AE, and PD, and the focus on *homogenous* markets continued the tradition established by Smith (who studied all-human markets) and by G&S (who studied markets homogeneously populated with either human, ZIU, or ZIC traders).

2.3.4. GD: Gjerstad & Dickhaut (1997)

Gjerstad's PhD studies of price formation in CDA markets also involved creating an algorithm that could trade profitably by adapting its behavior over time, in response to market events (Gjerstad & Dickhaut, 1998). In contrast to the ZI work, Gjerstad's trading algorithm uses frequentist statistics, gradually constructing and refining a *belief function* that estimates the likelihood for a bid or offer to be accepted in the market at any particular time, mapping from price of the order to its probability of success. Gjerstad did not explicitly name his strategy, but it has since become widely known as the GD strategy. In all other regards, as with Cliff (1997) and G&S (1993), Gjerstad's work was firmly in the S'62 tradition: homogenous markets of GD traders interacting in a CDA, buying and selling single items, with the metrics being Smith's α, AE, and PD. In a later paper, Gjerstad (2003) made some refinements to the GD algorithm and named it *HBL* (Heuristic Belief Learning), although the original GD remains by far the most cited.

2.3.5. MGD: Das *et al.* (2001)

In their landmark 2001 IJCAI paper, IBM researchers Das, Hanson, Kephart, & Tesauro studied the performance of GD and ZIP in a series of EE market experiments where, for the first time ever in the same market, some of the traders were robots while others were human (recall that the earlier work of Smith, of G&S, of Cliff, and of Gjerstad & Dickhaut had all studied homogeneous markets: either all-human or all-robot). Das *et al.* used a LOB-based market simulator called *Magenta*, developed by Gjerstad, and ran a total of six experiments, six market sessions, in which humans and robots interacted and where there were three shock-changes to P_0, i.e. four phases in any one experiment, each phase with a different P_0 value that was held static over that phase. The surprising result in this paper was that robot trading strategies could consistently outperform human traders, by significant margins: a result that attracted worldwide media attention. Both GD and ZIP outperformed human traders, and in the six experiments reported by Das *et al.* the results from the two robot strategies are so similar as to not obviously be statistically significant. A subsequent paper by IBM's Tesauro & Das (2001), reported on additional studies in which a *Modified GD* (MGD) strategy was exhibited what the authors described in the abstract of their paper as *"…the strongest known performance of any published bidding strategy*".

2.3.6. GDX: Tesauro & Bredin (2002)

Extensions to MGD were reported by IBM researchers Tesauro & Bredin (2002) at AAMAS 2002. This paper described extensions to MGD, using dynamic programming methods: the extended version was named *GDX* and its performance was evaluated when competing in heterogenous markets with ZIP and other strategies. Tesauro and Bredin reported that GDX outperformed the other strategies and claimed in the abstract of their paper that GDX *"...may offer the best performance of any published CDA bidding strategy*."

2.3.7. AA: Vytelingum (2006)

Vytelingum developed AA and documented it in full in his PhD thesis (2006) and in a major paper (Vytelingum *et al.,* 2008). The internal mechanisms of AA are described in greater detail in Section 3 of this paper. Although Vytelingum's work came a few years after the IBM publications, the discussion within Vytelingum's publications is phrased very much in terms of the S'62 methodology: the P_0 value in his AA experiments was either fixed for the duration of each market session, or was subjected to a single "price shock" partway through the session (as described in Section 2.3.1); and again the primary metrics studied are Smith's α, AE, and PD. Vytelingum presented results from heterogeneous market experiments where AA, GDX, and ZIP traders were in competition, and the published results indicated that AA outperformed both GDX and ZIP by small margins. In total, results from c.25,000 market sessions are presented in (Vytelingum *et al.,* 2008).

2.3.8. AA Dominates: De Luca & Cliff (2011)

As part of the research leading to his 2015 PhD thesis, De Luca used his LOB-based *OpEx* market simulator system (De Luca, 2012) to study the performance of AA in heterogeneous market experiments where some of the traders were AA, some were other robot strategies such as ZIP, and some were human traders sat at terminals interacting with the other traders (human and robot) in the market via the OpEx GUI, in the style introduced by the IBM team in their IJCAI 2001 paper. De Luca & Cliff (2011a) had previously published results from comparing GDX and AA in OpEx, at ICAART-2011; and the first results from AA in human-agent studies were then published in a 2011 IJCAI paper (De Luca & Cliff, 2011b), in which AA was demonstrated to dominate not only humans but also GDX and ZIP. For consistency with what was by then a well-established methodology, in De Luca's experiments the P_0 value was static for sustained periods with occasional "shock" step-changes to different values. Continuing the tradition

established by the IBM authors, the abstract of (De Luca & Cliff 2011b) claimed supremacy for AA: "*We… demonstrate that AA's performance against human traders is superior to that of ZIP, GD, and GDX. We therefore claim that… AA may offer the best performance of any published bidding strategy*". And, until the publication of Vach (2015), that claim appeared to be plausibly true.

2.4. Actually, AA doesn't dominate

Vach's 2015 Master's Thesis tells the story of his design of a new trading strategy based on ZIP and called ZIPOJA, which he then tested against AA, GDX, and ZIP. The testing revealed that ZIPOJA did not consistently outperform any of the three pre-existing strategies. But, in the course of that testing, as Vach checked and calibrated his implementations of the three pre-existing strategies, he found that AA could fail to dominate ZIP or GDX, depending on the proportions of the two strategies in the market: this runs counter to the established story that AA is the best-performing strategy. Tables 6.2 and 6.3 on p.47 of Vach's thesis show results from tests in which the performance of two trading strategies were tested in trials with proportions of the two trader strategies set at 6:0, 5:1, 4:2, 3:3, 2:4, 5:1, and 0:6. The ratios 6:0 and 0:6 are homogenously populated by one strategy or the other and hence are of little interest, because that one strategy necessarily dominates in those markets. In Vach's Table 6.2, AA is outperformed by ZIP when the ZIP:AA ratio is 1:5 – i.e, if one in six of the traders in the market are ZIP with the rest AA, then the ZIP traders will outperform the AAs: the efficiency of the ZIP traders was 99.5% while the efficiency of the AAs was 88.5%. In Vach's Table 6.3, AA is outperformed by GDX when the GDX:AA ratio is 3:3, 2:4, and 1:5.

Vach then performed three-way simulations systematically varying the ratios of AA:GDX:ZIP over all possible permutations and, in his Fig.6.1i (p.53) he shows a 2D simplex diagram which summarizes those results: a 28-node regular isometric mesh is drawn over the surface of the simplex as a co-ordinate frame, and AA is the dominant strategy in only 11 of those 28 nodes. Each of the three strategies is by definition dominant at the node representing a homogeneous ratio (i.e., either 1:0:0 or 0:1:0 or 0:0:1), so AA actually only dominates at 10 of the 25 nodes where it is actually contesting with the other two strategies: ZIP dominates one of the remaining nodes, and GDX dominates the remaining 14.

In a final four-way study, with AA, GDX, ZIP, and ZIPOJA competing against each other, Vach (2015, Table 6.7, p.60) declares GDX the overall winner although in that experiment the scores of GDX and AA are sufficiently close that the difference between the two is not obviously significant. Nevertheless, it is undeniable that in Vach's four-way study AA fails to clearly dominate. To the best of our knowledge, Vach's results are the first exhaustive study of AA's performance as the number and proportion of competitor strategies is systematically varied, and he was the first to demonstrate that AA is in fact not the best-performing strategy.

Cliff (2019) set out to replicate and extend Vach's results, using a finer-grained analysis, varying the proportions of AA, SHVR, ZIP, and ZIC, and also studying the effects of altering other aspects of the experiment design such as whether the replenishment of assignments to the traders is periodic or continuous-stochastic (see e.g. Cliff & Preist 2001); and whether the equilibrium price P_0 is largely constant with occasional shock-jumps, or continuously varying according to price-movements taken from real-world markets. Cliff's results from conventional S'62-style experiments, with periodic replenishment and largely constant, confirmed the established view: when AA was tested in the kind of simple market environment as has traditionally been used in the previous literature, AA scored just as well as well-known other trading strategies and was not dominated by them. But merely by altering the nature of the market environment to have continuous stochastic replenishment (which is surely what happens in real markets) and to have the equilibrium price P_0 continuously varying over time (which is also surely what happens in real markets), Cliff's results from AA were very poor indeed. Cliff (2019) concluded that AA's success as reported in previous papers seemed to be largely due to the extent to which AA's internal mechanisms are designed to fit exactly the kind of experiment settings first introduced by Vernon Smith: AA is very well suited to situations in which all assignments are issued to all traders simultaneously, and in which P_0 remains constant for sustained periods of time, with only occasional step-change "shocks". Real markets are not like this, and when AA is deployed in the more realistic market setting provided by BSE, Cliff (2019) demonstrated that AA's dominance disappears.

Cliff (2019) did not present results of exhaustive testing of AA against GDX, but Snashall's (2019) subsequent thesis did so, running more than one million simulation experiments. Snashall demonstrated that actually, even in the S'62 style of experiment that AA was first tested in, if AA is evaluated exhaustively in BSE across a wide range of proportions, then AA can be outperformed by GDX.

However, as Snashall (2019) argues, GDX's apparent superiority in some situations may itself be illusory, because of the computationally intensive nature of the GDX algorithm. In Snashall's study, GDX (which extends the original GD algorithm with techniques from dynamic programming) took roughly ten times longer than AA or ZIP to compute its response to a change in the LOB, every time the LOB data altered. Such a disproportionately long time spent "thinking" would most likely be a serious impediment to deploying GDX in today's electronic markets where speed of reaction time is a critical factor in determining the success or failure of an automated trading system: see Snashall & Cliff (2019) for further discussion.

3. MARKET MICROSTRUCTURAL ISSUES

The story told in Sections 2.1 to 2.4 should be of concern to anyone who cares about the use of simulation-based

evaluation in empirical science: a sequence of papers on performance of automated traders in simulated markets is published in leading international peer-reviewed AI conferences and journals, each building on and extending the work that had gone before it, and in which a set of apparently reasonable conclusions are drawn that, roughly a decade later, more sophisticated simulation studies call into serious question.

It is very important to note here that I am not calling into question the honesty or professionalism of the researchers involved in this sequence of papers. Each of the papers reviewed above was, as far as I know, prepared in good faith and then subjected to high standards of peer-review. There is no reason to doubt any of the results published in any of those papers. The issue is not with the results or the researchers, but rather with the wider methodological context, the "spirit of the time", within which these studies were conducted.

So, for example, there is no reason to doubt that Vytelingum's (2006, Vytleingum *et al.* 2008) published results showing AA outperforming GDX are genuine. They can be trusted as a fair representation of what happens when AA is placed in a S'62-style market simulator (as had also been used in Cliff's 1997 ZIP publication) in which all assignments are reissued periodically and there is no LOB. But as the results of Vach (2015), Cliff (2019), Snashall (2019), and Snashall & Cliff (2019) show, as soon as AA is deployed in a more realistic market simulator (such as BSE, in which there is a fully working LOB, and in which assignment updates arrive in a continuous random stream rather than in neat periodic bursts to all traders simultaneously), AA's dominance disappears.

AA, just like ZIP, GD, MGD, and GDX, has a simplicity in its specification that is a reflection of the simplicity of the underlying market simulators in which it was developed and tested. For example, if all traders' individual assignments are issued simultaneously in periodic bursts at the start of each trading period (each "day", as in the S'62 experiments), and each trader is given only one unit to buy or sell per assignment, then it makes perfect sense to also clear the LOB at the start of each trading period (also as in S'62) and hence the traders in the market never have to deal with order *cancellations* (i.e., where an existing limit order visible on the LOB is cancelled by the trader who issued that order, and the LOB is updated to reflect that change). As soon as the issuing of assignments is switched to be a continuous stochastic stream, occasions will arise in which a trader receives a new assignment and hence has to cancel a limit order that was previously issued to the exchange – e.g. because the new assignment has a radically different limit price to the previous one. As soon as cancellations are a routine occurrence, all of the trading algorithms described here need to be amended or extended to correctly distinguish between changes in the LOB that are the result of actual transactions taking place, and changes that are the result of cancellations.

So the real issue here instead seems (now, with hindsight) to be the extent to which the simplistic underlying market simulators were trusted by the entire community of researchers: the people doing the actual work and writing the papers; the people doing the reviewing; and the people who subsequently cited the published papers. (And, for the avoidance of doubt, I myself fall into all three of those categories: so, *mea culpa*). Smith's 1962 experiments used a non-computer-based manual simulation of the CDA found in real financial markets; once the S'62 methodology was established, it was natural for subsequent papers to use close copies of that methodology, to ease the comparison of new results with those already published. Hence it is natural that first G&S (1993), then Cliff (1997) and Gjerstad & Dickhaut (1998), and latterly Vytelingum (2006, 2008) all used essentially the same methodology.

Similarly, as Snashall (2019) has highlighted, the commonly-used simple single-threaded simulation approximations to what in real life is a parallel and asynchronous distributed system mean that the much longer compute-times required by GDX have not previously been highlighted by other researchers. Many real-world trading strategies have to be sensitive not only to the price that is agreed for a transaction, but also how much time is taken in arriving at an agreed price. While Kaplan's *Sniper* strategy (Rust *et al.* 1992) and Gjerstad's (2003) *HBL* strategy are both time-sensitive to varying degrees, these two strategies are the exception rather than the rule: yet the strategies described in the key papers reviewed here all essentially ignore timing issues and concentrate only on price. A very recent paper by Miles & Cliff (2019) discusses issues of latency and time-sensitivity in more detail.

And, in addition to considerations of price and time, there is a third factor that any real-world trading agent is likely to pay attention to: the quantity available at any particular price (also referred to as the *size* or the *volume*). In particular, if there is a much larger quantity available at the best price on one side of the LOB in comparison to the other side, that imbalance can be a strong indication that the market price is likely to move in the near future: excess supply pushes prices down; excess supply pushes price up. Market shocks can occur not only via sudden changes in prices (with quantities available remaining the same), but also by sudden changes in the quantities available (with the prices staying unchanged, initially at least). As with time-sensitivity, size-sensitivity was essentially ignored by the authors of the key papers reviewed above. A very recent publication by Church & Cliff (2019) discusses size issues in more detail.

All of these concerns are essentially *microstructural:* the details of how the assignments are distributed to the traders; how the traders' orders are processed by the exchange and what data the exchange then publishes to the traders for them to react to and act upon; and which factors of the orders on the LOB matter to traders (i.e: price, size, and timing). However, there are also broader issues, characterized here as *macrostructural*, discussed in the next section.

4. MARKET MACROSTRUCTURAL ISSUES

Smith's 1962 experiments are a fair approximation to an open-outcry *trading pit*, which was a common sight in financial exchanges prior to their wholesale automation in the last thirty-five years: a group of traders gathered in close physical proximity, shouting and gesturing at each other while trying to find a counterparty to trade with. It seems entirely reasonable for Smith to have set up a (manual) simulation of the financial exchanges of his day for his laboratory studies of the CDA, but that was more than half a century ago; surely today's market simulations should be structured in such a way that they are closer to the realities of today's actual financial markets?

The market simulators used in all the studies reviewed in Section 2 each essentially followed the S'62 pattern: some population of individual traders compete within a given market mechanism, and summary statistics (such as α, AE, and PD) are computed across the population of traders. So, in that sense, they are also simulations of a single trading pit. But the reliance on computing population-level statistics gives rise to a result that makes little sense: the minimally simple GVWY trading strategy, introduced in Section 2.2, seeks no profit at all and yet when its performance is measured in a population of traders, using standard population-level statistics, it can routinely outperform other more intelligent traders. In particular, GVWY traders (which always and immediately post orders on the LOB with prices equal to their private limit price, thereby guarantee them zero profit if their order is accepted by a counterparty) can often score an average profit per trader that is as good as, or better than, the average profit per trader of supposedly intelligent adaptive trading strategies such as AA or ZIP.

This counterintuitive result does have a rational explanation: an individual GVWY buyer(seller) can do well against any other type of trader if its private limit price L_G is sufficiently above(below) the best price on the ask(bid) side of the LOB. For example, say that you have a GVWY buyer G with limit price L_G =150, and the LOB's current best bid is 90 and best ask is 100, with that best ask coming from a seller whose private limit price is L_S=80): the GVWY trader G is polled to provide a quote and bids 150. This crosses the spread (because 150>100), and so the transaction goes through at price of current best ask, i.e. 100. That gives the seller who posted the ask a profit of 100-L_S=20, and the GVWY gets a profit of L_G -100=50. So, even though the GVWY order submitted to the market would have generated no profit for that trader *if it had been accepted at the indicated price*, instead the order crosses the spread and so the GVWY trader makes a nonzero profit equal to the difference between the best price on the counterparty side of the LOB. As an individual GVWY can never make a negative profit, its average score will be computed from the sum of a sequence of profits that are either zero or positive, and so the average profit per GVWY trader will often be a positive value. This shows how GVWY can make profit. Exactly how much profit GVWY can make depends quite a lot on the Supply/Demand schedule in the market at that time, and on the mix of other strategies that it finds itself competing with. But, crucially, adaptive strategies such as AA, ZIP, or GDX, may take some time to adjust their prices to the point where they successfully identify a counterparty to trade with, and in the time it takes an adaptive strategy to find a counterparty to trade with, a GVWY trader might have executed a sequence of several transactions. So although the GVWY average profit per *trade* may be small, and although its transactions may go through at prices some way distant from P0 (so its score for a may be poor in comparison to other strategies) by the end of an experiment its average profit per *trader* may actually be better than those of the adaptive "intelligent" trading strategies. Whether this happens or not depends on the proportions of the different trading staregies and the nature of the experimental market's supply and demand and updates to traders' assignments, but the fact that it can happen at all gives some pause for thought: in principle, when evaluated using the conventional techniques, GVWY could out-perform any of the more sophisticated trading strategies.

This calls into question the measuring of the performance of trading strategies at the population level: when monitored across a population, GVWY can come out as a good/dominant strategy because it's constrained to never post a quote price the wrong side of its limit price and so it never enters into loss-making deals. GVWY never loses, but occasionally wins big; aggregate this over lots of individuals doing lots of deals and GVWY's apparent profits can start to add up. Then couple that with the fact that GVWY is wholly incautious: it "goes for the deal" as fast as it can and either gets some profit or nothing, but then gets another assignment to deal with. Adaptive algorithms like ZIP and AA and MGD/GDX all spend a bit of time doing their adapting – they can often commence by quoting a price some distance from equilibrium and then gradually edge towards it, which means their count of trades per unit time is much lower than GVWY, which counts against them when they're ranked on profit-per-unit-time. Someone who studies trading strategies only in the conventional S'62 style of experiment might reasonably conclude that GVWY is a good strategy.

But this is not likely to go down well in a real trading situation, because real traders do not evaluate their performance at the population level: they are typically ruthlessly self-interested, and care only about the profit that they make as an individual. This is a key macro-level issue: while academic economists seem primarily interested in population-level statistics, individual traders tend instead to be singularly focused on their own personal profit-and-loss (P&L), and not much else.

From this it seems plausible to conjecture that the progression of ZIC-ZIP/GD-MGD/GDX-AA described in Section 2 may have been a result of people testing their trading agents in unrealistic environments and measuring them with the wrong metrics. If the underlying simulations had been more realistic, the tale of which strategy comes out as best (or whether *any* one strategy ever comes out as consistently better than the others) could plausibly have been a different story. And the best

way of testing that conjecture is to start building and using market simulators that more accurately reflect modern financial markets.

For instance, in many contemporary financial markets, the key players are no longer individual traders, single humans buying or selling, but instead the market is populated by some number *N* of trading *entities* which (although they could in principle still be individual humans) are more likely to be trading institutions such as banks or fund management companies. In most such trading entities, there will be a currently trusted, established trading system known as the *production system* (because it is deployed in live trading, in production) which is made up of *one or more* distinct trading strategies working in concert. Any one instance of a production system is likely to be well-suited to the market circumstances that prevailed at the time that it was introduced, but market circumstances rarely stay the same for long and so what is a profitable production system this week or this month may be generating much less income next week or next month, and for that reason it is common for trading entities to always also be working on a *development system*, again a set of *one or more* coordinated trading strategies, which is intended to be deployed as the eventual replacement for the current production system. Practitioners working with such coupled pairs of trading systems often use the abbreviations "prod" for the production system and "dev" for the development system: we'll use them in the rest of this paper too.

In the previous paragraph the phrase "one or more" was emphasized when introducing the notion of prod and dev systems to highlight the issue that many real-world trading entities now routinely employ what are commonly known as *algo-wheels*: systems that, for each order to be executed, automatically select one particular algorithmic trading strategy (the *algo*) from among a set of potentially applicable strategies (the *wheel*), depending on the nature of the order and the market conditions prevailing at the time. One of the first algo-wheel offerings to achieve widespread notice was the *Algo Switching Engine* from Pipeline Trading Systems LLC: see e.g. Stephens & Waelbroeck (2009).

A more accurate simulation model of contemporary automated markets would capture all of the above macrostructural factors: i.e., have the simulator allow for *N* trading entities, each trying to make profit from its current set of one or more prod trading strategies while also each working on one or more dev trading strategies that are intended to improve upon the current prod strategies. As in real markets, each entity should be able to monitor only its own (P&L) on its various strategies; for any one entity the technical details of its strategies and their individual P&L streams are private, not disclosed to any other trading entities, and hence any one entity has no knowledge of the technical details or the profitability of the trading strategies being deployed by the various other entities that it is competing with in the market. A simulator constructed in this way would be a major step towards providing a test-bed for simulation studies in which new trading strategies, expressed as algorithms, can be evaluated in environments that are reasonable approximations of today's financial markets. In the next Section, I describe how the freely-available open-source BSE market simulator has been extended to meet these needs.

5. BSE2: SIMULATING MODERN MARKETS

The BSE financial-market simulator (BSE 2012, Cliff 2018) has undergone major refactoring and extension, resulting in a "Version 2" of BSE, known as BSE2. The source code for this, written in Python, is being made freely available via the BSE GitHub repository.

BSE2 models current real-world financial markets in which some number *N* of profit-seeking legal entities each deploy proprietary automated trading systems. Each individual entity maintains an internal population of trading strategies, which it can select among for each order that is executed. In the simplest case, BSE2 can be configured so that each entity has an internal strategy population of size 1, so each entity is running a single specific strategy: this implements the S'62 style of modelling witnessed in the papers reviewed in Section 2.

To simulate a modern market scenario, each of the *N* entities simultaneously runs at least two trading strategies: one or more operational *prod* strategies which are to some extent trusted; and one or more *dev* strategies which are to some extent experimental improvements on that entity's current production strategies. The *N* populations of strategies, one population per entity, manifestly invite comparison with research studying meta-population dynamics. Taking the minimal case of each entity running only one prod and one dev strategy, as a first approximation the strategy innovation and improvement process within any one entity can be usefully modelled as a (1+1) Evolution Strategy (ES) optimization system: see e.g. Beyer & Schwefel 2002). The overall system is in fact *co-evolutionary* because the profitability (or, in the language of ESs, the *fitness function*) for any one entity's trading activity is defined at any one time largely by the nature of the *N-1* other entities' current trading strategies, which are simultaneously in the process of adapting. Preliminary simulation results studying co-adaptive dynamics in metapopulations of traders have been reported by Witchett (2004) using a version of the original ZIP market simulator (Cliff, 1997); and more recently by Hukerikar (2019) who worked with the original (Version-1) BSE studying the population dynamics of various simple trading strategies based on parameterized versions of ZIC, SHVR and GVWY, three of the built-in strategies in BSE.

BSE2 enables studying the evolutionary optimization of sets of parameter values for pre-established algorithms (see e.g. Cliff, 2009), operating in modern market settings; and could in further work be additionally extended to allow arbitrary trading strategies to be evolved and adapted via a genetic programming approach, revisiting early work such as that by Andrews

& Prager (1994), but in the modern market contexts provided by BSE2.

The various BSE2 source-code files can trivially be combined into a single Python script which can then readily be copied and launched across multiple virtual machines (VMs) from a commercial cloud service provider, so that many experiments can be conducted in parallel across a bank of VMs: the variability of market systems and the nature of the exhaustive testing required to establish rigorous results is likely forever more to require many millions of individual market sessions to be simulated, but this is a task that is *embarrassingly parallel* (i.e., when N i.i.d. simulation sessions are split across M machines, so that each machine is responsible for N/M simulated sessions, the speed-up factor is directly proportional to M).

The BSE2 GitHub repository, like the original BSE repository first created in 2012, offers not only the source-code for the BSE exchange simulator itself, but also the source-code for the trading algorithms ZIC, ZIP, GDX, and AA that were introduced in the review of key literature in Section 2, along with the source-code for simpler test strategies such as SHVR and GVWY. This served an important function of providing reference implementations for frequently-cited trading strategies, and is only notable because the authors of the papers introducing what were proposed (at the time) as world-leading algorithms such GDX and AA did not provide any sample code: the published descriptions of those two algorithms are both written only as narrative English text with occasional mathematical equations, and hence are open to some variation in interpretation. If the scientific study of trading strategies is to proceed smoothly, there is a need not only for a common and open freely-available up-to-date simulator for financial markets (as is now provided by BSE2) but also for freely-available open-source reference implementations of the major algorithms in the published literature (a point that was made forcefully over a decade ago by Toft, 2007), and of any other algorithms published in future.

6. SUMMARY

The BSE financial-market simulator (BSE 2012, Cliff 2018) has undergone major refactoring and extension, and is now known as BSE2. The BSE2 source-code is freely available on GitHub. BSE2 offers the facility not only for continuing to conduct experiments in the style of Smith (1962) but also allows trading strategies to be tested in simulations of modern-day markets where multiple trading entities, each in principle running more than one trading strategy at any one time, co-adapt via their interactions in the market. Because of the increased realism of BSE2, understanding the co-adaptive dynamics of market scenarios simulated in BSE2 is likely to help further our understanding of the dynamics of real-world financial markets, which are themselves inherently co-adaptive. Doing such work rigorously requires a shift in mind-set from the old view (prevalent in the key papers reviewed in Section 2) that a few tens of thousands of simulation sessions is sufficient to establish trusted results, to a new revised norm where it is routine for results to be published from tens or hundreds of millions of such experiments: the ready availability of cheap cloud computing services makes such an increase in CPU-cycles expended both practicable (because of the embarrassingly parallel nature of the simulations) and affordable (because of the economies of scale that have driven the development of commercial cloud service provisioning). As with its predecessor version, the BSE2 simulator has been made freely available via a standard MIT Open Source Licence on the GitHub public repository, and this includes not only the source-code for the exchange but also the source-code for various well-known trading strategies. The intention is that the BSE2 codebase becomes a common platform that is collectively refined and extended by the community of researchers interested in testing trading strategies in agent-based models of current real-world financial networks. If that does happen, then the effort expended in getting it this far will have been worthwhile.

ACKNOWLEDGMENTS

Many thanks to Francesco Longo for inviting me to write this position paper, and for his help with the deadline.

REFERENCES

Andrews, M., Prager, R., 1994. Genetic Programming for the Acquisition of Double Auction Market Strategies. in Kinnear, K. E. (d), Advances in Genetic Programming. MIT Press, pp.355-368.

Beyer, H.-G., Schwefel, H-P., 2002. Evolution Strategies: A Comprehensive Introduction. Natural Computing 1(1):3-52.

BSE, 2012. Bristol Stock Exchange. GitHub repo at https://github.com/davecliff/BristolStockExchange

Church, G., Cliff, D., 2019. A Simulator for Studying Automated Block Trading on a Coupled Dark/Lit Financial Exchange with Reputation Tracking. Proceedings EMSS2019.

Cliff, D., 1997. Minimal-Intelligence Agents for Bargaining Behaviours in Market-Based Environments. HP Labs Tech. Report HPL-97-91.

Cliff, D., Preist, C., 2001. Days Without End: On the Stability of Experimental Single-Period Continuous Double Auction Markets. Hewlett-Packard Labs Tech. Report HPL-2001-325.

Cliff, D. 2009. ZIP60: Further Explorations in the Evolutionary Design of Trader Agents and Online Auction-Market Mechanisms, IEEE Transactions on Evolutionary Computation. 13(1):3-18.

Cliff, D., 2018. An Open-Source Limit-Order-Book Exchange for Teaching and Research. Proceedings IEEE Symposium on Computational Intelligence in Financial Engineering (CIFEr); pp.1853-1860.

Cliff, D., 2019. Exhaustive Testing of Trader-Agents in Realistically Dynamic CDA Markets: AA Does Not Dominate. Proceedings ICAART2019.

Das, R., Hanson, J., Kephart, J., Tesauro, G., 2001. Agent-Human Interactions in the Continuous Double Auction. Proc. IJCAI-2001, pp.1169-1176.

De Luca, M., Cliff, D., 2011a. Agent-Human Interactions in the Continuous Double Auction, Redux. Proceedings ICAART2011.

De Luca, M., Cliff, D., 2011b. Human-Agent Auction Interactions: Adaptive-Aggressive Agents Dominate. Proceedings IJCAI-2011, pp.178-185.

De Luca, M., 2015. Adaptive Algorithmic Trading Systems. PhD Thesis, University of Bristol, UK.

Gjerstad, S., Dickhaut, J., 1997. Price Formation in Double Auctions. Games & Economic Behavior. 22(1):1-29.

Gjerstad, S., 2003. The Impact of Pace in Double Auction Bargaining. Working Paper, Department of Economics, University of Arizona.

Gjerstad, S., Shachat, J., 2007. Individual Rationality and Market Efficiency. Working Paper, Department of Economics, Purdue University.

Gode, D., Sunder, S., 1993. Allocative efficiency of markets with zero-intelligence traders. Journal of Political Economy, 101(1):119-137.

Hukerikar, N, 2019. Co-Adaptive Dynamics of Competitive Arms-Races Among Automated Trading Agents. Master's Thesis, Uni. of Bristol.

Le Calvez, A., Cliff, D., 2018. Deep Learning can Replicate Adaptive Traders in a LOB Financial Market. Proceedings of the IEEE Symposium on Computational Intelligence in Financial Engineering (CIFEr). SS1070, pp.1876—1883.

Miles B., Cliff, D., 2019. A Cloud-Native Globally Distributed Financial Exchange for Studying Real-World Trading-Latency Issues at Planetary Scale. Proceedings EMSS2019.

Rust, J., Miller, J., Palmer, R., 1992. Behavior of Trading Automata in a Computerized Double Auction Market. In Friedman, D., Rust, J. (eds) The Double Auction Market: Institutions, Theories, & Evidence. Addison-Wesley, pp.155-198.

Smith, V., 1962. An Experimental Study of Competitive Market Behavior. JPE 70(2):111-137.

Snashall, D., 2019. An Exhaustive Comparison of Algorithmic Trading Strategies: AA Does Not Dominate. Master's Thesis, Department of Computer Science, University of Bristol.

Snashall, D., Cliff, D., 2019 (in press). Adaptive-Aggressive Traders Don't Dominate. To appear in van den Herik, J., Rocha, A., Steels, L., (editors) Agents and Artificial Intelligence: Selected Papers from ICAART 2019. LNCS, Springer Verlag.

Stephens, C., Waelbroeck, H., 2009. Algorithm Switching: Co-Adaptation in the Market Ecology. The Journal of Trading, 4(3):59-73.

Stotter, S., Cartlidge, J., Cliff, D., 2013. Exploring Assignment-Adaptive Trading Agents in Financial Market Experiments, Proc. ICAART, 1:77-88.

Stotter, S., Cartlidge, J., Cliff, D., 2014. Behavioural investigations of financial trading agents using Exchange Portal (ExPo). In N. Nguyen et al. (eds), Trans. Computational Collective Intelligence XVII.

Tesauro, G., Das, R., 2001. High-performance Bidding Agents for the Continuous Double Auction. Proc. 3rd ACM Conf. on E-Commerce, pp.206-209.

Tesauro, G., Bredin, J., 2002. Sequential Strategic Bidding in Auctions using Dynamic Programming. In Proceedings AAMAS2002.

Toft, I., 2007. Adaptive Bidding Agents for Single Seller and Continuous Double Auctions. PhD Thesis, Computer Science, University of East Anglia, UK.

Vach, D., 2015. Comparison of Double Auction Bidding Strategies for Automated Trading Agents. MSc Thesis, Charles University in Prague.

Vytelingum, P., 2006. The Structure and Behaviour of the Continuous Double Auction. PhD Thesis, University of Southampton, UK.

Vytelingum, P., Cliff, D., Jennings, N., 2008. Strategic Bidding in Continuous Double Auctions. Artificial Intelligence, 172(14):1700-1729.

Witchett, D., 2004. Coadaptive Dynamics of Minimal Intelligence Trading Agents. Master's Thesis, University of Birmingham, UK.

AUTHOR'S BIOGRAPHY

Dave Cliff is a full Professor of Computer Science at the University of Bristol, UK, Before that, he held professorial posts at the University of Southampton, UK, and at the MIT Artificial Intelligence Lab in Cambridge, USA. He also spent seven years working in industrial technology R&D at Hewlett-Packard Labs in Palo Alto USA and Bristol UK, and as a Director & Trader in the Foreign Exchange Complex Risk Group at Deutsche Bank in the City of London. For more than 20 years Cliff's research has focused on applications of artificial intelligence and machine learning in financial markets, and on issues in systemic risk and stability in networked financial systems. Cliff served as a lead expert advisor to the UK Government Office for Science in its two-year "Foresight" investigation *The Future of Computer Trading in Financial Markets*, for its *Blackett Review* of the UK FinTech Industry, and is a founder member the Academic Advisory Council on Data Analytics for the UK Financial Conduct Authority, the main regulator of financial services activity in the UK. He is a Fellow of the British Computer Society and a Fellow of the Royal Society of Arts.

AGENT-MODEL BASED ON FLAME GPU FOR ASSESSING PUPAL PRODUCTIVITY OF THE TRANSMITTING VECTOR OF AEDES AEGYPTI INFECTIOUS DISEASES

Erica Montes de Oca[(a)], Remo Suppi[(b)], Laura De Giusti[(c)], Marcelo Naiouf[(d)]

[(a),(c),(d)]Computer Science Research Institute LIDI (III-LIDI), School of Computer Science, National University of La Plata, Scientific Research Agency of the Province of Buenos Aires (CICPBA), La Plata, Buenos Aires, Argentina
[(b)]Universitat Autònoma de Barcelona, Department of Computer Architecture and Operating Systems, School of Engineering, Campus Bellaterra, Cerdanyola del Vallès, Barcelona, 08193, Spain

[(a)]emontesdeoca@lidi.info.unlp.edu.ar, [(b)]remo.suppi@uab.cat, [(c)]ldgiusti@lidi.info.unlp.edu.ar
[(d)]mnaiouf@lidi.info.unlp.edu.ar

ABSTRACT
Dengue, Zika and chikungunya are among the infectious diseases that have emerged in recent years. The common denominator to these three is their transmitting vector: the Aedes aegypti mosquito. Due to sanitary reasons, it is highly important that the vector for transmitting these diseases be controlled through the implementation of strategies specifically designed for each situation. In this article, the creation of an agent-based simulation model that allows assessing control strategies and policies through parallel computing on GPU is proposed. High Performance Computing is necessary due to the large volume of data that has to be processed (hundreds of thousands of agents) to obtain results within an acceptable time frame. Model validation was done at small scale with an analogous model on CPU and NetLogo and using data from an real system. In this article, the implementation, scalability and potential of this model as decision support system (DSS) are presented.

Keywords: Infectious Diseases, Agent-Based Model, GPU, FLAME GPU

1. INTRODUCTION

The influence of human activity on the environment has resulted in the degradation of nature. The relation between climate change and the effects on human health are evident in the proliferation of infectious diseases on different locations around the globe (Barba-Evia 2016).
The World Health Organization (WHO) declared a public health emergency situation, describing it as "an extraordinary event which is determined [...] to constitute a public health risk to other States through the international spread of disease; and to potentially require a coordinated international response" (Maguiña-Vargas 2016). Dengue, Zika and chikungunya are among the infectious diseases that have emerged in recent years. The spread of these diseases has become a complex situation with high mortality levels (Álvares, Torres, Torres, Semper, Romeo 2018). The three diseases are transmitted through the Aedes aegypti vector, a mosquito species typical of tropical regions. Due to its adaptability, it has spread all over the globe through commercial and tourist routes. Currently, the three diseases that are transmitted by this mosquito are among those of greatest concern for public health (López-Latorre, Neira 2016).
Due to the importance for public health of controlling this vector, control strategies that are specifically designed for each particular situation are required. To this end, it is highly important to know mosquito populational characteristics and how they propagate (Albrieu-Llinás, Chiappero, Rodán-Dueñas, Gardenal 2016).
The modelization of complex systems through classic models based on continuous and derivable functions is not enough to describe the complexity of their components and relations. For this reason, computational models have gained relevance as a solution to study and research living entities systems (Ginovart 2015).
One of the tools that can assist in the decision-making process and help prevent diseases is simulation. Complex-system Agent-Based Modeling (ABM) has emerged to simulate the collective behavior of individuals in different scenarios, allowing researchers to study their characteristics in a short period of time (Izquierdo-Espinosa 2016). Agents are autonomous individuals that are familiar with the environment where they are, which allows them to interact with other agents. Agent behavior is determined by their inner state, a set of data that allows them to store information and act through temporary and historic perceptions. The environments where agents interact can be cities, towns, or any location in the real world that the researcher wants to model (López 2017). This allows this type of model to know system behavior through agent characteristics and behaviors as well as understanding

how the system itself affects the individuals it contains (Cantergiani, Gómez-Delgado 2016). Agents control their own behavior, but this behavior is not without any limits, but conditioned by the behavior of other agents. Agents are social actors with heuristic reasoning based on simple rules, and their behavior can be modified through learning from previous experiences. This means that they have limited memory to remember interactions and the results of their actions and strategies, and they use these to decide on new actions.

The amount of information generated through ABMs requires enough computational power in all its stages – screening, analysis and visualization. Simulating most complex problems requires using parallel or distributed solutions to obtain results within an acceptable time frame.

In recent decades, the advance of Graphics Processing Units (GPUs) in general-purpose programming (GPGPU), has allowed speeding up numerous applications that can be adapted to this architecture (Montes de Oca, De Giusti, De Giusti, Naiouf 2018). GPU is a *manycore* architecture whose approach is based on running parallel applications, with core number being doubled in each new generation. These devices have evolved and increased the calculation power in each floating point per second (initial design feature), which allows carrying out mass calculations of this type and in parallel. This parallel architecture has become one of the most widely accepted architectures by the scientific community, since its monetary cost is acceptable in relation to its computation power for the development of applications with high calculation demands (Wu, Deng, Jeon 2018).

GPU programming was benefited by the launch of CUDA (Compute Unified Device Architecture) by Nvidia in 2007, which freed programmers from having to consider parallel expressions (Aguilera, Silva-Aceves, Torres-Arguelles, Martínez-Gómez, Bravo-Martínez 2018). However, to exploit GPU performance to its maximum, programmers need to be familiar with the underlying architecture and optimize it as relevant, for instance, by using shared memory to reduce global memory latency time, kernel synchronization at the host, etc. (Nvidia 2019).

The implementation of a complex system such as the one presented here requires great dedication and effort to develop the ABM simulation kernel. Therefore, we decided to use the FLAME GPU framework (Chimeh, Richmond 2018). With this framework, we can disregard parallel simulation and focus on the simulation model for infectious diseases transmitting vector reproduction, obtaining results within an acceptable time frame.

FLAME GPU is a framework that allows simulating real-world individuals or objects into virtual agents. It is an extension of the FLAME (Flexible Large-scale Agent-based Modeling Environment) version. Its main goal revolves around expanding framework features to allow modeling GPU agents. It is specifically designed for high-performance parallel architectures, and it can create models with a massive number of agents (Richmond, Chimeh 2017). Agents are defined as finite state machines (with memory). Agents communicate via messages. Agent behavior is reflected on the functions it can perform, which change the agent's internal memory through inbound or outbound messages (Heywood, Maddock, Casas, García, Brackstone, Richmond 2018). Each agent's memory is permanent for each step of the simulation, but its message list is not. This provides internal communications that reflect the global behavior of the virtualized world.

The model to simulate is specified using a format called X-Machine Markup Language (XMML). It is about using XML to describe the agent and its behavior, the list of messages used by agents to communicate to each other, and the resulting control flow of the simulator (Chimeh, Heywood, Pennisi, Pappalardo, Richmond 2018).

The model defined in XMML generates the code for a set of templates through Extensible Stylesheet Transformations (XSLT). From those templates, a set of API dynamic templates is generated and added to agent features to produce the simulator (FLAME GPU 2018).

2. BACKGROUND

Aedes aegypti is a species of mosquito originally form Africa; it is believed that it used to breed inside tree trunk holes in the woods in the north of this continent. A period of intense drought forced the species to migrate to populated areas, favoring its fast adaptation in urban areas (Ruíz-López, Gonzalez-Maso, Vélez-Mira, Gómez, Zuleta, Uribe, Vélez-Bernal 2016). This adaptation to anthropic environments is due to the availability of the necessary food for their development and survival (Rossi, Almirón 2004). Currently, it has spread all around the globe due to commercial and tourist activities, coupled to the fact that mosquito reproduction control and elimination health programs are inefficient and there is an accelerated and non-planned growth of the population (Álvares, Torres, Torres, Semper, Romeo 2018).

The life cycle of Aedes aegypti has four stages: egg, larva, pupa and adult. Female mosquitoes need to take blood to make the eggs. They prefer low luminosity hours or night time to deposit their eggs in containers with water. Once the eggs are submerged in water, depending on temperature and humidity conditions, they hatch on the second or third day. They are resistant to drying out and can survive for months or years in dried out containers (Byttebier, De-Majo, Fischer 2014).

Larvae are aquatic and feed on the organic matter found in the container, including larvae from other species (or even their own). After fourteen days, at temperatures between 25ºC and 29ºC, larvae transform into pupae,

and remain at this stage for two to three days. In this state, they do not eat and use the energy accumulated during the larva period to transform. Finally, they experience anatomic and physiological changes and emerge as winged, dimorphous adults that seek humid places where there are no wind drafts. After 24 hs, they can already mate, and can live between 35-40 days (Rossi, Almirón 2004; Valendia-Romero, Olano, Coronel-Ruíz, Cabezas, Calderón-Peláes, Castellanos, Matís 2017).

The different types of containers they can use to breed depend on water supply, availability of food for the larvae, exposure to sunlight, and whether the container has a cover on top (Ngugi, Mutuku, Ndenga, Musunzaji, Mbakaya, Aswani, Irungu, Mukoko, Kitron, LaBeaud 2017).

When an adult mosquito bites, before it sucks any blood, it injects its own saliva, which contains a mix of anesthetic, blood thinners and histamine so that the host does not detect the bite and the blood does not coagulate, helping the mosquito get more of it and faster by reducing the amount of time required to be in contact with the host (Rossi, Almirón 2004).

Mosquitoes acquire the infection when they feed on a viremic person (Albrieu-Llinás, Chiappero, Rodán-Dueñas, Gardenal 2016). The virus is transmitted only through the mosquito bite, not from one individual to another. However, there have been some cases of Zika virus where transmission occurred through the sexual act or during gestation (from the mother to the fetus) (Gorodner 2016).

Aedes aegypti uses a large number of natural or artificial containers as habitat for its larvae. Some of these containers are more productive than others, meaning that vector control efforts should be aimed at eliminating the most productive ones, since these are the ones with higher epidemiological relevance. These strategies are correlated both vector local ecology and resident habits and attitudes in relation to containers (WHO 2009).

Any containers manufactured by man that are suitable for the development of mosquitoes in their non-mature stages are called "artificial microenvironments." Their distinctive characteristics are (Grech, Ludueña-Almeida 2016):

- Small size in comparison with natural environments such as swamps, irrigation channels, retention ponds, etc;
- They support a low number of species with reduced populational sizes;
- They do not generate organic matter;
- They are temporary environments with a lower frequency of predators.

The World Health Organization (WHO) recommends surveillance of the mosquito population as a method for analyzing and assessing prevention and control actions for the diseases transmitted by this vector (Cromwell, Stoddardt, Barker, Van-Rie, Messeer, Meshnick, Morrison, Scott 2017). Entomological surveillance is used for operational and research ends, to determine the following (WHO 2009):

- Vector geographical distribution;
- Obtaining measurements about vector population;
- Facilitating appropriate and timely decisions in relation to interventions.

There are several surveillances methods, and using each of them depends on the availability of skills and resources and, oftentimes, the level of infestation. The WHO recommends a sampling frequency of "weeks to months." Sampling methods vary based on the stage in the life cycle of the mosquito (Cromwell, Stoddardt, Barker, Van-Rie, Messeer, Meshnick, Morrison, Scott 2017). The most common surveillance methodologies use sampling procedures for larvae or pupae, rather than capturing eggs or adult mosquitoes. The basic sampling unit is the household, where searches are carried out to detect containers filled with water. Typically, a laboratory analysis is required to confirm the species. To regulate infestation levels, the following indexes are normally used:

- Household Index (HI): percentage of houses infested with larvae and/or pupae.
- Container Index (CI): percentage of water-holding containers infested with larvae and/or pupae.
- Breteau Index (BI): number of positive containers per 100 houses inspected.
- Pupae Per Container Index (PCI): total number of pupae for every 100 containers inspected.

The main recommendation for dengue campaigns is eradicating the most productive breeding environments based on the use of local evidence of pupal productivity in the containers. This strategy has been successfully used in many countries (Villegas-Trejo, Che-Mendoza, Gonzalez-Fernandéz, Guillermo-May, Gonzalez-Bejarano, Dzul-Mansilla, Ulloa-Gracia, Danis-Lozano, Manrique-Saide 2011).

2.1. Classic Epidemiological Models

Numeric modelization is a useful and pretty accurate tool that allows studying, analyzing and drawing conclusions about the system being considered, which allows making decisions towards controlling epidemic outbreaks (Medina-Arce, Ramos-Tapias 2017). The most commonly used models are compartment-based ones. Compartments represent a region with a group of evenly distributed individuals. While the individuals are static, it is the disease that moves in space through states (López 2017), which are based on the states through which the individuals go through. Available literature describes several mathematical models aimed at simulating infectious disease propagation. In relation

to the transmission of dengue, Zika and chikungunya, the following references can be found, for example:

- Ross-MacDonald model, based on ordinary differential equations, describing the dynamics for the relation between mosquitoes and humans, to model dengue disease in Cali, Colombia (Sepúlveda-Salcedo, Vasilieva, Martínez-Romero, Arias-Castro 2015);
- Mathematical model that uses ordinary differential equations to describe the evolution in time of human and mosquito populations in relation to the chikungunya virus in Guatemala (Ponciano, Chang, Quiroga 2018);
- Compartment-based mathematical model on a subpopulational network focused on dengue, Zika and chikungunya diseases (Anzo-Hernández, Velásquex-Castro, Bonilla-Capilla, Soto-Bajo 2018).

2.2. Agent-Based Epidemiological Modeling

Agent-based models are models that can be implemented on computers and allow describing problems whose analytical solution results are difficult to interpret or problems where finding a solution takes too much time (Rodriguez-Zoya, Roggero 2015). They have been widely used by the scientific community to understand how diseases are transmitted. References of epidemiological models based on agents developed for GPU architecture were gathered for this work; the most relevant are:

- ABM to simulate an activated sludge reactor on GPU (Pereda-García 2014);
- ABM modeling of an immune system using the FLAME GPU framework (Tamraker 2015);
- Implementation of a hybrid model using Verlet integration method and ABM to simulate agents involuntary and intentional interactions within a given mass; simulation results being presented on a GTX 750 GPU (Gutierrez-Milla, Borges, Suppi, Luque 2016);
- ABM for Reynolds' Boids simulation using GPU; a comparison with already implemented models is made. The results obtained show great speedup and a better understanding of agent behavior (Hermellin, Michel 2016);
- ABM that allows finding a treatment for vocal chord inflammation. The ABM proposed is run on GPU and, specifically in this work, a 3D visualization feature is added to the model (Seekhao, Jaja, Mongeau, Li-Jessen 2017);
- Simulation architecture called ParaCells, used to model biological problems. It uses the concepts of ABM and leverages the parallelism in GPU (Song, Yang, Lei 2018);
- Simulation of the pilgrimage to Makkah using an underlying model with agents, implemented with CUDA for GPU (Majid, Hamid, Rahiman, Zafar 2018).

There were no references in the literature reviewed for this work to models similar to the one presented here implemented on a high-performance architecture, based on GPU.

3. MODELING AND SIMULATION

In this article, an agent-based model is proposed for assessing the pupal productivity of Aedes aegypti mosquito. The model considers mosquito reproduction, which is dependent on container productivity. Results obtained in (Borges, Gutierrez-Milla, Suppi, Luque, Brito-Arduino 2015) show the effects on infectious disease propagation if container productivity is considered when developing prevention and control actions. As a first approximation, (Borges, Gutierrez-Milla, Suppi, Luque, Brito-Arduino 2015) developed a model on NetLogo; however, given the characteristics if this environment (based on Java), obtaining results for a large number of individuals and different scenarios within an acceptable time frame is not possible. To overcome these issues, a new agent-based model that can be run on a high-performance architecture, based on the FLAME GPU environment, is proposed. This new model, suitable for high-performance parallel environments, will allow modeling larger areas and simulate different scenarios within acceptable time frames. In the following sections, both approximations are presented, and the results obtained with the new version of the model proposed here are discussed.

3.1. Model on NetLogo

In (Borges, Gutierrez-Milla, Suppi, Luque, Brito-Arduino 2015), an agent-based model that allows assessing the pupal productivity of Aedes aegypti mosquito species was proposed. It is based on the control actions proposed by the WHO, considering the traditional method proposing division into areas, called strata, containing houses and breeding sites for Aedes aegypti. The mosquito gets the infection if it bites an infected individual and then, when it bites another individual who is not infected, it may transmit the disease. Only female mosquitoes can bite; they use the blood they suck to lay their eggs. Considering container productivity can help sanitation agents come up with more efficient actions, since they can focus on removing those containers whose productivity is higher, reducing the population of Aedes aegypti and, therefore, minimizing the risk for contagion.

The model was validated with data obtained from the real system through a study carried out in São Sebastião, which is on the northern coast of the state of São Paulo, 220 Km away from the state capital (Brito-Arduino 2014). In this study, a percentage of container productivity was obtained through field work. Containers were classified into removable and non-removable. This percentage is compared with the

average number of pupae produced in each container in several simulation runs.

The main limitations of the model developed on NetLogo revolve around its inability to simulate large areas. In (Brito-Arduino 2014), the study area encompassed 400.4 km^2. NetLogo was able to simulate 30,600 m^2, which represents a very small fraction of the total. Also, NetLogo only allows sequential runs, which means that a very long time is required to reach results statistical stability, since this type of models has a very high computational demand when the number of individuals is increased.

3.2. Model on FLAME GPU

Even though the types of agents to be represented and their interactions are similar to those implemented on the NetLogo model, their development is completely different, since the language used for modeling and interaction is entirely different. The same as NetLogo, FLAME GPU allows creating multiple agents (with internal memory and a set of functions that represent their actions) that interact through messages.

The framework, using an XML file and a *functions.c* file (containing agent actions implementation), is responsible for creating all necessary files (with *.cu* extension) that will form the simulator.

The developer only has to focus on implementing the details that are specific to the problem to be modeled, since the framework not only translated the code into CUDA language, but it also carries out all necessary optimizations to achieve maximum performance for the underlying GPU architecture.

The model simulates three types of agents: people, mosquitoes and containers. Some important parameters in the model are: the number of people and/or mosquitoes initially infected, the sizes of the populations of people and mosquitoes, the number and types of containers, the mosquitoes life time (by stages), the time of incubation of the disease, the size of the virtual world, the number of days of simulation, among others. The mosquitoes, in the initial state, are in the egg-stage, after completing this period, go through the larva-stage and then to pupa-stage and adult-stage. An adult mosquito will move through the virtual world in a maximum of 100 meters from the container where its birth, looking for a person to bite. Then, the mosquito, will explore for containers within the radius of action to deposit eggs.

In Figure 1, shows the three types of agents in the model. The types of containers can be tires, vases, glass containers, plastic, or metal, among others. If an uninfected mosquito bites a person not infected, neither the person nor mosquito acquire any disease. If the person is infected and the mosquito is not infected, the person will infect the mosquito, and after the incubation period, when the mosquito bites a not infected person, this person, after an incubation period, will be infected. An infected mosquito will remain (all his life) infected, contributing to spread the disease.

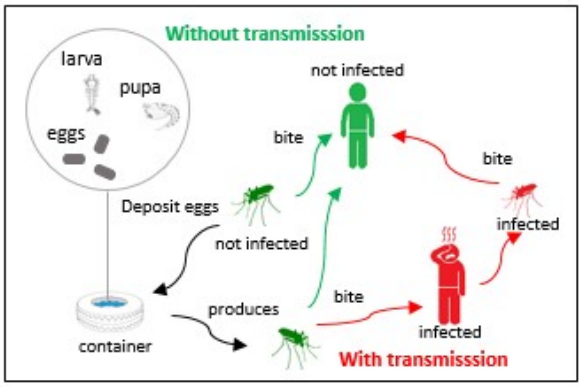


Figure 1: Model Agents and transmission cycles

The model can be configured to define the productivity of the containers. This productivity will determine the amount of eggs that a mosquito can put in each container. The productivity is used in the model to compute the pupal index. The pupal index (or the larval index) are used to estimate the number of mosquitoes that will reach the adult stage. The model can be configured to use any of these indices, however, the choice of the pupal index is more accurate since the mortality rate of the pupal stage is much lower than that of the larval state. On the other hand, if the productivity of the containers is not considered, the amount of eggs placed in each container will transform in a 100% of adult mosquitoes (worst case).

3.3. Experimental work and results

To verify the results of the model developed on FLAME GPU with the one implemented on NetLogo, 50 simulation runs were executed on both languages using the same number of individuals and model conditions. The number of pupae obtained in the experiments correspond to an initial setup of 300 mosquitoes and a total of 100 simulated days; the number of pupae in the container was established after each run and then averaged for the 50 runs. The models consider two possible scenarios: In the first scenario, the container has a certain productivity percentage that represents the number of non-mature stages of the mosquito that can reach adulthood. The second scenario does not consider container productivity – it allows 100% of the eggs laid in it to reach adulthood. Figure 1 shows the average of pupae per container for the 50 simulation runs carried out for each of the environments (bars) and their variance (*+ marks).

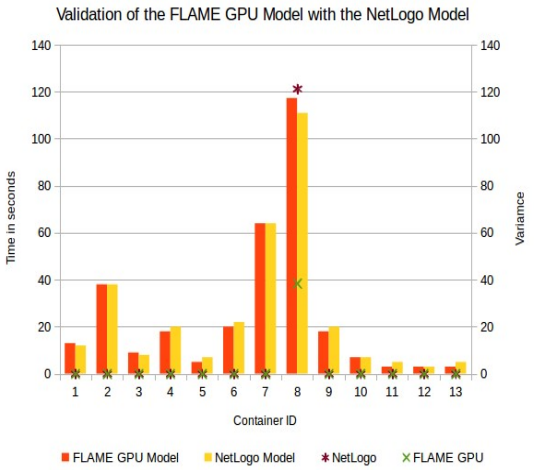


Figure 2: FLAME GPU vs. NetLogo. Average pupae per container for both models, and their variances

The number of eggs that a mosquito lays in a given container is randomly generated. It can be clearly seen that the model implemented on FLAME GPU yields results that are similar to those obtained with NetLogo, despite the fact that both models use different probability distributions for the generation of random numbers. NetLogo generates random numbers based on the normal distribution, whereas FLAME GPU uses (after an analysis of equivalences) a uniform distribution. Based on the results obtained, it can be stated that the probability distributions used are equivalent and they have no effect on result.

The variation in the number of pupae per container is reflected on the variance in each model. As it can be observed in Figure 2, the variances for both models converge. The peak observed with NetLogo in container 8 shows results variability in relation to the mean expected value; however, the average values obtained are similar in both environments.

All this confirms that the variability in the number of pupae per container is given by productivity, with randomness in the number of eggs that a mosquito can lay in each container being a contributing factor as well.

To validate this, the percentages of pupae per container obtained with these models were compared with those resulting from the field study detailed in (Brito-Arduino 2014). To obtain the percentage of pupae per container in (Brito-Arduino 2014), containers were inspected in the months during which mosquito eggs, larvae and pupae are produced, in consecutive seasons of vector proliferation, between 2002-2004.

As it can be seen in Figure 3, the largest difference between the compared results is 3.9% in container 8, followed by container 7 with a difference of 1.4%. In all remaining containers, this difference is smaller than 1%. These results show that the model proposed is viable and capable of yielding results that will be similar to those of a real system (model of reference in the fig. 2).

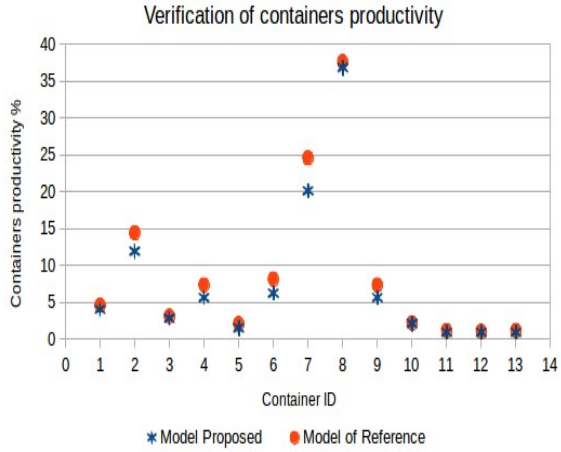


Figure 3: Comparison between pupal productivity in the containers of the real system (reference model) with the average percentage of pupae per container obtained with the model being proposed

After this validation, a new experiment was carried out, this time to show the advantages of the use of GPU as platform to run the model. Figure 4 shows the results in relation to time measurements for runs done on both models (average for the 50 runs). The initial number of mosquitoes was changed, simulating a total of 80 days. As it can be seen, for 30,000 mosquitoes, the simulation without-productivity could not be run on NetLogo, since the model exceeded memory capacity (Java Virtual Machine limitations).

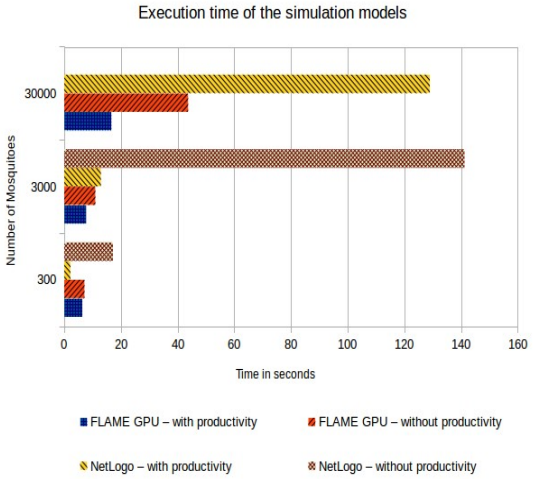


Figure 4: Measured execution time (in seconds) for both simulation models, increasing the initial number of mosquitoes in the simulation

It can be observed that the model developed on FLAME GPU, when the initial number of individuals is small, yields a higher execution time (considering container productivity) than that of the simulation run on NetLogo. In a GPU architecture, all cores must be working and if the number of thread is greater than the data to process, the GPU will distribute the work in order to avoid not idle threads. Therefore, the time-increase in the FLAME GPU model occurs because the

total number of individuals in the simulation is reduced due to container productivity. In this case, the use of a GPU architecture is not recommended. However, when the initial number of mosquitoes is increased, execution time increases, yielding better results when container productivity is not considered. The limitations of the model developed on NetLogo is given by the limitations that are inherent to the environment itself, since it limits the number of agents that can be used to experiment and, consequently, only very small areas can be simulated.

These results help consider the scalability of the model proposed on FLAME GPU in relation to the initial number of mosquitoes. Figure 5 shows how the models respond when container productivity percentage is changed, with a fixed initial number of mosquitoes (set to 3,000). To do this, the most productive containers (2, 7 and 8) were selected. The result is the average of the times measured, in seconds, for 50 simulation runs in both models, with the following settings: considering original productivity, Case A (12.6%, 21.5% and 32.9%, for containers 2, 7 and 8, respectively), increasing productivity for one container at a time (considering all possible combinations: Case B-1:50%, 21.5% and 32.9%; Case B-2: 12.6%, 50% and 32.9 %; and Case B-3: 12.6%, 21.5% and 50%, for containers 2, 7 and 8, respectively), increasing the productivity for two containers at the same time (with the following combinations: Case C-1: 50%, 50% and 32. 9%; Case C-2: 50%, 21.5% and 50%; Case C-3: 12.6%, 50% and 50%, for the three containers mentioned above, as applicable), and lastly, Case D, where productivity was increased to 50% for all three containers simultaneously.

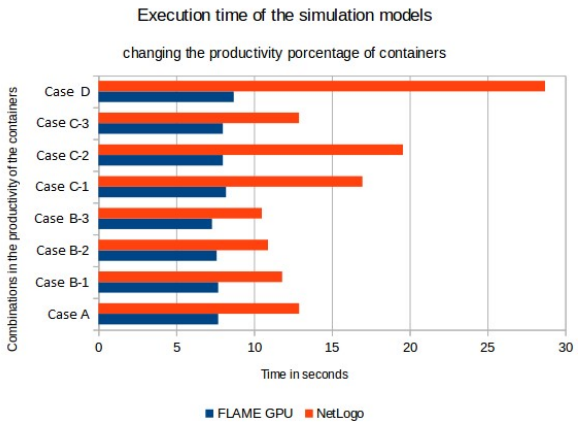


Figure 5: Measured execution time (in seconds) for both simulation models, increasing the productivity of containers 2, 7 and 8

Based on the runs carried out, the model on FLAME GPU is more stable in relation to changes in container productivity. The model on NetLogo presents an increase in the average simulation time (especially in these combinations: Case C-1: 12.6%, 50% and 50%, and Case C-2: 50%, 50% and 32.9%). Even though this increase is explained by container distribution in NetLogo, since container are close to each other and, as a consequence, they affect the number of pupae in other containers, this increase in simulation time is not proportional to the time measured for other types of containers.

4. CONCLUSIONS

The agent-based model developed with FLAME GPU offers clear advantages over its predecessor, implemented on NetLogo. With FLAME GPU, the developer can focus entirely on modeling the problem instead of having to devote time and effort to implement a complex system on GPU.

The implementation of an agent-based model on FLAME GPU to analyze how infectious diseases transmitted by Aedes aegypti spread exhibited an excellent behavior from the point of view of high-performance simulation and was successfully validated against a real system. The results obtained yielded values that are very close to the actual numbers of pupae per container produced by Aedes aegypti. Being able to reduce simulation time allow running larger simulations, be it in relation to the area to be simulated or to the number of individuals included in the simulation.

As open research lines and future work, the following topics of interest can be mentioned: a) assessing energy consumption for high-performance simulations on FLAME GPU; b) fine-tuning and validating the model using different geographical areas in Argentina; c) interactive visualization and adaptation to transform the model into an interactive tool for decision-making; d) migration to the Cloud to offer a high-performance simulation service to the scientific community.

ACKNOWLEDGMENTS

The authors would like to thank Francisco Borges for his contributions as developer of the first model on NetLogo, and especially Silvana Gallo for her collaboration with the implementation of the model on FLAME GPU. This research has been partially supported by the Agencia Estatal de Investigación, Spain and the FEDER-UE (contract TIN2017-84875-P) and collaboration agreement with the Fundación EUG.

REFERENCES

Barba-Evia J.R., 2016. Cambio Climático, globalización y su efecto sobre enfermedades infecciosas. ¿La fiebre por virus Ébola es una amenaza latente?, Rev. Latinoam Patol. Clin. Med. Lab. 63 (3), 124–132.

Maguiña-Vargas C., 2016. Zika la nueva enfermedad emergente en América. Rev. Med. Hered. 27 (1), 3-6.

Álvares E.M.C., Torres Á.A., Torres Á.A., Semper G.A.I., Romeo A.D., 2018. Dengue, Chikungunya, virus de Zika. Determinantes sociales. Rev. Med. Electórnica 40 (1), 120-128.

López-Latorre M.A., Neira M., 2016. La influencia del cambio climático en la biología de Aedes aegypti

(Diptera: Culicidae) mosquito transmisor de arbovirosis humanas. Rev. Ecuatoriana de Medicina Y Ciencias Biológicas 37 (2), 11-21.

Albrieu-Llinás G., Chiappero M.B., Rodán-Dueñas J.C., Gardenal C.N., 2016. Reconstrucción de una invasión: pasado y presente de poblaciones de Aedes (Stegomyia) aegypti en la Argentina. Preceedings of the X Jornadas Regionales sobre Mosquitos, 65-71. 15 y 16 de Septiembre de 2016, Mar del Plata, Buenos Aires, Argentina.

Ginovart M., 2015. ¿Qué pueden ofrecer los modelos basados en agentes en el contexto docente?. Modelling in Science Education and Learning 8 (2), 5-26.

Izquierdo-Espinosa G.V., 2016. Implementación de un modelo de agentes para estudiar la propagación del virus de la fiebre Chikungunya. Thesis (MD). Universidad San Francisco de Quito.

López L.R., 2017. Modelado de epidemias utilizando sistemas auto-organizados. Thesis (PhD). Universidad Nacional del Litoral.

Cantergiani C., Gómez-Delgado M., 2016. Diseño de un modelo basado en agentes para simular el crecimiento urbano el corredor del Henares (Comunidad de Madrid). Boletín de la Asociación de Geógrafos Españoles 70, 259-283.

Montes de Oca E., De Giusti L., De Giusti A., Naiouf M., 2018. Análisis de Consumo Energético en Cluster de GPU y MultiGPU en un problema de Alta Demanda Computacional. Preceeding of the XXIV Congreso Argentino de Ciencias de la Computación, 103-112. Tandil, Buenos Aires, Argentina.

Wu J., Deng L., Jeon G., 2018. Image autoregressive interpolation model using GPU-parallel optimization. IEEE Transactions on Industrial Informatics 14 (2), 426-436.

Aguilera F.J.E., Silva-Aceves J.M., Torres-Arguelles S.V., Martínez-Gómez E.A., Bravo-Martínez G., 018. Utilización de GPU-CUDA en el procesamiento digital de imágenes. CULCyT 15 (66), 65-79.

Nvidia, 2019. CUDA C Programming Guide. NVIDIA. Available from: https://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf [accessed April 2019].

Chimeh M.K., Richmond P., 2018. Simulating heterogeneous behaviours in complex systems on GPUs. Simulation Modelling Practice and Theory 83 (2018), 3-17.

Richmond P, Chimeh M.K., 2017. FLAME GPU: Complex System Simulation framework. Preceedings of the International Conference on High Perfomance Computing & Simulation, 11-17. Genoa, Italy

Heywood P., Maddock S., Casas J., García D., Brackstone M., Richmond P., 2018. Data-parallel agent-based microscopic road network simulation using graphics processing units. Simulation Modelling Practice and Theory 83, 188-200.

Chimeh M.K., Heywood P., Pennisi M., Pappalardo F., Richmond P., 2018. Paralell Pair-Wise interaction for Multi-agent inmune systems modelling. Preceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 1367-1373. Madrid, Spain

FLAME GPU, 2018. FLAME GPU Documentation Release. FLAME GPU. Available from: https://buildmedia.readthedocs.org/media/pdf/flamegpu/latest/flamegpu.pdf. [accessed April 2019].

Ruíz-López F., Gonzalez-Maso A., Vélez-Mira A., Gómez G:F:, Zuleta L., Uribe S., Vélez-Bernal I.D., 2016. Presencia de Aedes (Stegomyia) aegypti (Linnaeus, 1762) y su infección natural con el virus dengue en alturas no registradas para Colombia. Rev. Bomédica 36 (2), 303-308.

Rossi G.C., Almirón W.R., 2004. Clave ilustrada para la identificación de larvas de mosquito de interés sanitario encontradas en criaderos artificiales en Argentina. Serie Enfermedades Transmisibles. Buenos Aires, Argentina: Fundación Mundo Sano.

Byttebier B., De-Majo M.S., Fischer S., 2014. Low temperature hatching response of Aedes aegypti eggs: effects of hatching media and storage conditions. Journal of Medical Entomology 51 (1), 97-103.

Valendia-Romero M.L., Olano V.A., Coronel-Ruíz C., Cabezas L., Calderón-Peláes M.A., Castellanos J.E., Matís M.I., 2017. Detección del virus del dengue en larvas y pupas de Aedes aegypti recolectadas en áreas rurales del municipio de Anapoima, Cundinamarca, Colombia. Biomédica 37 (2), 193-200.

Ngugi H.N., Mutuku F.M., Ndenga B.A., Musunzaji P.F., Mbakaya J.O., Aswani P., Irungu N.W., Mukoko D., Kitron U., LaBeaud A.D., 2017. Characterization and productivity profiles of Aedes aegypti (L.) breeding habitats across rural and urban landscapes in western and coastal Kenya. Rev. Parasites & Vectors 10.

Gorodner J.O., 2016. Dengue, fiebre Zika y fiebre Chikungunya. Patologías conminantes y cambio climático en América. Rev. Asociación Médica Argentina 129 (1), 30-32.

WHO, 2009. Dengue: Guidelines for diagnosis, treatment, prevention and control. World Health Organization. Available from: https://apps.who.int/iris/bitstream/handle/10665/44504/9789995479213_spa.pdf;jsessionid=4E7C3EAE339679F1547C9B8CBE44CCB4?sequence=1 [accessed April 2019].

Grech M.G., Ludueña-Almeida F.F., 2016. Mosquitos que crían en microambientes artificiales. Proceedings of the X Jornadas Regionales Sobre Mosquitos, 142-155. Mar del Plata, Buenos Aires, Argentina.

Cromwell E.A., Stoddardt S.T., Barker C.M., Van-Rie A., Messeer W.B., Meshnick S.R., Morrison A.C., Scott T.W., 2017. The relationship between entomological indicators of Aedes aegypti

abundance and dengue virus infection. PLOS Neglected Tropical Diseases 11 (3).

Villegas-Trejo A., Che-Mendoza A., Gonzalez-Fernandéz M., Guillermo-May G., Gonzalez-Bejarano H., Dzul-Mansilla F., Ulloa-Gracia A., Danis-Lozano R., Manrique-Saide P., 2011. Control enfocado en Aedes aegypti en localidades de alto riesgo de transmisión de dengue en Morales, México. Salud Pública de México 53 (2), 141-151.

Medina-Arce Y., Ramos-Tapias J.A., 2017. Modelo matemático que explica mejor la afectación eidentifica el patrón relevante en la difusión para el dengue en la zona urbana del municipio de Neiva. Entornos 30 (2), 121-131.

Sepúlveda-Salcedo L.S., Vasilieva O., Martínez-Romero H.J., Arias-Castro J.H., 2015. Ross McDonald: Un modelo para la dinámica del dengue en Cali, Colombia. Revista de Salud Pública 17 (5), 749-761.

Ponciano J.A., Chang J.D., Quiroga F., 2018. Modelo epidemiológico para el estudio regional de la chikungunya. Ciencia, Tecnología y Salud 5 (1), 63-72.

Anzo-Hernández A., Velásquex-Castro J., Bonilla-Capilla B., Soto-Bajo M., 2018. Modelado y simulación de epidemias transmitidas por mosquitos en redes meta-poblacionales. Proceedings of the Congreso Nacional de Ingeniería Biomédica, 445-448. Ciudad de León, Guanajato, México.

Rodriguez-Zoya L.G., Roggero P., 2015. Modelobasado en agentes: aportes epistemológicos y teóricos para la investigación social". Rev. Mexicana de Ciencias Políticas y Sociales LX (225), 227-262.

Pereda-García M., 2014. Agent-based modeling and swarm intelligence in systems enfineering. Thesis (PhD). Universidad de Valladoid.

Tamraker S., 2015. Performance Optimization and Statistical Analysisof Basic Immune Simulator (BIS) Using theFLAME GPU Environment. Thesis (MD). University of Wisconsin–Milwaukee.

Gutierrez-Milla A., Borges F., Suppi R., Luque E., 2016. Crowd turbulence with ABM and Vertet integration on GPU cards. Procedia Computer Science 80, 2428-2432.

Hermellin E., Michel F., 2016. GPU Environmental delegation of Agent Perceptions: Application to Reynolds's Boids. In: Springer, eds. Multi-Agent Based Simulation XVI. Verlag Berlin, Heidelberg, Springer, 71-86.

Seekhao N., Jaja J., Mongeau L., Li-Jessen N.Y.K., 2017. Insitu visualization for 3D Agent-Based Vocal Fold inflammation and Repair Simulation. Supercomput Front Innov. 4 (3), 68-79.

Song Y., Yang S., Lei J., 2018. ParaCells: a GPU architecture for Cell-Centered models in Computational Biology. In: IEEE, eds. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 1-1.

Majid A.R.M.A., Hamid N.A.W.A., Rahiman A.R., Zafar B., 2018. GPU-based optimization of Pilgrim Simulation for Hajj and Umrah Rituals. Pertanika J. Sci. & Technol. 26 (3), 1019 –1038.

Borges F., Gutierrez-Milla A., Suppi R., Luque E., Brito-Arduino M., 2015. An Agent-Based Model for assessment of Aedes aegypti pupal productivity. Proceedings of the 2015 Winter Simulation Conference, 159-169. December 2015, Huntington Beach, CA, USA.

Brito-Arduino M., 2014. Assessment of Aedes aegypti pupal productivity during the Dengue Vector Control Program in a Costal Urban Centre of Sao Pulo State, Brasil. Journal of Insects 2014, 1-9.

AUTHORS BIOGRAPHY

ERICA MONTES DE OCA is Technician in High-Performance Computation and GRID Technology, fellow of the National University of La Plata (UNLP, Argentina) and is currently taking courses for her Doctorate in Computer Science. She is also a member of several research projects of the Computer Science Research Institute LIDI (III-LIDI). Her research is focused on the topics of GPU, GPU clusters, MultiGPU, Agent-based Modeling, High Performance Computing, High Performance Simulation, Green Computing, Energy Consumption and Performance.

REMO SUPPI received his diploma in Electronic Engineering form Universidad Nacional de La Plata (Argentina), and a PhD degree in Computer Science form Universitat Autonoma de Barcelona (UAB) in 1996. At UAB, he spent more than 25 years researching on topics including computer simulation, distributed, systems, high performance and distributed simulation applied to ABM or individual-oriented models. He has published several scientific papers on the topics above and has been an associate professor at UAB since 1997. He is also a member of the High Performance Computing for Efficient Applications and Simulation Research Group (HPC4EAS) at UAB.

LAURA DE GIUSTI has a PhD in Computer Science at Universidad Nacional de La Plata (UNLP, Argentina), and is Professor at the School of Computer Science of UNLP. In addition, she is a member of several research projects at the Computer Science Research Institute LIDI (III-LIDI) and has authored several research works published in national and international journals and conferences.

MARCELO NAIOUF is a Chair Professor at the Computer Science Research Institute LIDI (III-LIDI) at National University of La Plata, Argentina. Head of the PhD Program at the School of Computer Science, supervisor of 8 PhD theses, invited lecturer at various Universities in Argentina, and leader in several research

projects. His research interests include high performance computing and parallel algorithms. He has co-authored 3 books and more than 120 full-reviewed technical papers in journals and conference proceedings.

NONPARAMETRIC FREQUENCY POLYGON ESTIMATION FOR MODELING INPUT DATA

Stephen Hague[(a)], Simaan AbouRizk[(b)]

[(a),(b)]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada

[(a)]steve.hague@ualberta.ca, [(b)]abourizk@ualberta.ca

ABSTRACT

To construct valid probability distributions solely from input data, this paper compares three nonparametric density estimators: (1) histograms, (2) Kernel Density Estimation, and (3) Frequency Polygon Estimation. A pseudocode is implemented, a practical example is illustrated, and the *Simphony.NET* simulation environment is used to fit the nonparametric frequency polygon to a set of data to recreate it as a posterior distribution via the Metropolis-Hastings algorithm.

Keywords: nonparametric density estimation, input modeling, frequency polygon

1. INTRODUCTION

Input modeling in simulation studies can be divided into two broad approaches. A classic approach, whereby standard statistical distributions are used to model underlying input data using a standard approach of:

1) Selecting one of the standard statistical distributions,
2) Parameterizing the distribution, (e.g., using the method of moments or the method of maximum likelihood),
3) Examining the goodness-of-fit (e.g., using a standardized test such as the Kolmogorov-Smirnov or Chi Square tests), and
4) Repeating as necessary until an acceptable fit is found.

Another approach for input modeling is to use nonparametric modeling techniques to construct valid probability distributions directly by defining a probability density function (PDF) and cumulative distribution function (CDF) solely from input data.

The advantages of the classical approach are numerous:

- A wide array of probability distributions exists.
- Efficient algorithms for the evaluation of the PDF and CDF, as well as the generation of random deviates, are readily available for a variety of platforms.
- Storage requirements are minimal as, once fit to the data, only parameters are required for simulation.
- Many random processes are known to follow certain distributions.

While these advantages have resulted in the widespread use of the classical approach for input modeling, certain datasets (e.g., multi-modal, Monte Carlo simulation outputs and Markov chain Monte Carlo (MCMC) algorithm outputs) are not well-suited for this approach. Multimodal data sets are a good example of what happens when classical approaches become limited.

Consider the example adapted from Scott (1985), in which 400 samples are generated from the bimodal mixture density

$$\frac{3}{4}Normal(0.00, 1.00) + \frac{1}{4}Normal(1.75, 0.25) \quad (1)$$

where 75% of the samples are generated from a normal distribution with a mean of 0 and a standard deviation of 1, and the remaining 25% of the samples are generated from a normal distribution with a mean of 1.75 and a standard deviation of 0.25. A histogram and CDF of the samples versus the theoretical distribution are illustrated in Figure 1 (*top left and right, respectively*).

The bimodal nature of the distribution renders the fitting of a standard probability distribution difficult in practice. Indeed, using the method of maximum likelihood, a triangular distribution with a low value of −2.9769, a high value of 2.4815, and a most likely value of 1.7790 was selected; see Figure 1 (*bottom left and right*).

Despite the triangular distribution being selected as the best fit, the fit was rejected when tested using the Kolmogorov-Smirnov goodness-of-fit test (Table 1), demonstrating that the classical input modeling approach is not well-suited for modeling this dataset.

Nonparametric input models are not limited in form the same way that classic statistical distributions are. The main limitations are related to 1) sampling during simulation studies from distributions that have been identified as nonparametric, and 2) using nonparametric models when data used for characterizing the input model are not sufficient to properly identify the underlying distribution of a given phenomenon (e.g. when limited data are collected to model machine breakdowns and have been known to follow an exponential distribution, but do not reveal such).

In this paper we discuss how to address the first limitation.

2. NONPARAMETRIC DENSITY ESTIMATORS

Nonparametric density estimators construct a density function directly from a given set of samples $\{x_1, \cdots, x_n\}$. Three commonly applied methods are compared here, namely: the histogram, kernel density estimation, and the frequency polygon.

2.1. Histograms

The most basic nonparametric estimator is the histogram, which is constructed by choosing an origin, x_0, and a bin width of $h > 0$. The bins of the histogram are the intervals

$$[x_o + mh, x_0 + (m+1)h) \quad m \in Z \tag{2}$$

and the histogram itself is defined by

$$\hat{f}(x) = \frac{1}{nh}(number\ of\ x_i\ in\ same\ bin\ as\ x) \tag{3}$$

(Silverman 1986). The choice of bin width h is subjective, greatly affecting the usefulness of the resulting histogram. Readers are referred to Scott (1979) for a detailed discussion of the challenges associated with the use of histograms.

2.2. Kernel Density Estimation

A well-known improvement of the histogram is the kernel density estimate, based on the concept of Rosenblatt (1956) and Parzen (1962). Intuitively, the kernel density estimate surrounds each data point in a sample with a small "bump" of density. The density estimate is the sum of these "bumps."

The kernel density estimator is defined as

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4}$$

where $h > 0$ is a smoothing parameter—known as the bandwidth—and K is a non-negative function—known as the kernel—that integrates to one.

The bandwidth, h, is typically chosen to be as small as the data allow. The kernel is commonly represented by the density function of a probability distribution, including any one of the following:

- Uniform(-1, 1);
- Triangular(-1, 0, 1);
- Beta(2, 2, -1, 1), also referred to as Epanechnikov or parabolic; or
- Normal(0, 1).

Using a bandwidth of h = 0.12 and Normal(0, 1) as the kernel, the kernel density estimator for the data sampled from Equation 1 and the corresponding CDF are illustrated in Figure 2 (*top left and right, respectively*).

In contrast to the triangular distribution selected using the classical approach, the CDF generated using the nonparametric kernel density approach resulted in an acceptable fit (Table 1) when examined using the Kolmogorov-Smirnov test.

In cases where the kernel is a probability density, generation of a random deviate from the kernel density estimator is straightforward. First, a random deviate, u,

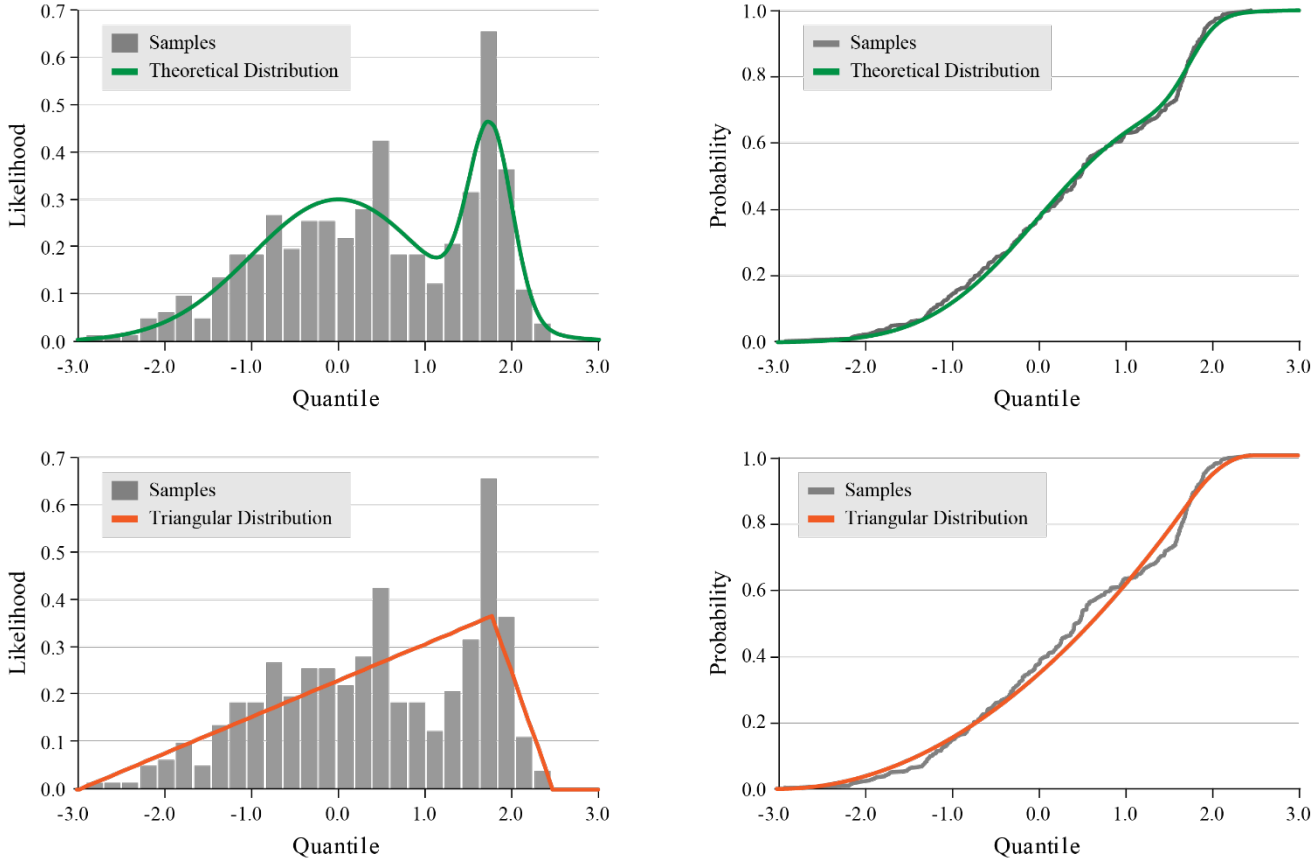


Figure 1: Histograms (*left*) and Cumulative Distribution Functions (*right*) of Generated Samples versus Theoretical Distribution (*top*) and Triangular Distribution (*bottom*).

is generated from the kernel. Next, an element x_j is randomly chosen from the sample set $\{x_1, \cdots, x_n\}$. Then

$$hu + x_j \tag{5}$$

is a random deviate from the kernel density estimator.

2.2.1. Limitations of Kernel Density Estimation

Although kernel density estimation is available in a variety of software packages, including *R* (R Core Team 2019) and MATLAB (MathWorks 2019), its functionality is limited in conditions where n is large because (i) evaluation of the PDF and CDF becomes computationally intensive, as the kernel must be evaluated at every sample point, and (ii) the estimate requires considerable amounts of storage, as every sample point must be available.

2.3. Frequency Polygon Estimation

The frequency polygon can be thought of as a generalization of the triangular distribution. Its PDF is constructed "from a histogram by connecting with straight lines the mid-bin values of the histogram" (Scott 1985). Its CDF is then a piecewise quadratic function.

As with the histogram, the frequency polygon is dependent on the choice of bin width ($h > 0$). Scott (1985) notes that the bin width for an optimal frequency polygon will generally differ from that of the optimal histogram.

Using a bin width of $h = 0.21$, the frequency polygon for the data sampled from Equation 1 and the corresponding CDF are illustrated in Figure 2 (*bottom left and right, respectively*). As with the kernel density estimate, the Kolmogorov-Smirnov test determined that the CDF generated using the frequency polygon resulted in an acceptable fit (Table 1).

Table 1: Results of the Kolmogorov-Smirnov Test

	Test Statistic	Fit*
Triangular Distribution	0.07501	Rejected
Kernel Density Estimator	0.02861	Accepted
Frequency Polygon	0.02353	Accepted

*At significance α = 0.05, with critical value = 0.06791.

Once constructed, computation of the frequency polygon does not require the availability of the original data points, rendering it efficient even in conditions when n is large.

3. PSEUDO CODE IMPLEMENTATION

For the pseudocode conventions used herein, see Cormen et al. (2009). From an implementation perspective, the frequency polygon consists of three

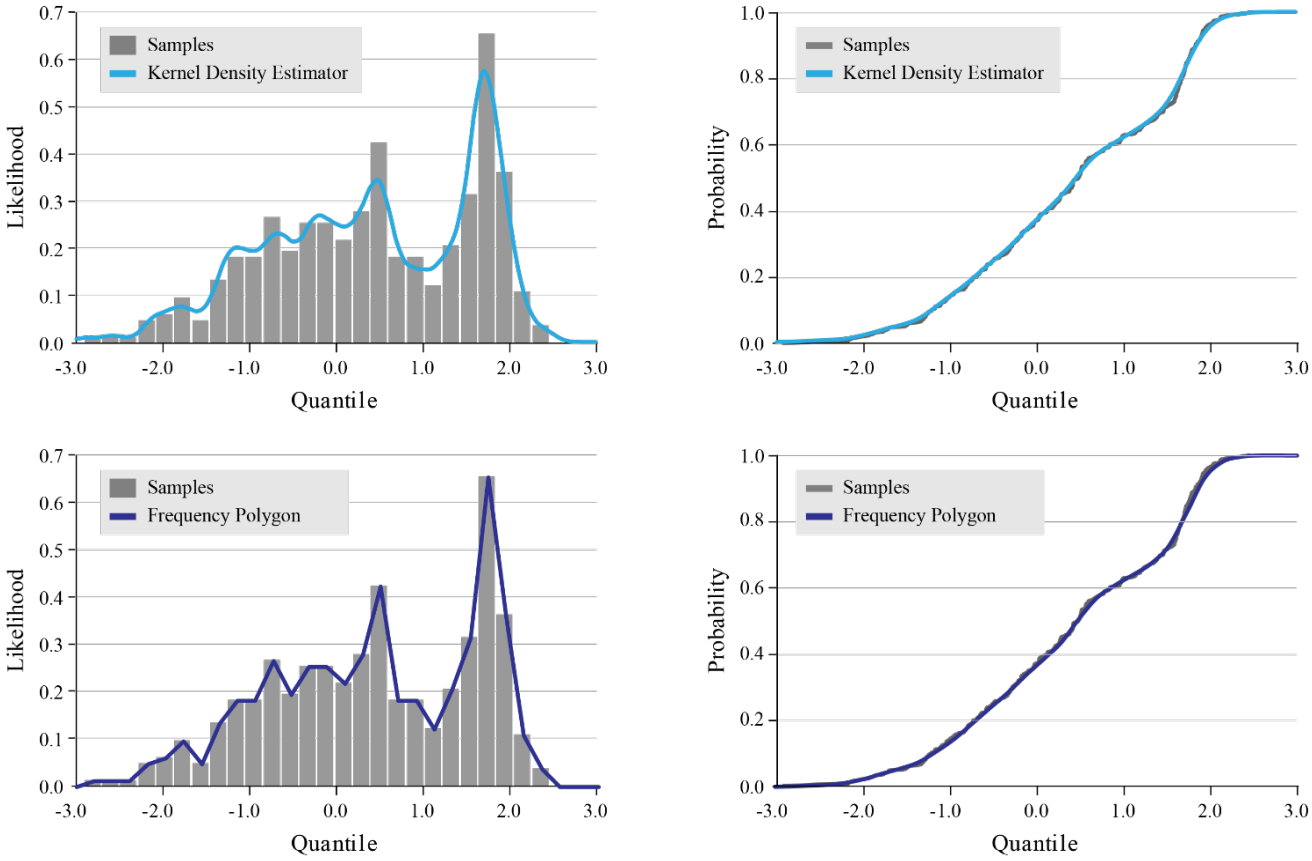


Figure 2: Histograms (*left*) and Cumulative Distribution Functions (*right*) of Generated Samples versus the Kernel Density Estimator (*top*) and Frequency Polygon (*bottom*).

arrays:

$$X[0 \cdots k+1]$$
$$L[0 \cdots k+1] \quad (6)$$
$$C[0 \cdots k+1]$$

where k is the number of bins in the original histogram. The array X is defined by

$$X[i] = \text{mid-point of bin } i \qquad 0 \le i \le k+1 \quad (7)$$

the array L (the PDF values corresponding to the entries in X) is defined by

$$L[i] = \begin{cases} 0 & \text{if } i = 0 \\ \text{mid-point value of bin } i & \text{if } 1 \le i \le k \\ 0 & \text{if } i = k+1 \end{cases} \quad (8)$$

and the array C (the CDF values corresponding to the entries in X) is defined recursively by setting $C[0] = 0$ and

$$C[i] = C[i-1] + \frac{h}{2}(L[i-1] + L[i]) \quad (9)$$

for $i > 0$, where h is the bin width of the original histogram.

Note that for persistence purposes, only arrays X and L need to be stored, as C can be reconstructed if necessary.

3.1. Binary Search Procedure

The existence of a procedure, termed BINARY-SEARCH(x, T, p, r), is then assumed. This procedure takes a key, x, together with a sorted subarray, $T[p \cdots r]$, and returns one of the following:

- If $T[p \cdots r]$ is empty $(r < p)$, then the index p is returned.
- If $x \le T[p]$ and, therefore, less than or equal to all of the elements of $T[p \cdots r]$, then the index p is returned.
- If $x > T[p]$, then the largest index q in the range $p < q \le r+1$ is returned, such that $T[q-1] < x$.

Example pseudocode for a binary search procedure is detailed in Cormen et al. (2009).

3.2. Initialize Procedure

In this procedure, the data are first binned, and the values for the three arrays are then constructed. Specifically, the INITIALIZE procedure initializes the X, L, and C arrays given the array of data T, to which the polygon is to be fit.

```
INITIALIZE(T, X, L, C)
 1 determine the number of bins (k)
 2 initialize X[0 ⋯ k+1] as a new array
 3 initialize L[0 ⋯ k+1] as a new array
 4 initialize C[0 ⋯ k+1] as a new array
 5 min = min(T)
 6 max = max(T)
 7 h = (max − min)/k
 8 for i = 0 to T.length − 1 // bin the data
 9     j = min(⌊(T[i] − min)/h⌋ + 1, k)
10     L[j] = L[j] + 1
11 for i = 0 to k + 1 // construct the arrays
12     X[i] = min + h(i − 0.5)
13     L[i] = L[i]/T.length/h
14     if i > 0
15         C[i] = C[i − 1] + h(L[i − 1] + L[i])/2
```

3.3. Slope Procedure

The SLOPE procedure is a helper method that calculates the slope of the PDF line segment between $X[i]$ and $X[i+1]$.

```
SLOPE(X, L, i)
 1 return (L[i + 1] − L[i])/(X [i + 1] − X [i])
```

3.4. Probability-Density Procedure

The PROBABILITY-DENSITY procedure evaluates the PDF of the frequency polygon for real argument, x. It begins by determining the largest index, i, such that $X[i] < x$, and then uses linear interpolation to calculate the value of the PDF at x.

```
PROBABILITY-DENSITY(X, L, x)
 1 if x ≤ X[0] or X[X.length − 1] ≤ x
 2     return 0
 3 else
 4     i = BINARY-SEARCH(x, X, 0, X.length − 1) − 1
 5     m = SLOPE(X, L, i)
 6     h = x − X[i]
 7     return L[i] + mh
```

3.5. Cumulative-Distribution Procedure

This procedure evaluates the CDF of the frequency polygon for real argument, x. It begins by determining the largest index, i, such that $X[i] < x$. The procedure then adds the area of the trapezoid formed by the PDF between $X[i]$ and x to the cached CDF value $C[i]$.

```
CUMULATIVE-DISTRIBUTION(X, L, C, x)
 1 if x ≤ X[0]
 2     return 0
 3 elseif X[X.length − 1] ≤ x
 4     return 1
 5 else
 6     i = BINARY-SEARCH(x, X, 0, X.length − 1) − 1
 7     m = SLOPE(X, L, i)
 8     h = x − X[i]
 9     y = L[i] + mh // value of PDF at x
10     return C[i] + h(L[i] + y)/2
```

3.6. Quantile Procedure

The QUANTILE procedure evaluates the inverse CDF (i.e., quantile function) of the frequency polygon for real argument y. It begins by determining the largest

index, i, such that $C[i] < y$. If the slope of the PDF line segment between $X[i]$ and $X[i+1]$ is non-zero, then the quantile is the solution to a quadratic equation. Otherwise, linear interpolation is used to calculate the quantile.

QUANTILE(X, L, C, y)
1 **if** $y < 0$ **or** $1 < y$
2 **return** NaN
3 **elseif** $y = 0$
4 **return** $X[0]$
5 **elseif** $y = 1$
6 **return** $X[X.length - 1]$
7 **else**
8 $i =$ BINARY-SEARCH$(y, C, 0, C.length - 1) - 1$
9 $m =$ SLOPE(X, L, i)
10 $h = y - C[i]$
11 **if** $m \neq 0$ // solve quadratic
12 **return** $X[i] + (\sqrt{(L[i]^2 + 2mh)} - L[i])/m$
13 **else** // linear interpolation
14 $m = (X[i+1] - X[i])/(C[i+1] - C[i])$
15 **return** $X[i] + mh$

3.7. Sample Procedure

Using the inverse transform method, this procedure generates a random deviate from the frequency polygon.

SAMPLE(X, L, C)
1 generate random number $y \in [0,1]$
2 **return** QUANTILE(X, L, C, y)

4. PRACTICAL EXAMPLE

As discussed previously, the frequency polygon is best suited for density estimation of large data sets, such as those acquired from automated sensors, the outputs of Monte Carlo simulations (e.g., risk analyses), or outputs of Markov chain Monte Carlo (MCMC) algorithms (e.g., posterior distributions generated using Bayesian statistics).

To demonstrate the functionality of the proposed approach, the frequency polygon method was applied to generate the CDF of the outputs of a MCMC algorithm obtained by Ji and AbouRizk (2017). In their study, Ji and AbouRizk (2017) modeled the number of nonconforming (i.e., failed) welds in a pipe weld inspection process using a binomial distribution $B(n, p)$, where n was the sample size, and p was the probability of nonconformance (i.e., weld failure). The prior distribution $P(p)$ of parameter p was modeled as Beta(0.5, 0.5) and, after observing D nonconforming welds in n inspections, the posterior distribution, $P(p|X)$, was determined to be

$$Beta(D + 0.5, n - D + 0.5) \tag{10}$$

In particular, if $n = 100$ and $D = 10$, the posterior distribution is

$$Beta(10.5, 90.5) \tag{11}$$

It is important to note that, in this particular case, a closed-form (i.e., analytical) solution for the posterior distribution exists; however, a closed-form solution is often difficult to derive or does not exist for many posterior distributions. Therefore, for the purposes of demonstrating the functionality of the frequency polygon approach, a numerical solution was instead determined using the Metropolis-Hastings algorithm—a common MCMC method (Metropolis et al. 1953, Hastings 1970)—together with the frequency polygon. From Bayes' Theorem, the posterior distribution is

$$P(X) = \frac{L(X|p)P(p)}{P(X)} \tag{12}$$

where $L(X|p)$ is the likelihood function. As $P(X)$ is independent of p,

$$P(X) \propto L(X|p)P(p) \tag{13}$$

The Metropolis-Hasting algorithm is then used to generate random samples from the probability distribution when a function is proportional to its PDF. For the distribution denoted in Equation (12),

$$L(p) = p^D(1-p)^{n-D} = p^{10}(1-p)^{90} \tag{14}$$

and $P(p)$ is the PDF of $Beta(0.5, 0.5)$, therefore

$$P(X) \propto p^{10}(1-p)^{90}\frac{p^{0.5}(1-p)^{0.5}}{Beta(0.5, 0.5)} \tag{15}$$

Using Equation 17 and a starting p value of 0.1, the Metropolis-Hastings algorithm was used to generate 10,000 samples from the posterior distribution. A histogram consisting of 42 bins was constructed from the samples, and a frequency polygon was generated in *Simphony.NET* (AbouRizk et al. 2016) using the pseudo code approach detailed above. The resulting histogram and CDF are illustrated in Figures 3 and 4, respectively.

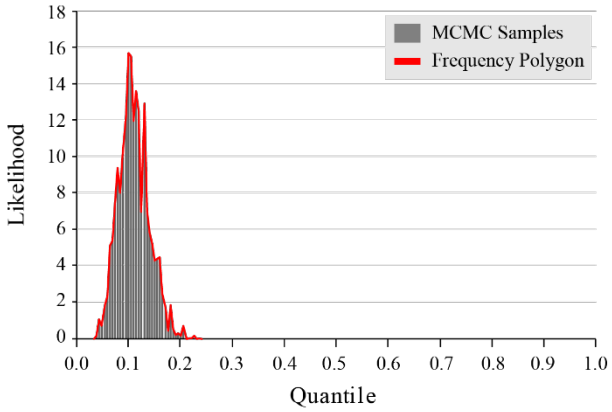


Figure 3: Histogram of Samples Generated using an MCMC-Based Method versus the Frequency Polygon

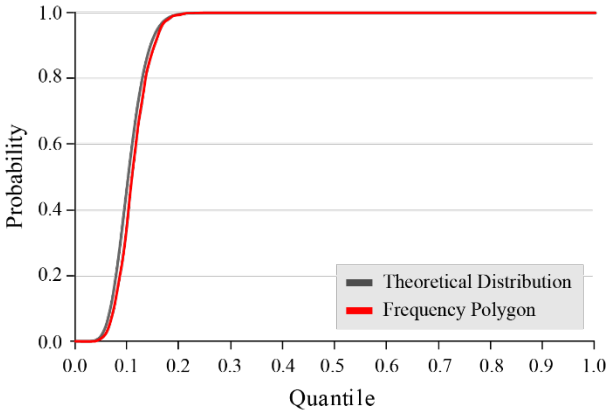


Figure 4: Cumulative Distribution Function of Theoretical Distribution versus the Frequency Polygon

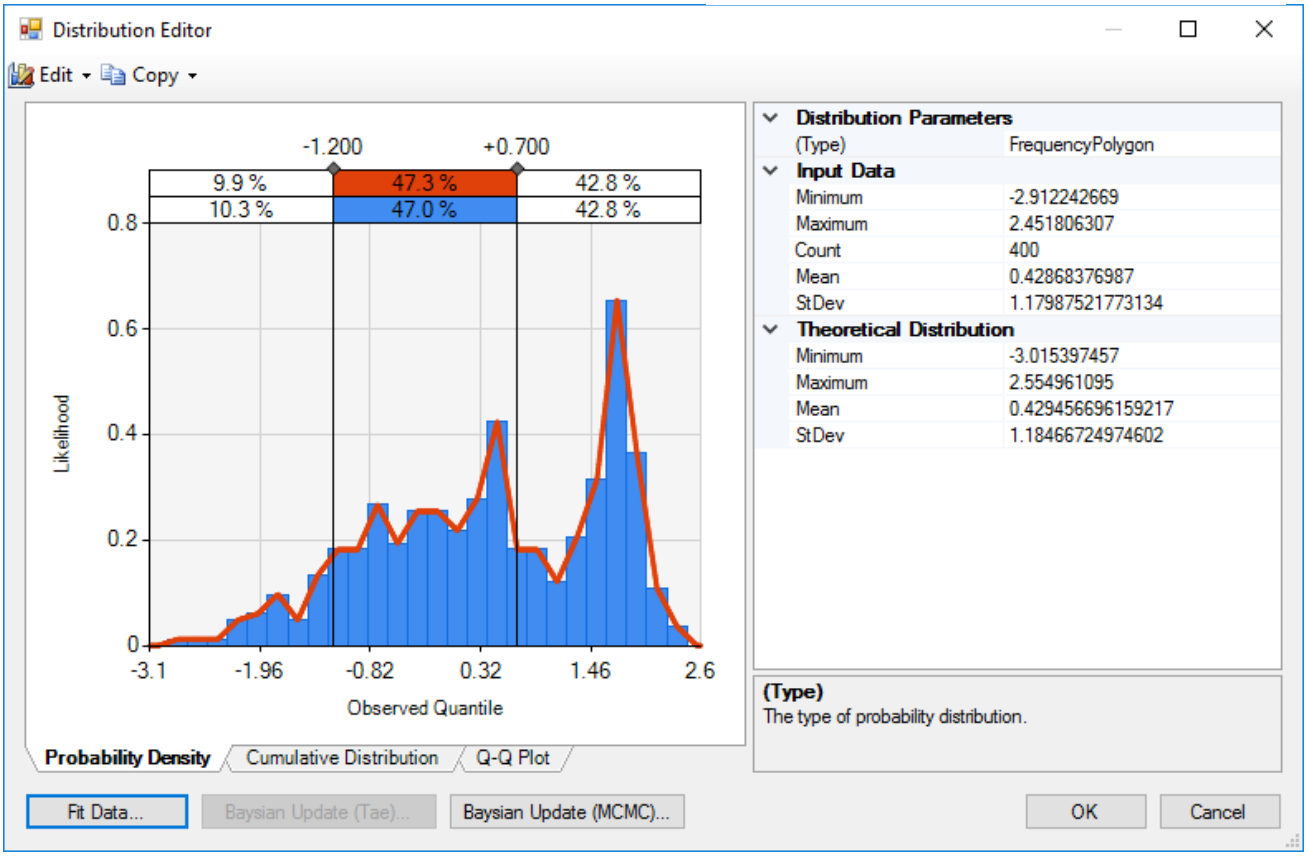


Figure 5: Output Results of the Implementation of the Frequency Polygon Pseudocode Approach in *Simphony.NET*.

5. IMPLEMENTATION IN *SIMPHONY.NET*

The frequency polygon has been implemented within the *Simphony.NET* simulation environment (AbouRizk et al. 2016). *Simphony.NET* is a modeling environment comprised of simulation services and a modeling user interface. Based on modular and hierarchical concepts, *Simphony.NET* provides a medium for developing and deploying simulation modeling templates that are general purpose by design, while featuring a number of special purposes templates.

The *Simphony.NET* environment now supports the fitting a frequency polygon to a set of data and the creation of a frequency polygon as a posterior distribution via the Metropolis-Hastings algorithm. A screenshot of the output results of the frequency polygon approach in *Simphony.NET* are illustrated in Figure 5.

6. CONCLUSIONS

The paper demonstrates how to implement nonparametric input modeling techniques to augment classic methods during simulation studies. The approach covers defining a probability density function

from random data and estimating both the cumulative density function and the inverse distribution. The latter facilitates sampling during simulation. The implementation was summarized in pseudo code to facilitate its use by others within their own systems.

ACKNOWLEDGMENTS

This project was supported by a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Sciences and Engineering Council of Canada.

REFERENCES

AbouRizk S., Hague S., Ekyalimpa R., Newstead S., 2016. Simphony: A next generation simulation modelling environment for the construction domain. Journal of Simulation, 10(3), 207–215.

Cormen T.H., Leiserson C.E., Rivest R.L., Stein C., 2009. Introduction to algorithms. 3rd ed. Cambridge, MA: The MIT Press.

Hastings W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1), 97–109.

Ji W., AbouRizk S., 2017. Credible interval estimation for fraction nonconforming: analytical and numerical solutions. Automation in Construction, 83(2017), 56–67.

MathWorks, Inc., 2019. MATLAB. Available from: https://www.mathworks.com/products/matlab.html?s_tid=hp_products_matlab [Accessed 03 April 2019].

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics, 21(6), 1087–1092.

Parzen E., 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33(3), 1065–1076.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: https://www.R-project.org [Accessed 03 April 2019].

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, 27(3), 832–837.

Scott D.W., 1979. On optimal and data-based histograms. Biometrika, 66(3), 605–610.

Scott D.W., 1985. Frequency polygons: Theory and application. Journal of the American Statistical Association, 80(390), 348–354.

Silverman B.W., 1986. Density estimation for statistics and data analysis. London, UK: Chapman & Hall.

AUTHOR BIOGRAPHIES

Simaan AbouRizk, PhD, PEng, FRSC, FCAE, NAC is a Distinguished University Professor, Tier 1 Canada Research Chair in Operations Simulation, and Chair of the Department of Civil and Environmental Engineering at the University of Alberta. As a renowned expert in the development and application of computer simulation for construction planning, productivity improvement, constructability review, and risk analysis, AbouRizk's research focuses on developing innovative information technologies for modeling, analyzing, and optimizing operations in the construction industry. His contributions to academia and industry have been recognized by numerous organizations including the American Society of Civil Engineers and the Canadian Society for Civil Engineering.

Stephen Hague is a System Analyst in the Department of Civil and Environmental Engineering at the University of Alberta. Steve has been instrumental in the development of numerous simulation and software tools that have been implemented throughout the industrial construction sector in Alberta, Canada. He has also played an integral role in the development of *Simphony.NET*, which continues to be extended at the University of Alberta through Dr. AbouRizk's research program.

A SIMPLIFIED PROBABILISTIC VALIDATION OF PRODUCTION FLOWS

Kadir Emir[(a)], David Kruml[(b)], Jan Paseka[(c)], Iveta Selingerová[(d)]

[(a),(b),(c),(d)]Department of Mathematics and Statistics, Masaryk University
Kotlářská 2, 611 37 Brno, Czech Republic

[(a)]emir@math.muni.cz,[(b)]kruml@math.muni.cz,[(c)]paseka@math.muni.cz,[(d)]selingerova@math.muni.cz

ABSTRACT
This paper extends the stack validation algorithm in a probabilistic way. In other words, we introduce a new model for stack validation when the production parameters are random variables and the result is compared with a confidence interval. The major outcome of this simplified probabilistic model is to determine random variables merely by mean, variance, and skewness. This straightforwardly enables some direct, fast and consistent calculations by using certain properties of these moments.

Keywords: categoric simulation of production flow, signal, stochastic processing times, probabilistic stack validation

1. INTRODUCTION

In the industrial factory planning practice; planning capacity, material in volatile or changing demand situations often require high capacities and also cause material plan instability for both suppliers and factories. The decisions of the planner can be based on many factors such as period machine capacity, profit margins, holding costs, etc. One of the biggest challenges of any planner is dealing with the level of capacity (input, output, or stack) during any step of the production flow. At this point, standard and traditional deterministic methods are of little use. However, probabilistic and statistical methods come into play for predicting the underlying risk of capacity and material in planning, in advance.

The category theoretical description and simulation of production flow is developed in Kruml and Paseka (2018) by introducing the *signal model*. This model is perfectly suitable for modeling the production in an algebraic point of view, which is mass and regular on processes. But these processes do not have to be synchronized. That is, such processes can operate under different speeds and batch sizes, and we need high capacity stacks to eliminate such differences. Every stack is filled by outputs of preceding processes and drawn by inputs of the next processes. We refer *signals* for these time functions. In regular mass production, signals can be effectively encoded by *formal words* where we write x for a produced unit, o for a time unit, and powers to mean lengths and repetitions. For instance, the word $(o^5x)^{10}o^{30}$ stands for the signal A (Fig. 2) where production cycle o^5x comprises cycle time 5 and 1 produced unit, is repeated 10 times, and followed by the inactivity of time 30. Expansion of the word draws a walk in the mass-time, where x is a move in mass and o is a move in time.

Such a decomposition and composition of signals is a part of general principles formulated for all aspects of the production flow and formalized by terms of category theory, see Coecke and Paquette (2011). Moreover, we know that a production flow can be modeled within three modes – time, space (network), and mass (material and products). The composition principle enables us to figure out the flow as a tree-organized collection of subordinated jobs and solve them independently, simultaneously, and with predefined accuracy.

However, the crucial problem in our methodology is the behavior of the stack. The flow is always considered to be valid regarding the stack, in the given simulation interval, and the stock needs to be kept balanced within the lower and upper bounds of the stack. The stock function is defined as a sum of all input and output signals. Its evolution could look chaotic and computationally hard for a complete determination. Instead of this, the *stack validation algorithm* (SVA) is developed based on tree organization of signals. The algorithm starts with rough approximations on large time intervals and continues with detailed approximations only on smaller parts where the rougher methods fail to decide the stack validity. In other words, it uses the recursive search of critical moments and avoids detailed inspection elsewhere that makes it *lazy* and *effective*.

2. STATEMENT OF THE PROBLEM

In this paper, we skip the process of recursion and mainly focus on the inner step of SVA. Here, we have the signals (Fig. 1) with lower and upper linear approximations, and we also use their sums as approximations of the sum signal. This follows a discussion of whether the approximations overlap the bounds. We extend this idea to the probabilistic level and describe an analogous step of *probabilistic stack validation algorithm* (PSVA). The recursive part of the algorithm remains similar to that of SVA.

Afterward, we propose to use our signal model to incorporate the probabilistic and statistical issues into the planning process. To our knowledge, no similar formulation exists in the literature, which is directly applicable to the planning situation discussed above.

Such modeling situation results in stochastic process-

ing times, and it is the typical one where queueing effects Manitz (2008) or stochastic-flow network models Fiondella, Lin and Chang (2015) occur. Nevertheless, we can still use the regularity of processes.

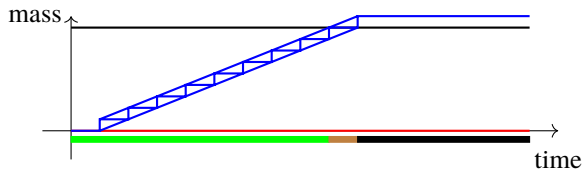


Figure 1: Stack validation

The lines enveloping the signal provide upper and lower approximations, which are much easier for calculations. Crossings with the upper bound (black) and the lower bound (red) of a stack indicate states of SVA (here we assume that there is no other signal affecting the stack). The green interval is surely correct, the black interval is surely incorrect (overfill), and the brown interval will be inspected with higher accuracy in a further recursion step.

We also expect that our probabilistic model could be useful also for any other composed probability systems including job production or project management – e.g., the network analysis technique PERT (see Pohl and Chapman (1987); Heagney (2016)).

This paper addresses the issues given above and makes the following contributions:

1. We extend the signal model from Kruml and Paseka (2018) to a probabilistic version without substantial changes.

2. Our approach can check the behavior of the whole system at each stack separately and hence to provide validity tests in a fast and reliable manner.

3. The method does not use the experimental approach (Monte Carlo simulation). Predictions are made by direct calculations based on fast transformations between probabilistic moments and quantiles.

3. STOCHASTIC SIGNALS

A geometric interpretation of the production flow is introduced in Kruml and Paseka (2018) through a certain surface in a space-time-mass. The model describes an ideal type of production with no errors or uncertainty. Here, we would like to describe a probabilistic extension of that model in the sense that any flow parameters could be considered as random variables. In this case, the resulting flow should not be imagined as a surface anymore but as a "fuzzy cloud" inside the space-time-mass – i.e., representing the density of possible runs of a given plan. We know that random effects are often dependent. For instance, an error of one process can affect the behavior of other processes or waste in series due to inappropriate machine setting. However, we suppose that many types of dependencies vanish when the random effects are resolved to elementary ones and positioned to appropriate segments in the model hierarchy.

Among other possibilities, the flow can be validated by the inspection of stacks. Namely, we test whether a stack keeps acceptable storage of produced units all the time, so it is never overfilled or lacked concerning defined bounds. We call this a *signal model* because the stack validation algorithm evaluates the input and output *signals*, i.e., time functions representing mass processed in time. The signal can be defined as a *formal word* and has a natural breakdown structure of subwords or subsignals. The main advantage of the model is that mass production is highly regular and repetitive. Moreover, the repetitions can be effectively encoded as *powers* in words. We claim that a large class of random effects can be modeled merely by adjusting these powers (Fig. 3, Fig. 4 and Fig. 5). The idea is quite general and also could be useful in planning or simulation of less regular applications, e.g., job production or project management.

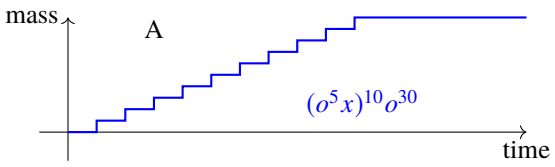


Figure 2: Signal A

The letters x, o mean mass (product) and time (delay) units, respectively. Signal A models a machine working in a regular regime that produces 1 mass unit per 5 time units. It totally produces 10 units and stops for 30 units.

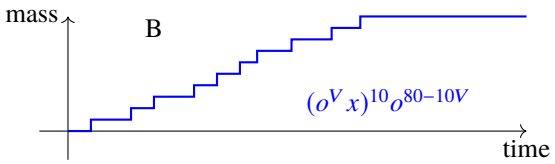


Figure 3: Signal B

Signal B is affected by irregularity in cycle time which is defined as random variable V.

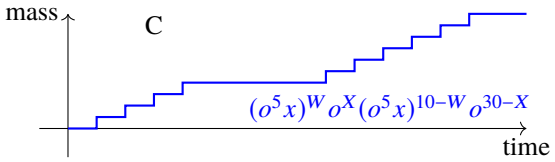


Figure 4: Signal C

In signal C, the machine works again in regular cycle time but the season is randomly divided to two parts by some stoppage. Here, we have two random variables: W represents the number of repetitions before the stoppage, and X is the time of the stoppage.

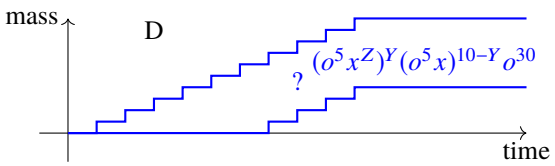


Figure 5: Signal D

Finally, signal D represents a mixture of two scenarios: in the first one the machine works well, in the second one the season starts with Y repetitions producing waste (and no products x). A scenario is selected by a random variable Z with Bernoulli distribution (see Ross (2014)) – the first scenario is performed if Z takes value 1, the second one if Z takes 0. All examples can be combined.

4. PROBABILISTIC STACK VALIDATION

The stack validation algorithm (SVA) respects the breakdown tree hierarchy of signals and iteratively search only on intervals where rough approximations do not provide decisive answers. This approach makes it more effective than straightforward simulation methods, e.g., discrete-time simulation. When we like to extend the algorithm to the probabilistic environment, we must reformulate the statement of the problem as well as to develop a fast calculus for manipulations with random variables.

First of all, *we give up any attempt to predict "global probability" that a given plan is capable*. This is a quite hard aggregation problem where all dependencies, hidden decisions, objective and subjective preferences should come into account. Instead of this, we still assume that the *plan evolves by small improvements made on particular processes, jobs, and orders.* The system communicates with the planner by predicting whether affected stacks perform well until the performance is considered satisfactory. But such a conclusion is completely under the responsibility of the planner. We must emphasize this fact because the probabilistic results are hardly 100% decisive and the planner should anticipate results like "the stack will be overfilled with a probability smaller than 5%".

According to the philosophy that "makes the planner responsible for everything", we let him/her set these *confidence intervals* for each stack separately and provide the validation concerning them. This can reflect thinking in a sense "*Overfill of that stack would mean a small temporal disorder next to the cutter but this happens often, and nobody bothers*" or "*Shortage of that stack is unpleasant, but I have some extra reserves of material to supply*".

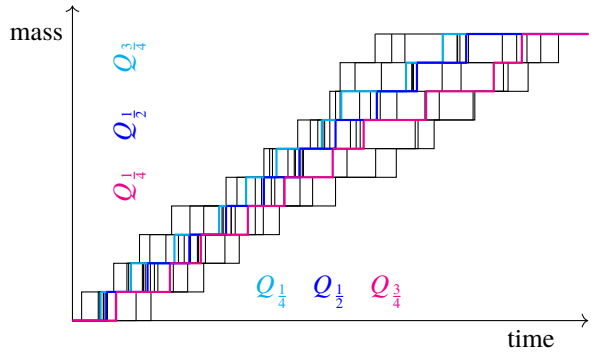


Figure 6: "Fuzzy cloud" representing a density of possible runs of a given plan

The "cloud" in Fig. 6 is drawn by multiple runs of a probabilistic process. The first, second, and third quartile is depicted by a cyan, blue, and magenta lines, respectively (From the mass point of view, the order is reversed).

Our main goal is to find *reliable approximations of the flow* such that we would certainly know that several possible runs out of such range are less than a given percentage. In other words, *we compare bounds of the stack with given quantiles of (random and time-varying) stock.* Since we consider only one-way transitions in the production network (two-ways transitions can be separated into two channels), the mass development of any signal has an obvious orientation (positive on inputs, negative on outputs). Of course, time only moves in a positive direction. Thanks to this, we find obvious but important fact about the flow: For a given *state* (t, N) (i.e., the event that Nth unit passed in time t) there is the same number of runs "on the right" (those which reach N after time t) and of runs "below" (those which reach N before time t). One can consider the flow as a collection of random walks or an analog of two-dimensional distribution. In this formalism, the above fact states that the marginal distributions along time and the marginal distributions along mass *share quantiles*, i.e., distribution of mass in time t reach in value N the same quantile like distribution of time in mass N in value t (Fig. 6 and Fig. 8). (To be more precise, for a quantile Q_p we should rather speak about the complementary quantile Q_{1-p} because of the orientation of the time and mass axes.)

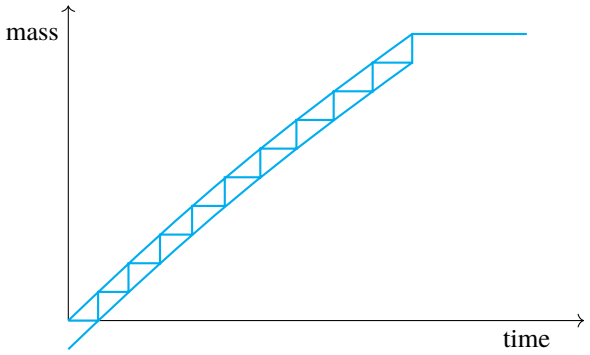


Figure 7: Upper and lower approximations of a quartile

By theory, the quartiles evolve as sequences linearly depending on n and $\sqrt{n}$ where n is the number of repetitions. Thus, they have upper and lower approximations by certain quadratic curves. These can be interpreted as quartiles of a *continuous random process*.

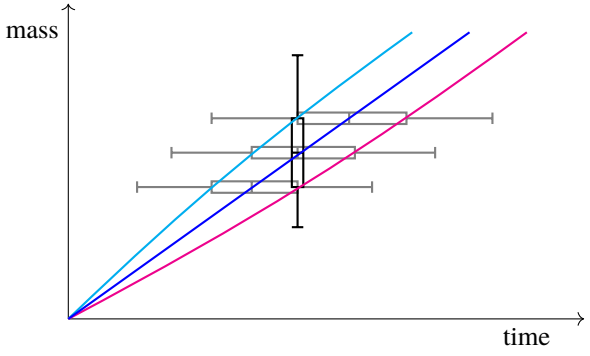


Figure 8: Swap of axes (from time to mass)

In Fig. 8 quartiles for fixed mass (gray) yield quartiles for fixed time (black). Consequently, we use them to estimate the vertical distribution.

Of course, this is not the case of densities – they are partial derivations of the cumulative density functions,

and there is no reason to be equal. Moreover, the two orthogonal distributions may have quite different shape properties. For example, if the signal is random in cycle time and we assume that all repetitions have the same distribution and are independent, then (after quite a small number of repetitions) the sum of times tends to a normal distribution, regardless of the type of distribution on one cycle (central limit theorem). On the other hand, the corresponding orthogonal distribution need not be symmetrical even in a large number of repetitions – we can see that there is more pointed density for small N and less pointed for large N because later units are produced with higher dispersion in time. Thus, the *skewness* of the distribution will be positive.

The described swap of axes enables us to study the mass (vertical) distribution using quantiles of time (horizontal) distributions. No matter which distributions enter the validation and how they are combined, we use the idea at least for the end of each iteration step in the SVA because we compare the mass distribution with the bounds of the stack. But it is also important for parallel summation of more than one random signals, which is a typical situation occurring at stacks.

The swap is known for certain classes of distributions. Namely, Poisson distribution is, in this sense, inverse to exponential distribution and normal distribution to a certain type of generalized inverse Gauss distribution. Our method is not precise, but it is simple and general.

5. L-ESTIMATORS AND MOMENTS

In practice, the random variables presented in the flow model are obtained from statistical analysis of real production datasets or guesses resulting from long-term experience. In many cases, the planner is not an expert neither on probability theory nor on mathematical statistics. Therefore he/she will be only able to use simple methods for the probability distribution.

A popular method is the 3-point estimation where just *minimum*, *middle*, and *maximum* values of the distribution are considered. The middle value is interpreted as a location parameter depending on context or planner's preferences; it can be mean, median or mode of the distribution. The minimum and maximum points determine variability. Since the minimum and maximum are very inclinable to chance, they can be replaced by much more robust quartiles Q_1 and Q_3, namely the quantiles of 25% and 75%. When we adopt the median Q_2 for the role of middle value, we model the distribution by triple quartiles (Q_1, Q_2, Q_3), that is, the values obtained by dividing the ordered dataset to four quarters. Similar approaches are examined in various works such as Bland (2015); Wan, Wang, Liu and Tong (2014) and also Hozo, Djulbegovic and Hozo (2005).

Important characteristics of the distribution can be guessed by the appropriate *L-estimators*. Namely, we will use the following three ones:

- $\frac{Q_1 + Q_3}{2}$ ... *midhinge*, location measure, replacement for mean,
- $Q_3 - Q_1$ ... *interquartile range*, scale or dispersion measure, replacement for standard deviation,
- $\frac{Q_1 + Q_3}{2} - Q_2$ ... difference of midhinge and median, *skewness measure*.

Such simplifications could be very inaccurate in many situations, e.g., when the modeled distribution is discrete or of exceptional shape. (A deterrent example is the Bernoulli distribution Z from signal D in Fig. 5.) Nevertheless, our experiments with many natural types of continuous unimodal distribution demonstrated that the L-estimators are in a surprisingly good coincidence with moments, and they provide an elementary calculus for summation of random variables. Notice that the transformation from quartiles to L-estimators is reversible. Hence one can also recover the quartiles from the L-estimators. (This is also the reason why we work only with a triple of quartiles instead of popular *five-point* or *seven-point* cases. Here the transformation would be more accurate but not reversible.) Thanks to this, we can also invert the procedure for special distributions including the discrete ones – we first find the moments by precise mathematics *and then* get *virtual quartiles* by the inverse transformation. The remaining step converting moments to the L-estimators is a linear adjustment of each of the moments and can be done concerning an assumed kind of distribution. That is, one can create a "dictionary" of frequently used distributions with specified values of the coefficients. In this manner, we can handle any distribution in two equivalent ways – either by the three quartiles (regardless they have the correct meaning, or they are just "ghosts") or by three moments, namely mean, variance, and skewness. The former triple is needed for the swap of axes, the later for a quick and stable summation of random variables which is based on the well-known *additivity* of the three moments (see Renyi (2007)):

Theorem 1 *Let X_1, X_2 be two independent random variables with the triples $(\mu_1, \sigma_1^2, \tau_1), (\mu_2, \sigma_2^2, \tau_2)$ of moments, i.e., mean, variance, and third central moment. Let $X = X_1 + X_2$. Then*

$$\mu = \mu_1 + \mu_2, \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2, \qquad \tau = \tau_1 + \tau_2$$

are moments of the random variable X.

In particular, let $X_i, i = 1, \dots n$ be independent identically distributed random variables, and define:

$$X = \sum_{i=1}^{n} X_i.$$

If the fixed triple $(\mu_1, \sigma_1^2, \tau_1)$ will be the moment triple of any X_i, then

$$\mu = n\mu_1, \qquad \sigma^2 = n\sigma_1^2, \qquad \tau = n\tau_1$$

are moments of the random variable X.

For a random variable X with a moment triple (μ, σ^2, τ) we define $\gamma = \tau/\sigma^2$. We can think of γ as a parameter for *skewness*. If X_1, X_2 are independent random variables with moments triples $(\mu_1, \sigma_1^2, \tau_1), (\mu_2, \sigma_2^2, \tau_2)$,

$\gamma_1 = \tau_1/\sigma_1^2$, $\gamma_2 = \tau_2/\sigma_2^2$ and $X = X_1 + X_2$ then we obtain the parameter γ for X as

$$\gamma = \frac{\gamma_1\sigma_1^2 + \gamma_2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

The midhinge is a direct guess of the mean and thus we will also denote it by $\widehat{\mu}$. The interquartile range is assumed to be a multiple of the standard deviation, i.e., the square root of the variance. To make the use more intuitive, we will consider half of the interquartile range and denote it by $\widehat{\sigma}$. Finally, the difference between midhinge and median will be denoted by $\widehat{\gamma}$. The parameter γ is assumed to be proportional to the scale in the sense that third central moment τ linearly depends on $\gamma\sigma^2$. This property yields direct formulas for the moments. Hence for a prediction of a sum of two independent random variables we will use the following rules:

$$\widehat{\mu} = \widehat{\mu}_1 + \widehat{\mu}_2, \quad \widehat{\sigma} = \sqrt{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}, \quad \widehat{\gamma} = \frac{\widehat{\gamma}_1\widehat{\sigma}_1^2 + \widehat{\gamma}_2\widehat{\sigma}_2^2}{\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2}. \ (*)$$

Consequently, the sum of n independent identically distributed random variables will be predicted as follows:

$$\widehat{\mu} = n\widehat{\mu}_1, \qquad \widehat{\sigma} = \sqrt{n}\widehat{\sigma}_1, \qquad \widehat{\gamma} = \widehat{\gamma}_1.$$

Let us recall that

$$\widehat{\mu} = \frac{Q_1 + Q_3}{2}, \quad \widehat{\sigma} = \frac{Q_3 - Q_1}{2}, \quad \widehat{\gamma} = \frac{Q_1 + Q_3}{2} - Q_2,$$

and derive the inverse transform

$$Q_1 = \widehat{\mu} - \widehat{\sigma}, \qquad Q_2 = \widehat{\mu} - \widehat{\gamma}, \qquad Q_3 = \widehat{\mu} + \widehat{\sigma}.$$

6. EXAMPLE

To demonstrate the model we consider a stack with the input signal $s_1 = o^{20}(xo^Y)^{120}$ and the output signal $s_2 = o^{25}(x^6o^Z)^{20}$ where Y, Z both have triangular distributions with min/max/mode values $0/1/0.5$ or $0/6/3$, respectively. This yields quartiles

$$Q_{11} = \frac{1}{2\sqrt{2}} \approx 0.354, \qquad Q_{12} = \frac{1}{2},$$
$$Q_{13} = 1 - \frac{1}{2\sqrt{2}} \approx 0.656, \quad Q_{21} = \frac{3}{\sqrt{2}} \approx 2.121,$$
$$Q_{22} = 3, \qquad Q_{23} = 6 - \frac{3}{\sqrt{2}} \approx 3.879,$$

and L-estimators

$$\hat{\mu}_1 = 0.5, \qquad \hat{\sigma}_1 = 0.156, \qquad \hat{\gamma}_1 = 0,$$
$$\hat{\mu}_2 = 3, \qquad \hat{\sigma}_2 = 0.879, \qquad \hat{\gamma}_2 = 0.$$

Let us look at the finishing time of s_1. The model predicts that the quartiles evolve to

$$Q_{11,13} = 20 + 120\hat{\mu}_1 \mp \sqrt{120}\hat{\sigma}_1 \approx 20 + 60 \mp 2.236$$
$$= 77.764, 82.236,$$
$$Q_{12} = 20 + 120\hat{\mu}_1 - \hat{\gamma}_1 = 20 + 60 + 0 = 80.$$

Five simulations, each with 1000 runs, provided datasets with:

experiment	Q_{11}	Q_{12}	Q_{13}
1	78.445	79.916	81.545
2	78.464	79.911	81.222
3	78.436	80.086	81.938
4	78.468	79.993	81.689
5	78.537	79.944	81.511

The result has a significantly different interquartile range (or standard deviation). This difference is caused by the fact that one cycle distribution is assumed to be triangular while its 120th power is almost normal. But ratio $\hat{\sigma}/\sigma$ is about 0.717 for triangular distribution while 0.675 for normal distribution. When the difference is multiplied by $\sqrt{120}$, we get an error 0.470 which agrees with the difference between modeled values and experimental results. (We can conclude from this that the input quartiles of cycle times should be rather obtained from repeated cycles or large datasets than from a single cycle and triangular expectation.)

Now let us look on stock at time 40 for both signals and calculate quartiles in the mass axis. The quartile guesses appear as solution of equations, cf. Fig. 8:

$$\hat{\mu}_i\frac{Q_{i1}}{n_i} + \hat{\sigma}_i\sqrt{\frac{Q_{i1}}{n_i}} = 40 - t_i,$$
$$\hat{\mu}_i\frac{Q_{i2}}{n_i} - \hat{\gamma}_i = 40 - t_i,$$
$$\hat{\mu}_i\frac{Q_{i1}}{n_i} - \hat{\sigma}_i\sqrt{\frac{Q_{i1}}{n_i}} = 40 - t_i.$$

where $i = 1, 2$ is an index of signal, $n_1 = 1, n_2 = 6$ the batch sizes, and $t_1 = 20, t_2 = 25$ the start times.

This yields

$$Q_{11} \approx 38.190, \qquad Q_{12} = 40, \qquad Q_{13} \approx 41.896,$$
$$Q_{21} \approx 26.319, \qquad Q_{22} = 30, \qquad Q_{23} \approx 34.195,$$

hence

$$\hat{\mu}_1 \approx 40.043, \qquad \hat{\sigma}_1 \approx 1.853, \qquad \hat{\gamma}_1 \approx 0.043,$$
$$\hat{\mu}_2 \approx 30.257, \qquad \hat{\sigma}_2 \approx 3.938, \qquad \hat{\gamma}_2 \approx 0.257.$$

For the subtraction signal $s = s_1 - s_2$ we get from (*)

$$\hat{\mu} = 9.786, \qquad \hat{\sigma} \approx 4.352, \qquad \hat{\gamma} \approx -0.203,$$

and by transforming it back

$$Q_1 \approx 5.434, \qquad Q_2 \approx 9.989, \qquad Q_3 \approx 14.138.$$

Notice now that PSVA still works with envelopes of

signals where the quartile guesses are shifted up and down by constants depending on the stair step (batch size) of each of signals. Here the upper guess should be shifted by 1 (batch of positive s_1) and the lower guess by 6 (batch of negative s_2).

Again, from experiments of 1000 runs we get datasets

experiment	Q_{11}	Q_{12}	Q_{13}
1	3	8	12
2	3	8	11
3	3	8	11
4	3	7	11
5	3	8	12

which are close to the centre of expected signal range, and display expected variance and skewness. (The mass distribution represents stored products, thus it is discrete with support on integers.)

The model provided an excellent performance also for other settings.

7. APPROXIMATING CURVES

Like in the ideal production SVA, we need to determine intervals where stock is in given bounds and where it is not. The PSVA assumes that the enveloping approximations of a certain quantile are always quadratic curves of form $at + b\sqrt{t + c} + d$. In the simplest case of one signal of a regularly working process, this has an obvious interpretation: a is mean production speed, b, c are related to standard deviation, b is positive for Q_3 and negative for Q_1, and d comprises actual stock at the beginning of the tested time interval and eventually the skew component γ. Since the Q_1 and Q_3 differ only on the sign of b, the function $(Q_3 - Q_1)^2/4$ (an L-estimator corresponding to variance) is only a linear function of time. Summing of such signals would be very simple because we could use the formulas (*) to calculate the coefficients a, b, c, d.

But in general situations, there is no relation between the values b for Q_1 and Q_3. For example, when the process start is random, part of runs already draws the parabolic shape of Q_3, i.e., the parameter b is non-zero, while the other runs still "do nothing" and the function Q_1 is constant with $b = 0$ on the initial interval of production. Evolution of L-estimators is non-linear, and summed signals could have complicated formulas for their exact values.

However, we still consider that the L-estimators are replacing mean, variance, and the third central moment and that they evolve linearly. This yields formulas

$$\begin{aligned}\hat{\mu} &= (1-t)\hat{\mu}_0 + t\hat{\mu}_1,\\ \hat{\sigma}^2 &= (1-t)\hat{\sigma}_0^2 + t\hat{\sigma}_1^2,\\ \hat{\gamma} &= \frac{(1-t)\hat{\gamma}_0\hat{\sigma}_0^2 + t\hat{\gamma}_1\hat{\sigma}_1^2}{(1-t)\hat{\sigma}_0^2 + t\hat{\sigma}_1^2}\end{aligned}$$

where the index 0 stands for begining of a time interval, index 1 for its end, and $t \in [0, 1]$. Thus we extrapolate the values of L-estimators to any time just from the values at some critical points. It is reasonable to consider all the points where some of the quartile curves are broken.

The assumption is correct for the central part of the process activity but not exact outside it, as the Fig. 9 shows.

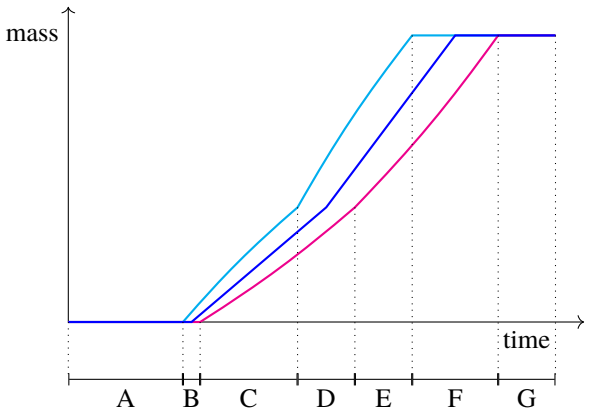


Figure 9: Evolution of L-estimators

In A, there is no activity. In B, the process may have started or not yet. In C, the process surely works in one regime. In D, the process still works in the first regime or perhaps it has switched to the second (faster) regime. In E, the process surely works in the second regime. In F, the process still works in the second regime, or it has already stopped. In G, the process has undoubtedly finished. The periods A, C, E, and G are parametrized exactly, the periods B, D, F are simplified. However, the error caused by this assumption is relatively small concerning the error caused by the simplified calculus of L-estimators.

Another objection to the model is that it does not assume that the process could start sooner than Q_1 or finish later than Q_3, but both cases may quickly happen. Such a problem can be fixed by adding two more parameters Q_0 and Q_4 according to Tukey (see Tukey (1977)), representing reasonable bounds of the distribution. They could be used in situations when $Q_1 = Q_2 = Q_3$, that is, when the distribution seems to be trivial (and the process to be deterministic) but we know that it is not. These Q_0 and Q_4 could be interpreted as "0th" and "4th" quartile and identified with $\hat{\mu} \pm 3\hat{\sigma}$. Then there are analogous transformation rules between quartiles and L-estimators and Q_0, and Q_4 hold the quadratic nature like Q_1 and Q_3.

Anyway, the proposed model approximates the probability density of a signal at all points by simply parametrized curves. Both quartiles and L-estimators have quadratic parametrization, and hence the same is true for sums of signals. The validity of a stack concerning a given confidence interval is then answered by solving an appropriate quadratic equation.

8. CONCLUSION

The described method may resemble a Fourier transform of functions. Summation of general random variables can be provided by discrete or continuous convolution. However, this can be very difficult or at least computationally hard. Instead of this, we parametrize the variables just by the three quartiles, convert them to certain L-estimators and work with those as with moments. Assuming the additivity, we quickly get L-estimators of the sum and convert them back to quartiles (see Fig. 10).

The quartile parametrization enables to swap between the time and mass axes easily. The only thing we need to do is to find the intersections of quartile curves with a line. Since mean evolves linearly and the skew parameter is constant for a repetitive process, we consider such parametrization for a general description of all approximations. Consequently, the crossing points can be found as solutions of certain linear or quadratic equations. The result is converted back to any of the forms and used for the next operation or to recover the resulting distribution.

Using the method, we can extend SVA to the probabilistic environment without substantial changes in its general principles and reasoning. The derived calculus is rough and simplified but still robust and highly effective for repetitive processes. It completely avoids the traditional Monte Carlo approach in simulation. Our experiments show that the proposed model is not only quick but also very realistic and consistent.

From another point of view, we know from Kruml and Paseka (2018) that the production flow without error or uncertainty forms a monoidal category structure with tensor product (Coecke and Paquette (2011)). In this paper, we now replace the certainty with uncertainty and some probabilistic errors. On the other hand, it is well known that there is a close relationship between fuzzy sets and probability theory; see Dubois, Nguyen and Prade (2000) for more details. Consequently, there might be again a monoidal category structure somehow related to fuzzy categories (Walker (2004)). In such a case, it is an interesting question how fuzzy categories and monoidal categories play a role together to model this structure in the light of category theory.

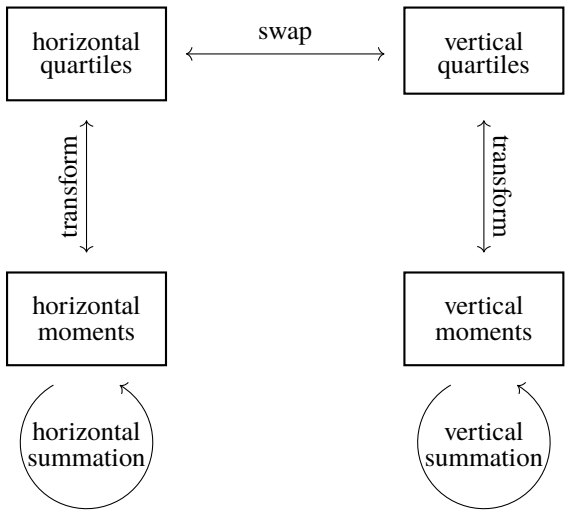


Figure 10: Computational strategy

ACKNOWLEDGEMENTS

We thank the anonymous referees for the extensive reading and contributions to improve our presentation of the paper.

Research of the first author was supported by the project Mathematical structures 8 (MUNI/A/1186/2018). Research of the second (corresponding) author was supported by the project New approaches to aggregation operators in analysis and processing of data, Nr. 18-06915S by Czech Grant Agency (GAČR). Research of the last two authors was supported by the project Research and Development of Information System for Value Flow Optimization and Production Planning financed by the Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest concerning the research, authorship, and publication of this article.

REFERENCES

Bland M., 2015. Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. International Journal of Statistics in Medical Research 4 (1), 57–64.

Coecke B., Paquette É. O., 2011. Categories for the practicing physicist. In: Coecke B., ed. New Structures for Physics. Berlin: Springer, Lecture Notes in Physics, vol. 813, 173–286.

Dubois D., Nguyen H. T., Prade H., 2000. Possibility theory, probability and fuzzy sets. Misunderstandings, bridges and gaps. In: Fundamentals of fuzzy sets. Dordrecht: Kluwer Academic Publishers, 343–438.

Fiondella K., Lin Y., Chang P., 2015. System performance and reliability modeling of a stochastic-flow production network: A confidence-based approach. EEE Transactions on Systems, Man, and Cybernetics: Systems 45 (11), 1437–1447.

Heagney J., 2016. Fundamentals of project management. McGraw-Hill.

Hozo S. P., Djulbegovic B., Hozo I., 2005. Estimating the mean and variance from the median, range, and the size of a sample. BMC Medical Research Methodology, 5:13.

Kruml D., Paseka J., 2018. Categoric simulation of production flows. Proceedings of the 17th International Conference on Modeling and Applied Simulation, pp. 68–75. 17-19 September 2018, Budapest (Hungary).

Manitz M., 2008. Queueing-model based analysis of assembly lines with finite buffers and general service times. Computers & OR 35 (8), 2520–2536.

Pohl J., Chapman A., 1987. Probabilistic project management. Building and Environment 22 (3), 209–214.

Renyi A., 2007. Foundations of probability. Dover: Dover Publications, Dover Books on Mathematics.

Ross S.M., 2014. A first course in probability. 9th ed., international ed. Boston, MA: Pearson.

Tukey J.W., 1977. Exploratory Data Analysis. Boston, MA: Pearson.

Walker C.L., 2004. Categories of fuzzy sets. Soft Computing 8 (4), 299–304.

Wan X., Wang W., Liu J., Tong T., 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology 14 (1), DOI: 10.1186/1471-2288-14-135.

AUTHORS BIOGRAPHY

KADIR EMIR

is a Ph.D. student at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic; under the main supervision of Professor Jan Paseka. He received his first Ph.D. in 2016 in the field of algebraic topology. His current research area spans the category theoretic approach to quantum theory and production networks.

DAVID KRUML

is an assistant professor at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic. He received his Ph.D. in 2003. He is primarily interested in algebraic quantum structures, category theory, and game theory. He currently works as a key person on the project "Research and Development of Information System for Value Flow Optimization and Production Planning" financed by Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370 (2017–2019).

JAN PASEKA

is an associate professor at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic. He received his Ph.D. in 1990. His professional field of interests includes a non-commutative generalization of topology, quantum, and many-valued structures. He was a principal investigator of several research projects. He currently supervises at Masaryk university the project "Research and Development of Information System for Value Flow Optimization and Production Planning" financed by Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370 (2017–2019).

IVETA SELINGEROVÁ

is a researcher at Masaryk Memorial Cancer Institute and lecturer at the Department of Mathematics and Statistics of Masaryk University in Brno, Czech Republic. She received her Ph.D. in 2015. Her research activities include biostatistics, survival analysis, clinical trials, kernel estimation, or numerical methods. She worked within the project "Research and Development of Information System for Value Flow Optimization and Production Planning" financed by Ministry of Industry and Trade: project CZ.01.1.02/0.0/0.0/15_019/0001370 (2017–2019) on modeling and testing for simulation of production flows with a stochastic character.

INTEGRATED ERGONOMIC AND PRODUCTIVITY ANALYSIS FOR PROCESS IMPROVEMENT OF PANELISED FLOOR MANUFACTURING

Chelsea Ritter[(a)], Regina Dias Barkokebas[(b)], Xinming Li[(c)], Mohamed Al-Hussein[(d)]

[(a),(b),(d)] Hole School of Construction Engineering, University of Alberta, Edmonton, Canada
[(c)] Mechanical Engineering Department, University of Alberta, Edmonton, Canada

[(a)]critter1@ualberta.ca, [(b)]rdbarkokebas@ualberta.ca, [(c)]xinming.li@ualberta.ca, [(d)]malhussein@ualberta.ca

ABSTRACT
Workers in the construction manufacturing industry are often exposed to labour-intensive tasks with ergonomic risks such as awkward body posture, forceful exertion, and repetition motion. Due to the increased productivity and increased repetitive motions resulting from improvement initiatives implemented in offsite construction, the investigation of ergonomic risks associated with these changes is needed. In this context, this paper explores an existing panelised floor production line aiming to minimize its ergonomic risks while improving its current productivity rate. Information on human body motion and productivity are extracted from video recordings. The ergonomic risks associated with specific tasks are identified using an existing ergonomic risk assessment tool (i.e., Rapid Entire Body Assessment (REBA)). The information extracted from the simulation model pertaining to ergonomic risks and productivity supports the decision-making process and aids in the prioritization of changes to improve the working environment.

Keywords: workstation design, work measurement, ergonomics, decision support system, productivity improvement

1. INTRODUCTION

Workers' unsafe behaviour is responsible for 80% of construction accidents (Li et al. 2015). In 2017, the manufacturing and construction industries accounted for the second and fourth highest number of diseases and lost time injuries, and the second and first highest number of fatalities among all industries in Canada, respectively (Association of Workers' Compensation Boards of Canada (AWCBC) 2019). Construction and manufacturing workers are often exposed to the three primary causes of work-related musculoskeletal disorders (WMSDs): awkward body posture, forceful exertion, and repetitive motion (Canadian Centre for Occupational Health and Safety 2017; Public Services Health & Safety Association (PSHSA) 2010; Xu et al. 2012). WMSDs, which are often caused by bad workplace ergonomics, are responsible for higher absenteeism and injury rates thus resulting in significant loss of productivity and increased production cost (Botti et al. 2017a; Rajabalipour Cheshmehgaz et al. 2012). Hence, the investigation of physical demands of body motion is needed for workstation design and manufacturing processes in offsite construction facilities to minimize WMSDs and reduce any negative impacts on company productivity.

Several approaches have been developed to identify and assess ergonomic risks by analysing two main factors: (a) body posture (e.g., body angles, force load, and interaction between the human body and other working elements such as tools and machines) (Golabchi et al. 2016), and (b) biomechanical analysis, which focus on internal and musculoskeletal loads, and stresses on joists (Armstrong et al. 1996). For instance, Rapid Entire Body Assessment (REBA) (Hignett & McAtamney 2000), Rapid Upper Limb Assessment (RULA) (McAtamney & Nigel Corlett 1993), and Ovako Working Posture Analysing System (OWAS) (Karhu et al. 1977) focus on body posture analysis, while 3D Static Strength Prediction Program (The University of Michigan Center for Ergonomics 2017) and OpenSim (OpenSim 2019) use biomechanical analysis as their primary assessment approach. In research studies, REBA and RULA are often applied to conduct ergonomic risk assessment (Li 2017).

Meanwhile, the construction industry is known for low productivity rates (Ikuma et al. 2011). One approach to improve lead times and productivity for construction projects is to shift to offsite construction and utilise lean construction principles (Abbasian-Hosseini et al. 2014; Dotoli et al. 2015; Jia et al. 2013; Yu et al 2013). To ensure continuous improvement of productivity, many facilities focus on continuous improvement initiatives (Aqlan and Al-Fandi 2018). These initiatives need to consider the impact of any changes made on the business, including the cost, productivity, ergonomic, and public perception implications of making changes to the production process. The considerations when determining if a process change will have desirable effects can be conflicting, making it difficult for management to determine which changes should be implemented. Although implementation of lean construction principles results in improved productivity, studies indicate that it also leads to an increase in

physical workload and motion repetition, which are factors associated with WMSDs (Botti et al. 2017a; Colombini et al. 2002; Hochdörffer et al. 2018; Mossa et al. 2016). Investigating the implementation of lean principles in industrialized construction is thus needed to explore its impacts not only on productivity, but also on ergonomic risks.

Few studies investigate ergonomic risks in industrialized construction. Inyang et al. (2012) and Abaeian et al. (2016a) propose a framework to perform ergonomic assessment on a residential construction production line. The application of three-dimensional (3D) models to automatically identify and evaluate awkward body posture to reduce WMSDs in manufacturing plants is explored by Golabchi et al. (2015) and Li et al (2017a). Li et al. (2017b) investigate muscle activity during repetitive material handling, which is also explored by Abaeian et al. (2016b). An improved physical demand analysis based on ergonomic risk in manufacturing construction is developed by Li et al. (2019). The application of ergonomic principles in the design of work places to reduce injury rates and exposure to ergonomic risks while increasing productivity is explored in several studies (Battini et al. 2011, 2015; Bortolini et al. 2018; Botti et al. 2017a,b; Golabchi et al. 2018; Ikuma et al. 2011; Mossa et al. 2016).

In this paper, an existing panelised floor manufacturing line is investigated with the objective of minimizing ergonomic risks while maintaining/increasing the production rate.

2. METHODOLOGY

This study combines two common research areas in offsite construction and simulation to recommend decision alternatives that consider not only the productivity and performance aspects of the proposed change, but also the effect on the ergonomic risk for the workers who are expected to complete these tasks. This analysis is done by integrating the REBA score for the various postures required to complete a floor construction task into a simulation model that represents the production times and possible task alternatives for the process. The methodology can be broken down into three areas, which are discussed in more detail in the following sections.

2.1. Production Observations and Simulation

To carry out a full analysis of the production activities for the floor panel construction process, it was first necessary to observe and understand the current process. This was done by in-person observations, as well as through video cameras installed in the facility. Workers were notified of the study and were aware of the cameras. It was communicated to the workers in the area that the analysis was for the process and not for critique of their work habits.

Throughout the observations, timings for individual tasks, differences in the way tasks were being completed, and the number of people working on a task at a time were recorded. This information was used to create a simulation model of the process and to test proposed improvements for the purposes of productivity increase, as detailed in a previous study (Ritter et al. 2016). The production alternatives shown in Table 1 were identified for investigation in this study, as they have been selected as possible productivity improvement measures that involve a change in the way people complete the production tasks.

Table 1: Identified Production Alternatives

Task	Current State	Possible Alternative
Place joists on jig	One person placing large joists (half of the time)	Require two people to place large joists
Apply glue on joists	Apply glue with automated multi-function bridge	Apply glue manually when multi-function bridge is being utilized on the other jig
Place sheathing on joists	Sheathing placed with vacuum lift (1 person)	Sheathing delivered with material delivery bridge
	Sheathing placed with vacuum lift (2 people)	

2.2. Ergonomic Risk Assessment

As identified through literature review and discussions with industry partners, the ergonomic risks associated with a production change are a priority factor, along with the cost and production interruption, when deciding whether to implement the change. An ergonomic assessment is performed using REBA to identify the ergonomic risks associated with the current process to which the proposed changes will be compared in terms of ergonomic risks and productivity. REBA is selected in this study as its total score encompasses information on upper and lower limbs as well as force load and coupling, thus covering the majority of human body movements encountered in the construction task explored in this paper (Hignett & McAtamney 2000).

The inputs for the REBA assessment are collected based on observation of workers' postures and motions extracted from video recordings of the panelised floor manufacturing process. Video recording is a cost-effective approach to acquire posture-based information and it also has the advantage of not disrupting the workforce during data collection (Li & Buckle 1999). An example of the REBA assessment conducted in this study is presented Figure **1**. The overall REBA score will be calculated in the simulation model using the formula shown in Equation 1, where d_i is the duration of each task. According to the REBA score, the risk level of a task and necessity of action is obtained as summarized in Table 2.

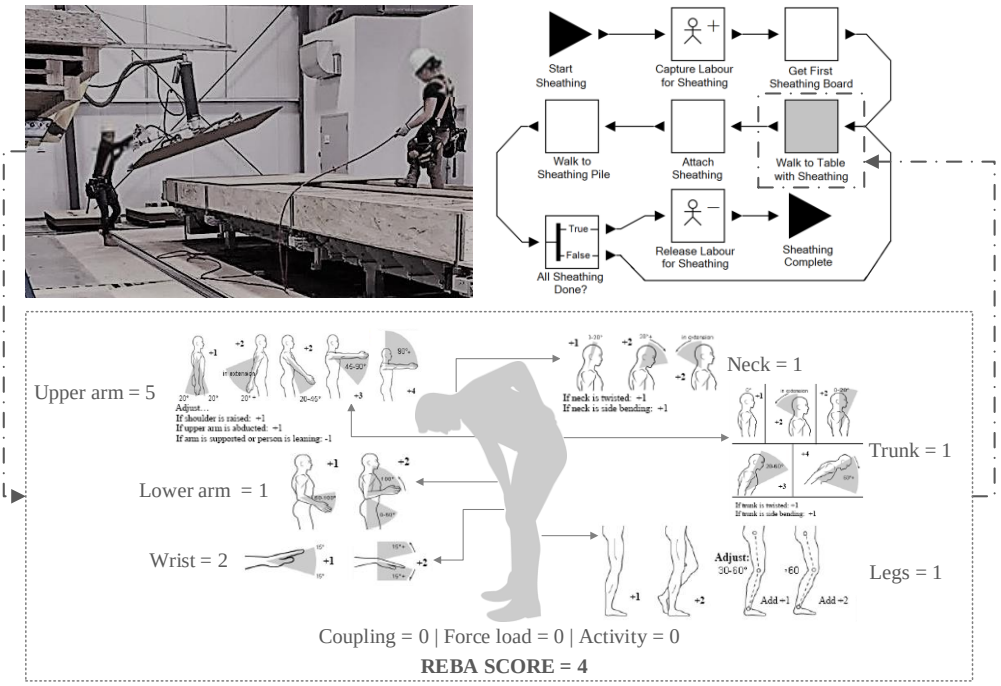


Figure 1: REBA score of walking to table with sheathing task

$$REBA_{sim} = \frac{\sum_{i=1}^{n}(REBA_i \times d_i)}{\sum_{i=1}^{n} d_i} \quad (1)$$

Table 2: REBA score, risk level, and action (adapted from Hignett & McAtamney 2000).

REBA Score	Risk Level	Action
1	Negligible	Not necessary
2-3	Low	May be necessary
4-7	Medium	Necessary
8-10	High	Necessary soon
11-15	Very high	Necessary now

2.3. Implementation Recommendations for Production Alternatives

Once a productivity impact and an ergonomic risk assessment were both complete for the current and possible future alternatives for the identified tasks of the floor panel production process, a more complete quantification of the impact of the proposed changes can be made. To determine a thorough estimate of the effect that a change may have on the production line, it is necessary to combine the ergonomic risk factor with the productivity change. To do this a detailed simulation model with the times for individual activities within each task was made to include the REBA score along with the time to complete the task, so that an overall idea of the extent to which each score was experienced for the variable tasks could be realized. The ergonomic risk as well as the expected average time to complete the task for each alternative considered in this study are detailed in Table 3.

Table 3: Ergonomic risk and production time

Task	Maximum Postural Hazard Score (REBA)	Expected Average Task Time (mins)
Place joist with one person (done half of the time)	10	0.4
Always place joist with two people	10	0.4
Always use MFB to apply glue	1	3.8
Glue manually when MFB is in use	6	5.0
Place sheathing with one person using a vacuum lift	6	1.3
Place sheathing with two people using a vacuum lift	8	0.9
Place sheathing with two people using material delivery bridge	4	0.6

As observed in Table 3 and Figure 2, a relationship between the number of workers performing a task and the task's REBA score is not identified in this paper. For instance, having one or two people placing the joists in the workstation did not result in different REBA scores. This happens because the weight of the joists is greater than 10 kgs, even when shared between two workers, and thus it receives the highest score in the load/force category of REBA. The REBA score might be different in cases of lighter materials. Furthermore, more than one person can be sharing a

task but not necessarily performing the same subtask, as illustrated in Figure 2 (right side), which results in the distinct REBA scores of each worker. It is important to clarify that the maximum REBA score is used to conduct the analysis in this study.

Figure 2: Samples of workers' body movements used to conduct ergonomic analysis.

Workers' behaviour and movement preferences are also found to be a key aspect when conducting ergonomic risk analysis, as exemplified in Figure **3**, which illustrates one and two workers attaching the sheathing. While the worker in the image on the left is standing with his/her legs slightly angled, the worker in the image on the right is squatting to perform the activity, which results in a higher REBA score due to the angle of his/her legs and thus increasing the risk of WMSD, especially if accounting for the fact that this posture is repeated several times until completion of the task. This shows the importance of providing training for workers with focus not only on the operational aspect of the task but also on how to perform it minimizing ergonomic risks. To reduce the discrepancy between the REBA scores of different workers, standardization with respect to how to conduct tasks in an ergonomic manner is recommended.

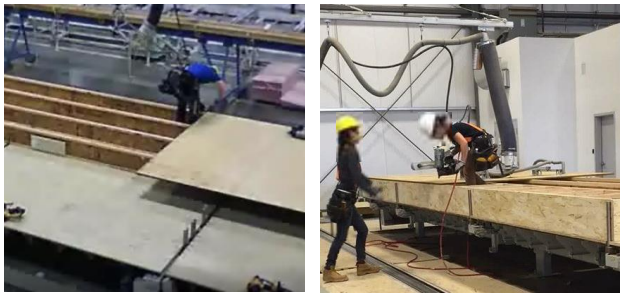


Figure 3: Attaching sheathing with one worker (left) and two workers (right).

2.4. Quantification of Results

In the case of competing goals, it may be necessary to apply a metric to simplify the process of selecting a final decision from the possible alternatives. First, the factors are normalized using equation 2 and 3, where F_E is the ergonomics factor, F_P is the productivity factor, P_{sim} is the simulated productivity for the scenario, and P_{goal} is the goal productivity, which was 2.1 panels per hour.

$$F_E = \begin{cases} 1, if\ REBA_{sim} \leq 3 \\ 1 - \frac{1}{8}(REBA_{sim} - 3), if\ REBA_{sim} > 3 \\ 0, if\ REBA_{sim} \geq 11 \end{cases} \quad (2)$$

$$F_P = \begin{cases} \frac{P_{sim}}{P_{goal}}, if\ P_{sim} < P_{goal} \\ 1, if\ P_{sim} \geq P_{goal} \end{cases} \quad (3)$$

As can be seen by these equations, F_E is 1 (meaning the ergonomics goal is met), when the REBA score is between 1 and 3 and 0 when there is a high ergonomic risk, or a REBA score of 11 or above. Between these values, F_E is assumed to be a linear incremental risk factor. Similarly, F_P is 1 when the productivity goal is met or exceeded, while an F_P of 0 would indicate that no productivity change was accomplished.

Next, the factors are combined. For this scenario, the two factors are assumed to be equal. For other cases where there are multiple factors and they may not have equal weights on the final decision, surveys of the stakeholders to determine the relative weighting of each factor may be required. The formula for the final metric, or decision factor (F_D) is shown in equation 4.

$$F_D = 0.5F_E + 0.5F_P \quad (4)$$

A decision factor of 1 shows that both goals of minimizing the REBA scores of the activities and improving the productivity are met.

3. RESULTS

When ergonomics are taken into consideration along with other common decision making factors, such as capital cost, operating cost, implementation time, and expected productivity increase, it is found that the decision as to whether or not to implement a production alternative may change, or further investigation may be required before an alternative can be recommended for implementation. The results for the total productivity and average REBA score for each of the twelve alternatives created by combining the station alternatives in various ways can be seen in Table 4.

Based on these results, it can be determined that options 3, 6, 9, and 12 result in the highest productivities while options 3, 6, 4, and 10 result in the lowest REBA scores. In this case, two of the most ideal options for each of the considerations overlap, so options 3 and 6 are recommended; however, options 9 and 12 are also considered ideal as they result in similar productivities with a minimal increase in REBA score. It is important to also note that the two options with the lowest REBA scores (options 4 and 10) also result in the lowest productivities of all of the options. Due to the minimal difference between the REBA scores for these options and the REBA scores for the options with the highest productivities, these two options are not recommended for implementation. These results also illustrate the importance of analysing how the production changes affect the two variables together. An example of this is

Table 4: Simulation results for suggested production alternatives

Option	Place Joists Method	Gluing Method	Sheathing Method	Total Productivity (panels/hour)	Average REBA Score	Decision Factor (F_D)
1	Place with one person (50% of panels) and two people (other 50% of panels)	Always use MFB to apply glue	Place with one person using a vacuum lift	1.737	2.881	0.91
2			Place with two people using a vacuum lift	1.859	3.890	0.89
3			Place with two people using a material delivery bridge	2.054	2.083	0.99
4		Glue manually when MFB is in use	Place with one person using a vacuum lift	1.473	1.957	0.85
5			Place with two people using a vacuum lift	1.985	2.546	0.97
6			Place with two people using a material delivery bridge	2.137	2.016	1.00
7	Always place joists with two people	Always use MFB to apply glue	Place with one person using a vacuum lift	1.740	3.049	0.91
8			Place with two people using a vacuum lift	1.870	4.070	0.88
9			Place with two people using a material delivery bridge	2.063	2.288	0.99
10		Glue manually when MFB is in use	Place with one person using a vacuum lift	1.473	1.957	0.85
11			Place with two people using a vacuum lift	2.004	2.724	0.98
12			Place with two people using a material delivery bridge	2.165	2.185	1.00

the difference between option 4 and option 6. Option 6 may be assumed to have a lower overall REBA score due to the lower score of the sheathing method (the differentiating activity) chosen in option 6; however, the simulation analysis shows that option 4 actually results in the lower REBA score of the two options due to the difference in the number of people and duration of the two different methods of installing the sheathing.

While, in this scenario, the best options for implementation can be determined by looking at the simulated results for the overall productivity and REBA score, the combined decision factor (F_D) can be utilized to make this conclusion faster and clearer. As seen in Table 4, options 6 and 12 satisfy both goals, as evidenced by their decision factor of 1, while options 3 and 9 are reasonable options as well, both with a factor of 0.99.

Future work for this analysis will involve investigating the calculation of the decision factor further. Currently, a REBA score of 4, or one point above the acceptable score range will decrease the decision factor by the same point as a productivity that is 12.5% below the goal productivity. Additionally, the effect of the REBA scores on the decision factor are currently being considered as linear, which requires further investigation.

4. CONCLUSION

The case study covered in this exercise shows that, when compared to considering only the effects on productivity that are made by process changes, including the ergonomic impacts of a process change will allow management at offsite construction facilities to make more informed decisions.

While the productivity impacts, cost, ergonomic influence, production interruption, and other factors are all important when making a decision to change a process, it is difficult to determine the relative importance of each consideration and the overall collective impact they have on the operations of a facility. In this case, the ergonomic and productivity impact have been considered together to analyse the decision alternatives and, in the future, more variables can be added to continue to simplify the decision-making process for management at these facilities. Another future improvement to the model will be to consider the weight that each decision variable has in the final decision.

ACKNOWLEDGMENTS

The authors appreciate the technical writing assistance of Jonathan Tomalty and Kristin Berg. In addition, special thanks to everyone at the ACQBuilt factory in Edmonton, AB, Canada, for allowing us to observe the actions at the factory, both in person and through video cameras, in order to collect the data required for this study. This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Industrial Research Chair program (grant file no. IRCPJ 419145-15).

REFERENCES

Abaeian H., Inyang N., Moselhi O., Al-Hussein M., El-Rich M., 2016a. Ergonomic Assessment of Residential Construction Tasks Using System Dynamics. Proceedings of the 33rd Annual International Symposium on Automation and Robotics in Construction, pp. 258-256. July 18-21, Auburn, AL, USA.

Abaeian H., Moselhi O., Al-Hussein M., 2016b. System Dynamics Model Application for Ergonomic Assessment of Manual Material Handling Tasks. Modular and Offsite Construction Summit Proceedings, pp. 240-247. September 29-October 1. Edmonton, AB, Canada.

Abbasian-Hosseini S., Nikakhtar A., Ghoddousi P., 2014. Verification of Lean Construction Benefits through Simulation Modeling: A Case Study of Bricklaying Process. KSCE Journal of Civil Engineering. 18(5):1248-1260.

Armstrong T., Buckle P., Fine L., Hagberg M., Haring-Sweeney M., Martin B., Punnett L., Silverstein B., Sjogaard G., Theorell T., Viikari-Juntura E., 1996. Musculoskeletal Disorders: Work-related Risk Factors and Prevention. International Journal of Occupational and Environmental Health. 2:239–246.

Association of Workers' Compensation Boards of Canada (AWCBC). 2018. 2017 Statistics. Available from: awcbc.org/?page_id=14 [accessed 14 April 2019].

Aqlan F., Al-Fandi L., 2018. Prioritizing process improvement initiatives in manufacturing environments. International Journal of Production Economics. 196:261-268.

Battini D., Delorme X., Dolgui A., Sgarbossa F., 2015. Assembly line balancing with ergonomics paradigms: two alternative methods. 15th IFAC Symposium on Information Control Problems in Manufacturing, pp. 586–91. May 11-13. Ottawa, ON, Canada.

Battini D., Faccio M., Persona A., Sgarbossa F., 2011. New methodological framework to improve productivity and ergonomics in assembly system design. International Journal of Industrial Ergonomics. 41:30–42.

Bortolini M., Faccio M., Gamberi M., Pilati F., 2018. Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes. Comput. Ind. Eng.

Botti L, Mora C, Regattieri A. 2017a. Integrating ergonomics and lean manufacturing principles in a hybrid assembly line. Computers & Industrial Engineering. 111:481–491.

Botti L., Mora C., Regattieri A., 2017b. Application of a mathematical model for ergonomics in lean manufacturing. Data in Brief. 14:360–365.

Canadian Centre for Occupational Health and Safety. 2017. What are work-related musculoskeletal disorders (WMSDs). Available from: www.ccohs.ca/oshanswers/diseases/rmirsi.html [accessed 15 June 2019].

Colombini D., Grieco A., Occhipinti E., 2002. Risk Assessment and Management of Repetitive Movements and Exertions of Upper Limbs : Job Analysis, Ocra Risk Indices, Prevention Strategies, and Design Principles. Amsterdam ; Boston : Elsevier, ©2002.

Dotoli M., Epicoco N., Falagario M., Costantino N., 2015. An integrated approach for warehouse analysis and optimization: A case study. Computers in Industry. 70:5-69.

Golabchi A., Guo X., Liu M., Han S., Lee S., AbouRizk S., 2018. An integrated ergonomics framework for evaluation and design of construction operations. Automation in Construction. 95:72–85.

Golabchi A., Han S., Fayek AR., 2016. A fuzzy logic approach to posture-based ergonomic analysis for field observation and assessment of construction manual operations. Canadian Journal of Civil Engineering. 43(4):294–303.

Golabchi A., Han S., Seo J., Han S., Lee S., Al-Hussein M., 2015. An Automated Biomechanical Simulation Approach to Ergonomic Job Analysis for Workplace Design. Journal of Construction Engineering and Management. 141(8):1–12.

Hignett S., McAtamney L., 2000. Rapid Entire Body Assessment (REBA). Applied Ergonomics. 31(2):201-205

Hochdörffer J., Hedler M., Lanza G., 2018. Staff scheduling in job rotation environments considering ergonomic aspects and preservation of qualifications. Journal of Manufacturing Systems. 46:103–14.

Ikuma LH., Nahmens I., James J., 2011. Use of Safety and Lean Integrated Kaizen to Improve Performance in Modular Homebuilding. Journal of Construction Engineering and Management. 137(7):551-560.

Inyang N., Han S., Al-Hussein M., El-Rich M., 2012. A VR model of ergonomics and productivity assessment in panelized construction production line. Construction Research Congress 2012: Construction Challenges a Flat World, pp. 1084–1093. May 21-23. Lafayette, IN, USA.

Jia Z., Lu X, Wang W., Jia D., 2013. Design and Implementation of a Lean Facility Layout System of a Production Line. International Journal of Industrial Engineering. 20(7-8):502-514.

Karhu O., Kansi P., Kuorinka I., 1977. Correcting working postures in industry: A practical method for analysis. Applied Ergonoics. 8(4):199–201.

Li G., Buckle P., 1999. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. Ergonomics. 42(5):674–695.

Li H., Lu M., Hsu S-C., Gray M., Huang T., 2015. Proactive behavior-based safety management for construction safety improvement. Safety Science. 75:107–117.

Li X., 2017. Ergonomic Risk Assessment in Construction Manufacturing Facilities. Thesis (PhD). University of Alberta.
Li X., Gül M., Al-Hussein M., 2019. An improved physical demand analysis framework based on ergonomic risk assessment tools for the manufacturing industry. International Journal of Industrial Ergonomics. 70:58–69.
Li X., Han S., Gül M., Al-Hussein M., El-Rich M., 2017a. 3D visualization-based ergonomic risk assessment and work modification framework and its validation for a lifting task. Journal of Construction Engineering and Management. 144(1):4017093.
Li X., Komeili A., Gül M., El-Rich M., 2017b. A framework for evaluating muscle activity during repetitive manual material handling in construction manufacturing. Automation in Construction. 79:39–48.
McAtamney L., Nigel Corlett E., 1993. RULA: a survey method for the investigation of work-related upper limb disorders. Applied Ergonomics. 24(2):91–99.
Mossa G., Boenzi F., Digiesi S., Mummolo G., Romano VA., 2016. Productivity and ergonomic risk in human based production systems: A job-rotation scheduling model. International Journal of Production Economics. 171(Part 4):471–77.
Stanford University, 2012. OpenSim User's Guide.
Public Services Health & Safety Association (PSHSA). 2010. Repetitive work: Could you please repeat that … again and again and again? Available from: www.pshsa.ca/wp-content/uploads/2013/01/RepetitiveWorkInjury.pdf [accessed 21 April 2019].
Rajabalipour Cheshmehgaz H., Haron H., Kazemipour F., Desa MI., 2012. Accumulated risk of body postures in assembly line balancing problem and modeling through a multi-criteria fuzzy-genetic algorithm. Computers & Industrial Engineering. 63:503–512.
Ritter C., Li X., Al-Hussein M., 2016. Production monitoring and process improvement for floor panel manufacturing. Proceedings, Modular and Offsite Construction Summit, pp. 75–82. September 2—October 1. Edmonton, AB, Canada.
The University of Michigan Center for Ergonomics, 2017. 3D Static Strength Prediction Program™ Version 7.0.0: User's Manual.
Yu H., Al-Hussein H., Al-Jibouri S., Telyas A., 2013. Lean Transformation in a Modular Building Company: A Case for Implementation. Journal of Management in Engineering. 29(1):103-111.
Xu Z., Ko J., Cochran DJ., Jung M-C., 2012. Design of assembly lines with the concurrent consideration of productivity and upper extremity musculoskeletal disorders using linear models. Computers & Industrial Engineering. 62:431–41.

AUTHORS BIOGRAPHY

Chelsea Ritter is a PhD student in the Construction Engineering and Management program at the University of Alberta. Her current research focuses on process and layout analysis and improvement methods for offsite construction facilities.

Regina Dias Barkokebas is a PhD student in the Construction Engineering and Management program at the University of Alberta. Her current research focuses on improving health and safety in industrialized construction by assessing ergonomic risks and proposing solutions to minimize safety risks utilizing virtual reality (VR).

Xinming Li is an assistant professor in the Mechanical Engineering Department at the University of Alberta. Her research seeks improvements in industrialized construction by evaluating ergonomic risks and investigating corresponding corrective measures to secure the health and safety of workers and enhance workplace productivity. Her areas of focus include physical demand analysis, human body physiological measurement, ergonomic risk assessment based on 3D visualization modelling of operational tasks, and lean manufacturing in industrialized construction.

Mohamed Al-Hussein is a is a professor in the Hole School of Construction Engineering and Director of the Nasseri School of Building Science and Engineering at the University of Alberta, and holds the NSERC Industrial Research Chair in the Industrialization of Building Construction. A highly sought researcher and consultant in the areas of lean manufacturing, construction process improvement, CO_2 emission quantification and reduction, and building information modelling (BIM), Dr. Al-Hussein has successfully applied lean principles to improve work methods and productivity standards for various industries and projects. His research has developed best practices for panelised building systems, lean production, and modular construction, and has been published in approximately 200 peer-reviewed journals and conference proceedings.

WEARABLE MIXED REALITY SOLUTIONS FOR INDUSTRIAL PLANTS AND PRODUCTION LINES

Agostino G. Bruzzone[a], Marina Massei[b], Kirill Sinelshchikov[c], Francesco Longo[d], Matteo Agresta[e], Luciano Di Donato[f], Cesare Di Francesco[g]

[a] Simulation Team
[b], [c] DIME University of Genoa
[d] MSC-LES, University of Calabria
[e], [g] Liophant Simulation
[f] INAIL

[a]agostino.bruzzone@simulationteam.com, [c]kirill@simulationteam.com, [e]matteo.agresta@simulationteam.com,
[g]cesare.difrancesco@simulationteam.com
[b] massei@itim.unige.it
[d]francesco.longo@unical.it
[f] l.didonato@inail.it

ABSTRACT
In the present research project, the authors developed wearable and portable solutions capable to improve safety in production lines by taking advantages of availability of exhaustive and reliable data in modern industrial plants. Indeed, the synergy between Industry 4.0 and cutting edge devices, such as smartphones and headsets for Mixed Reality demonstrated to be potentially used to assist personnel on the shop floor, especially during critical and most dangerous operations. In this paper it is presented an ongoing project devoted to develop such support systems and to evaluate their efficiency in multiple industrial environments.

Keywords: Modeling, Simulation, Augmented Reality, Mixed Reality, Virtual Reality, Wearable devices

1 INTRODUCTION

Modern industrial plants and production lines are characterized by new generations of control systems that include usually quite advanced equipment solutions including numerous sensors as well as capability to collect and transfer an huge amount of reliable data. Typically. these data allows to better control the situation and to activate innovative service such as production optimization and predictive maintenance. therefore there is a significant need to develop Intelligent Solutions able to filter and elaborate these data in order to extract valuable information. So it is evident that these machines represent a potential value and enable new approaches; these "smart" machines could operate in centralized/decentralized mode, feeding directly other devices and systems with a variety of data. Indeed, these capabilities, often grouped under Industry 4.0 name, are crucial to be competitive in today's market. Obviously, the information obtained from these sources could be used not only to optimize costs and efficiency, but even to improve safety.

2 NEW DATA SOURCES: IIOT

Recent IIoT (Industrial Internet of Things) developments are interesting for their capability to extend also already existing capabilities and to gather large amounts of data (Big Data) that enable the introduction of new solutions able to integrate distributed systems and new components within networks. Due to these considerations, it is possible to develop innovative solutions for monitoring, tracking and maintenance in industrial plants and to support operations on them.
The availability of Big Data from production lines and industrial plants, especially regarding machines' state, it is the reason why nowadays it is quite common to test and introduce new solutions based on Artificial Intelligence (AI), Digital Twins, Augmented Reality (AR) and Autonomous multi-domain Vehicles (AxV) (Vignali et al. 2018, Bruzzone et al. 2017, Bruzzone et al. 2016). Indeed, the project proposed in this paper is mostly focused on use of Mixed Reality intended as interactive combination of Virtual and Augmented Reality and it is based on the development of an approach to guarantee reliable, up-to-date and exhaustive information from different data sources.

3 OBJECTIVES

As mentioned, the availability of information, generated from rich and reliable data sources, provides numerous opportunities of improvement for different aspects of the plants' activities, from performance and productivity upgrade to enhancement of security and safety. In this particular research the authors address several aspects; indeed, in the proposed case it is expected to improve both safety and work optimization. Information regarding systems' state could be used as warning about potentially dangerous conditions as well

as to reduce time of response; furthermore, there are new additional functionalities, such as remote assistance: respect this aspect, for instance, it turns possible to provide support to operators in front of the production lines through remote assistance from highly qualified personnel, available at any time in Headquarters; this aspect create a new kind of service able to guarantee highly qualified remote support whatever the distance between remote experts and on-site operators.

To address such aspects, the authors propose portable, mobile and wearable solutions in order to assist operators in performing their daily tasks on the shop floor, along production lines and inside industrial plants. In particular, the idea investigated is related to the development of flexible solutions able to be used over multiple platforms from HoloLens to smartphones and tablets, which might be used not only to display information, but also to acquire and store data about interventions and information related to lesson learned. This research is deriving from multiple industrial initiatives; among them it should be mentioned the W-ARTEMYS project (Wearable Augmented Reality for Employee safety in Manufacturing sYStems) that aims to analyze how these new tools and solutions impact on safety within industrial plants.

According to these considerations, the following objectives have been identified:

- To define Key Plant Indicators and most important risks faced by operators to be provided within the Mixed Reality (MR) in relation to crucial machines and lines in order to alert users on these aspects
- To develop innovative Software solutions able to integrate Industry 4.0 Equipment with Mixed Reality, Simulation and AI, developing also smart solutions for data acquisition, transfer, elaboration and visualization
- To carry out experiments in order to validate the approach and estimate the potential of alternatove solutions applied on site

4 AUGMENTED PLANT OPERATOR

As anticipated, this research focuses on testing innovative solutions to support supervision and maintenance for industrial production line using MR technologies; special attention has been paid to use of Augmented Reality to highlight criticalities and performance during operations. Indeed, one aim of this research was also to evaluate efficiency of various types of wearable and portable hardware and software solutions, considering also limitations imposed by the different kinds of production lines; for instance in some case, these environments could be characterized by excessive temperature, high level of noise, dust or even presence of chemical agents. In particular, the authors are testing solutions that are flexible and portable on different platforms including smartphones, tablets and HoloLens.

Augmented Reality and wearable technologies are very good to alert the shift managers and on-site operators about the necessity of immediate intervention. In this way it is possible to provide intuitive and quick access to the main plant indicators as well as to simplify retrieving of information regarding specific machines and production parameters. However, to create these functionalities it is necessary to integrate portable devices and to let them know their position relative to the line and access common databases. Depending on the particular requirements it would be possible to use Indoor Positioning System (IPS) or by attaching QR "anchors" to predefined places of the plant. IPS is one of the most widespread approaches for positioning inside buildings as well as outside. In fact, it is common to use the satellite positioning in many applications, for example GPS; however, within an industrial plant or along a production line, usually, these signals turn to be weak, there are reflections and bad interferences. Fortunately, today there are various tools that allow to finalize the position using transmitters installed in fixed positions within the plant. For example, IPS based on Wi-Fi and Bluetooth signals allow to find the position quite precisely and even with minimum infrastructure investment, otherwise there are specialized and more precise systems that use UWB (Ultra-Wide Band) signals (Yang & Shao, 2015). In facts, in the proposed case studies, considering they require an average precision level positioning, could use any of the mentioned systems. In some cases, absolute positioning on the production line is not required, hence, to avoid unnecessary investments it is possible to use QR codes.

Indeed it should be stated that in specific cases, the QR code scanning could be effectively used as way to confirm that the operator/mobile device has reach a specific point in the plant.

Indeed, in this case, all machines of interest are equipped with such tag, which aids finding exact relative position and rotations of the client device, once the code is found. Furthermore, mobile and wearable devices are typically equipped with Inertial Measurement Units (IMU), which allow to keep control over position even after the loss of tag tracking. Hence, even using tags combined with IMU it is possible, usually, to achieve sufficient positioning capabilities.

In any case, the positioning solutions allow to improve efficiency of the operators in many ways and are often a crucial element of a solution involving wearable technologies. First of all, in this way, it turns possible to provide up-to-date information of plant situation to the operators: i.e. alerts, abnormal behavior and machines' state as well as various "symptoms" of possible criticalities, production performance anomalies, increase on defects, etc.. In addition, by transmitting the information about the operator activities to the traditional control room, central office and/or engineering department, it is possible to improve efficiency of the remote supervision to operations. Indeed, the synchronization of wearable technology devices in this process allows to know exactly where an operator is located and even what he is looking at.

This approach could be further improved by creating a Digital Twin of the production line, or of the entire plant, and to move operators in this virtual world. In such case, remote supervisor could have access to all plant indicators in a virtual 3D environment and assist the operator on site, having the same vision of the plant (Virtual) as the on-site operator (Real), (see figure 1) (Bruzzone et al., 2019a). Obviously, such synergy of technologies could enable even more benefits.

Figure 1: HoloLens for remote supervision, laboratory demonstration

4 SCENARIOS OF INTEREST

In order to develop useful and efficient solutions based on this approach, the authors identified several possible cases of application: one related to hollow glass production, while the second is mainly focused on food & beverages.

Glass Production Line

Producing hollow glass operates within an "hostile environment"; indeed, the process includes melting the raw material in furnaces at high temperature, hot and cold treatment along noisy production lines, numerous controls and packaging systems. Obviously, the initial parts of such lines have a massive presence of machines and materials at high temperature, while, control and packaging deal with highly fragile and dangerous products, which often causes presence of cracked glass around. Furthermore, automated packaging and transportation systems require to address properly the risk of collisions. In general these environments are considered quite dangerous (BLS, 2017). An interesting peculiarity of this environment is that risks are caused by both machines and products themselves.

Food and Beverage Production Line

The second case study is related to a bottling line, which covers processes as cleaning and sterilization, bottling and sealing the containers once they are filled up with milk. In this case the main risk factors are related to presence of elevated temperature and pressures in the machines as well as to moving parts.

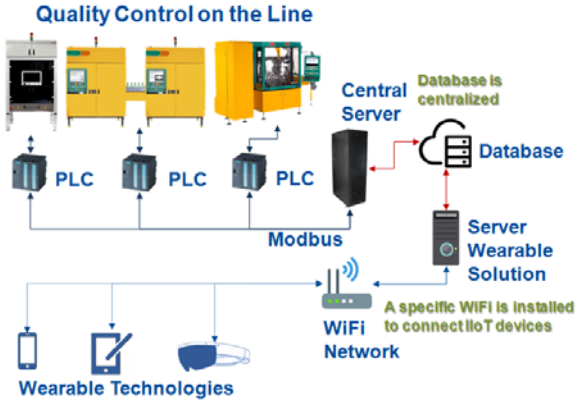


Figure 2: Architecture related of Quality Control along a Production Line

5 PROPOSED SYSTEM ARCHITECTURE

To develop flexible solutions for the proposed case studies, it is necessary developed an heterogeneous and robust support architecture integrating different equipment, models and wearable technologies as well as the operators themselves (Longo et al. 2012). Indeed the necessity to create reliable models of the production process it result critical to properly connect them with the virtual representation of the machines and components. An example for online quality control a production plant is summarized in figure 2. In order to succeed in such task and avoid multiple possible pitfalls, it is essential to define properly the whole architecture that allow this system to perform all required activities, maintaining high level of reliability, efficiency as well as cyber security. For instance, in order to limit access to machinery and keep data acquisition efficient, simple and consistent, it is preferable to have a single access point for such information, for example by providing a web API (Application Programming Interface) or database access. Such centralized approach avoids the redesign the data extraction modules for each different client device, which is very a important aspect considering the variety of access modes and data formats of typical production lines composed by multiple machines produced by different manufacturers. Furthermore, a centralized data source such as a database allows introduction of new functions as well as extension of existing ones, while the entire system could be easily adopted for other production lines and sites. Thus, this approach allows to access the data in a simpler and more effective way, making the project server and clients completely oblivious regarding particularities of access to the different machines and maintaining the highest possible level of security (Tanenbaum & Van Steen, 2017). Summarizing the proposed solution, the data are extracted and saved in centralized database,

periodically retrieved by the server and consequently sent to the clients. In this case the server performs also data elaboration, analyzing the alerts and proposing possible explanations and corrective actions.

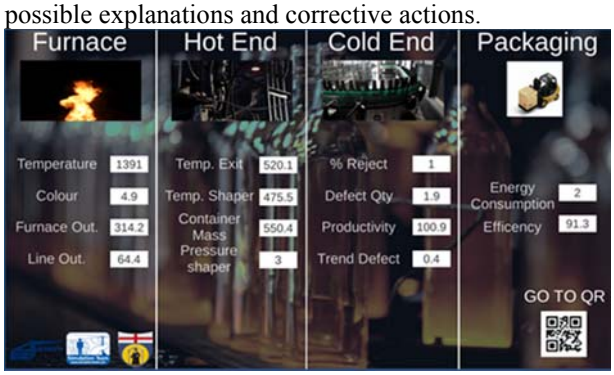


Figure 3: Mobile application displaying main plant indicators

Indeed, the server's functionality could be improved through an Artificial Intelligence module, which would allow, not only to indicate critical situations, but also to foresee them, reducing the risks both for operators and for the plant.

As mentioned, another important aspect of the project is to guarantee security considering that sensible data are shared and operators act inside dangerous environments and within highly cost production processes; in this framework the cyber threats aims not only to security, but also safety issues; for instance, the production line's data are taken directly from the company's servers. Obviously, the maximum degree of security would happen when all the systems were isolated from one another, making however any kind of integration, centralization, control and remote maintenance impossible. Hence, compromises must always be made between safety and efficiency. For these reasons the case study employs only portable solutions capable to maintain high level of data protection. Starting from the proposed architecture it is possible to evaluate the most critical points in terms of data protection and access control.

For instance, considering that in proposed cases, the main server could have access also to some parts of the control system of production lines, it is clear the need to control access to sensitive data. Partially, this issue is addressed by the read-only access to the databases, which reduce drastically possibility of unauthorized modifications to the machines' configuration. In the same time, all but the database's port of the server must be blocked by the firewall. Another criticality related to cyber security is related to the use of wireless networks (Wi-Fi) as support solution for the new system. From this point of view, it is very important to provide high level of data protection and integrity control (Ferguson, 2010). For example, WPA2 protection must be used for connections, while all connected devices must operate within a secure VPN (Virtual Private Network) network (Nguyen, 2018). Obviously, portable devices must be forced to use only the dedicated wireless network. In the scenario of interest such protection level could be sufficient, however, in cases when a public network is required, it could be possible to introduce additional encryption level in order to improve overall protection level, for instance, by introducing probabilistic cryptography layer, such as that one based on ElGamal.

Last, but not least: security, but also safety aspects are strongly related to human factors (Bruzzone et al. 2019b). In facts, all security measures were useless if the end user would access external or mobile networks, creating connection bridges between networks, or to install unauthorized applications. Considering this, it is essential implementing access control on the devices, introducing policies and encryption.

6 USE MODE AND EXPERIMENTATION

In particular, the authors have identified different possible use mode for these new systems in industrial plants. One is for sure related to support directly the operators within the plant; in this context AI and wearable technologies help the operator to understand quickly problems and criticalities and it is directed over there and guided to solve. Still in this framework the proposed solutions could support remote assistance to operator to solve more complex situation. Another important use mode is related to the idea to adopt these Mixed Reality Solutions to train operators off line and on line and to improve their familiarity with problem solving and interventions on the productions Lines.

From this point of view an example is proposed in relation to the combined use of smartphone and HoloLens to monitor state of the plant; vice versa headset and or CAVE could be also used for remote assistance and training. For instance it is possible to use a tablet to support operators during maintenance, providing them with data on machine alerts and action to be carried out/components to be checked. (see figure 3). Currently, the authors are still testing the new systems and experiencing other functionalities

CONCLUSIONS

The authors have carried out a research on modern mobile and wearable technologies and their possible integration with Production Plant. The experience was successful both with new industry 4.0 machines and old components inside different kind of production lines and it was possible to identify synergies with available information systems and to integrate data in modern plants in order to improve safety. The project is in active development phase and experimentation are ongoing; however, Subject Matter Experts (SME) have been involved since the beginning and currently they confirm their interest in these solutions, based on preliminary experimental results.

REFERENCES

BLS (Bureau of Labour Statistics) of United States (2017). Retrieved from www.bls.gov.

Bruzzone, A., Longo, F., Nicoletti, L., Vetrano, M., Bruno, L., Chiurco, A., Fusto, C. & Vignali, G. (2016). Augmented reality and mobile technologies

for maintenance, security and operations in industrial facilities. Proc. of 28th EMSS 2016, Larnaca (Cyprus), September 26-28, pp. 355-360.

Bruzzone, A. G., Massei, M., Longo, F., Sinelshchikov, K., Di Matteo, R., Cardelli, M. & Kandunuri, P. K. (2017). Immersive, Interoperable and Intuitive Mixed Reality for Service in Industrial Plants. Proceedings of the 16th International Conference on Modelling & Applied Simulation, MAS 2017, part of I3M 2017, Barcelona (Spain), September 18 -20, pp.208-213.

Bruzzone, A.G., Cianci, R., Sciomachen, A., Sinelshchikov, K. & Agresta, M. (2019a). A digital twin approach to develop a new autonomous system able to operate in high temperature environments within industrial plants. Proceedings of Summer Simulation Conference, Berlin (Germany), July 22 – 24.

Bruzzone, A. G., Fancello, G., Daga, M., Leban, B., & Massei, M. (2019b). Mixed reality for industrial applications: interactions in human-machine system and modelling in immersive virtual environment. International Journal of Simulation and Process Modelling, 14(2), 165-177.

Ferguson, N., Schneier, B. & Kohno, T. (2010). Cryptography Engineering: Design Principles and Practical Applications. Wiley.

Longo, F., Massei, M., & Nicoletti, L. (2012). An application of modeling and simulation to support industrial plants design. International Journal of modeling, simulation, and scientific computing, 3(01), 1240001.

Nguyen, H. G. B. (2018). Wireless Network Security: A Guide for Small and Medium Premises, Lathi University of Applied Science

Tanenbaum, A. S., & Van Steen, M. (2017). Distributed systems. Prentice-Hall, NYC

Vignali, G., Bertolini, M., Bottani, E., Di Donato, L., Ferraro, A. & Longo, F. (2018). Design and testing of an augmented reality solution to enhance operator safety in the food industry. International Journal of Food Engineering Volume 14, Issue 2, 23 February 2018, Article number 20170122

Yang, C., & Shao, H. R. (2015). WiFi-based indoor positioning. IEEE Communications Magazine, 53(3), 150-157.

BIOGRAPHIES

Agostino G. Bruzzone is full professor in Genoa University, President of Simulation Team and currently he serves as Council Chair of STRATEGOS the new MSc in Strategic Engineering and President of the Master Program in Industrial Plant Engineering and Technologies of Genoa University

Marina Massei is active in Modeling and Simulation from over 10 years with crucial roles in R&D Projects applied to Defense, Homeland Security, Industry and Service to the Society, she has highly qualifications in NATO and Academic Courses related to innovative technologies and acts as Project Controller in R&D initiatives

Kirill Sinelshchikov has completed the International Master Program in Industrial Plant Engineering and Technologies and he is active as M&S Researcher working on multiple projects related to industrial applications in Hot Metals, Glass Industry, Ports, Logistics, etc.

Francesco Longo is coordinator of MSC-LES Labs and serve as professor at UNICAL, he is very active in Modeling and Simulation Community since several years with special attention to industrial plant, cultural heritage and logistics.

Matteo Agresta completed his PhD in Transportation Engineering and works on use of M&S applied to Logistics, Industrial Plant Engineering and Technologies

Luciano Di Donato is a senior researcher in Italian National Agency for Safety on Work Place with special attention to use of innovative technologies in Industries.

Cesare Di Francesco is active in Liophant conducting researches on innovative solutions based on IoT, Virtual Reality and Simulation applied to Industrial Engineering field

INTELLIGENT AUTONOMOUS SYSTEM DEVOTED TO IMPROVE EFFICIENCY AND SAFETY IN INDUSTRIAL PLANT

Agostino G. Bruzzone[(a)], Matteo Agresta[(b)], Kirill Sinelshchikov[(c)], Marina Massei[(d)]

[(a), (b), (c)] Simulation Team, DIME University of Genoa
[(d)] SIM4Future

[(a)]agostino.bruzzone@simulationteam.com, [(b)]matteo.agresta@simulationteam.com,
[(c)]kirill.sinelshchikov@simulationteam.com
[(d)]marina.massei@sim4future.com

[(a),(b),(c)] www.itim.unige.it
[(d)]www.sim4future.com

ABSTRACT

The industrial plants, in process industries, are often characterized critical environments and dangerous conditions respect the utilization of human workforce.
These aspects are quite common in chemical, oil & gas, iron & steel plants where this combination of factors causes serious accidents. The aim of this study is to develop a new generation of intelligent autonomous system capable to support operations in these environments reducing the presence of workers in most risky operations. In order to succeed in this task, these new systems must be suitable to operate in complex industrial environments and to face related threats and risks. Indeed, this paper proposes an analysis of a complex industrial environment related to hot metal production by identify the most dangerous areas and operations as well as potential critical elements for autonomous vehicle use. Consequently, simulation solutions are proposed to support engineering of these innovative robotic system as well as development of new operations and procedures.

Keywords: Autonomous Systems, Virtual Prototyping, Safety

1 INTRODUCTION

Nowadays, thanks to the new technologies, especially the ones related to autonomous and robotic systems, it is possible to safely carry out operations in industrial areas where human presence could be dangerous (Bruzzone et al. 2017). Indeed, thanks to the advances of the autonomous and intelligent systems, sensors and robotics it became possible to introduce new solutions to already well-known problems, tailoring systems for specific sectors, which in the past were poorly covered, especially in such sector as metal production. Indeed, combination of "old economy" and modern autonomous systems could lead to significant improvement of safety and efficiency in this field and create new services. Such considerations are supported by the statistics and other data provided by experts, which confirms that in the hot metal production there are many work areas that are dangerous for humans and that lethal accidents rate is significantly higher than that in many other industrial fields (Shikdar & Sawaqed 2003). This information suggests introduction of new automated solutions, especially the ones based on autonomous systems, able to deal with complex operational procedures (e.g. monitoring activities, inspections, measurements, etc.) in order to stave off personnel on site from most dangerous zones, or at least to reduce time of exposure to hazards, hence, to reduce risks.

2 SIMULATION AND REQUIREMENTS

Considering complexity of the industrial plants, processes and operations, it is evident that the development of the new systems should be supported by M&S (Modeling & Simulation) to analyze the scenario complexities. From this point of view, it is evident that simulation have to include not only the platform for an UGV (Unmanned Ground Vehicle) or UAV (Unmanned Aerial Vehicle), but also its payload and systems, plus the production line or even of an entire plants.
Obviously, this scenario results pretty complex and it requires to develop not only "physical" parts (e.g. platform, hardware), but also new control systems and new dedicated procedures for the robotic solutions. Furthermore, potential limitations, pitfalls and requirements, related to these new systems, could be discovered by employing Modeling and Simulation. It is important to state that is not necessary always to adopt high fidelity models for these analysis, while resolution and confidence band should be consistent with the purpose of the simulation.
In addition, the data available, or supposed to be available when the simulators and its results are required (e.g. new plant not yet completed when the new autonomous system has to be designed) represent another major constraint in definition of simulation requirements. For instance, in this paper this issue is addressed considering the necessity to evaluate capabilities to operate in high temperature and aggressive environments without having too much detailed data from the field.
In this context, it result necessary to develop a Dynamic Virtual Interoperable Simulation able to combine 3D and thermodynamic models for the industrial scenario. In this way, it become possible to simulate the heating of the autonomous system due to many factors such as irradiation in different initial and boundary conditions. Hence, this is pretty crucial to quantify the dynamic impact of the environmental conditions, also related to the plant operational mode that could change along the

production phases, respect the vehicle and eventually vice versa impact of vehicle operations on the plant components. Indeed, this paper proposes a methodological use of different models to face possible thermodynamic issues much time before to finalize engineering and construction of the physical vehicle. In this way the definition of new operational modes for this new component of the industrial plant could be defined concurrently during design getting benefits of simulation dynamic environment.

3 PLANT OPERATION, TASKS AND CRITICAL ISSUES

As mentioned, process industry is a challenging framework and, specifically, hot metal production is often characterized by dangerous work conditions that impact human personnel, equipment and automated systems (Pardo & Moya 2013). Based on conducted study and experimentations it is identified that the most critical characteristics of the work environment are intensive heating by irradiation, spills of molten metal and occasional presence of corrosive and/or poisonous gas; it is evident that new autonomous systems have to deal with these issues and provide a major advantage for safety. Precisely, the authors focus their attention on simulation of a new UGVs to support inspection and monitoring in hot areas.

Obviously, in this environment a major critical issue is related to the high temperatures due to the proximity of molten metals (e.g. over to 1400ºC). However, the overall conditions are changing dynamically depending on several external factors: e.g. external temperature considering that also indoor area is mostly "open" and current operational mode of the industrial plant; indeed, these facilities are located in very diverse parts of the world, sometime in places with very high or very low temperature (e.g. Siberia and India).

In any case, a very significant part of the heating is caused by irradiation, due to the presence of white hot metal; for instance in runners (channels which are used to transport molten metal) located on a blast furnace during casting operations this has a big impact, while when the tap hole is close the boundary conditions are pretty different (Brimacombe 1999; Geerdes et al. 2015). Near white hot metal infrastructures and even the floor could assume quite high temperature (e.g. over 100 ºC). Hence, in zone adjacent to the most hot parts, the strong irradiation affects also personnel and equipment, but it could turn also critical for the autonomous systems considering the temperature limitation of its electronic components.

Indeed the temperature result to be a constraint also for operation with autonomous vehicles or robotic systems. Furthermore, same zones are characterized by presence of dust, toxic gas, whiffles, sparks and spills of molten metal, which could easily damage the robots. Considering these aspects, it is important to include into the simulation models also these elements.

In order to achieve sufficient fidelity of the simulation, it is essential to employ different interoperable models, able to cover different aspects of the operations. Indeed, the authors adopted the MS2G (Modeling, interoperable Simulation & Serious Game) paradigm to combine heat exchange model with 3D model of the UGV and environment. In such case, data obtained from simulation of spatial state evolution of the vehicle is combined with thermodynamic model which takes into account position of the system of interest respect to the sources of heat and current hot metal production phase on the plant. Hence, the model analyzes, not only accessibility constraints, procedures and interactions with other machines and present equipment, but also the impact of the most important boundary conditions on the plant. The paradigm allowed to employ previously developed solutions based on autonomous systems for emergency management in industrial plants and to facilitate creation of digital twins of these new autonomous systems. In facts these simulations could be used in virtual prototyping, but also for real tests of control systems and operational procedures as well as to improve the operations on site (Bruzzone et al. 2017).

4 DIGITAL TWINS AND DIGITALIZATION AS SUPPORT FOR PLANNING

Digital twin is a high fidelity virtual replica of a physical asset which allows to perform tests and experimentations even on not yet existing systems; this approach gained special popularity in the last years thanks to the technological advances. By this approach, it turns possible to check proposed modifications of real system, to evaluate its efficiency in new boundary conditions, or to use it for training or even to develop new procedures.

Due to these reasons this concept is now largely adopted and supported by simulation, where the digital twins are often considered as the next step after the simulation-based system design (Boschert & Rosen 2016).

In facts, availability of precise and realistic model of a system permits conduction of tests with high precision respect the use modes. In the same time, cost of modification of virtual model is usually much lower in comparison to the physical asset, making possible to carry out fast evaluations of the efficiency respect different configurations, as well as to obtain exploitation characteristics in different conditions.

Obviously, in such cases the cost of initial development of the model could be relatively high and requires high qualified skills and experience. Such solutions could be used not only to test single assets, but to validate the entire systems or even systems of systems (SoS), making it useful tool for strategic decision making on new developments (Massei et al. 2014, Bruzzone 2018).

Indeed, in the presented paper, the authors employed this approach since the beginning of the project, supporting basic and detailed engineering, acquisition of components and production of the new system.

Development of a digital twin could be quite challenging; in the presented case the UGV includes numerous elements and sub-systems, staring from sensors and up to robotic arm, which must be integrated and optimized for different boundary conditions.

For example, the authors analyzed different platforms in combination with distinct propulsion systems, sets of sensors, robotic arms, manipulators, tools, instruments and cooling system. Thanks to the digital twin and virtual prototyping approach the authors identified optimal architecture for the autonomous system within the plant and related operations. As mentioned, another important aspect of digital twins and virtual prototyping is the possibility to test assets in different boundary conditions; the authors used such opportunity to test the proposed solution in distinct layout of real plants in order to optimize configuration and procedures performed by the autonomous system. For example, there were analyzed accessibility constraints, which are often strict due to the presence of bulky equipment; in the same time different tools and instruments were analyzed respect to the required operations. In the same time particular attention was focused on thermodynamic analysis of the new system in such dangerous environment.

Indeed, the authors combined a 3D virtual simulator of the plant operations and procedures with thermodynamic model for estimating temperature dynamics and is impact of UGV. In facts, it is possible to simulate not only the movements of the UGV and to analyze how it is subjected to the heating by irradiation, but potentially to integrate these models with other ones already available to create a more comprehensive scenario (Bruzzone et al. 2016). Indeed, interoperability could be provided by HLA and could guarantee a standard based tight coupling of different models which represent distinct assets and that address different types of interactions. In some cases, it could be convenient to employ also standards such as FMI (Functional Mock-up Interface) in order to support data exchange (Garro et al. 2015). In particular, in this case it was adopted the concept of creating a digital twin to support the development of a new system and related procedures. At the same time this approach allowed to analyze critical issues such as overheating, which is a crucial aspect for this kinds of industrial plants. For example, the approach could be used to design properly an effective and adequate HVAC system (Heating, Ventilation and Air Conditioning) for the UGV as well as to find a correct combination of heat shields and other passive elements which could help to prevent overheating.

5 PRODUCTION PHASES AND RISK ANALYSIS

The presented study aims to stave off personnel from the riskiest operations, or at least to reduce the time it is exposed to the hazards. In order to achieve these goals, it was decided to conduct a specific case study to evaluate efficiency of proposed solution in a real plant.

The proposed case is related to a plant with casting; as presented in the figure 1, a typical cycle of casting is characterized by different hazards related to phases of the process. For example, at the end of the casting phase it is possible to observe whiffles of molten metal coming from the tap hole. Indeed, during the operation, hot metal is melted inside the furnace, forming sometime puddles; in order to discharge the hot metal, it is required to open an hole near the bottom part, wait until the level of liquid metal drops to the minimum (casting), then close the hole and wait until the level in the grows up to sufficient value. In ideal case, the level of metal does not go below the tap hole, keeping this part of the furnace over the molten metal, however, in practice, this level could be too low, opening partially or completely the hole. Obviously, in such case the gases exit from the furnace altogether with the metal, reducing efficiency of the furnace, furthermore, such combination causes a mixture of gas and liquid metal, creating a foam which could be blasted away from the furnace. Such situation is one of most dangerous events in the cast house because the whiffles could fly away to up to tens of meters, potentially damaging equipment or even causing injures of personnel. In such circumstances the personnel is required to perform inspections, sampling and measurements, which is often not very safe. Furthermore, shortly after finish of the casting, when the tap hole is already closed, but the molten metal is still present, it is necessary to carry out cleaning procedures, which could be quite dangerous due to the temperature of surfaces and strong irradiation caused by presence of still hot metal. Indeed, in these cases the temperature could overpass 1400°C.

Obviously, in such short, but dangerous interventions it is convenient to substitute human personnel with dedicated automatic and autonomous systems; in such case the list of activities to be done contains following operations: sampling, measurements, cleaning as well as periodic and emergency inspections (Sarcar et al. 1982; Mailliet & Metz 1993; Nelson 2014; Nelson & Hundermark 2016).

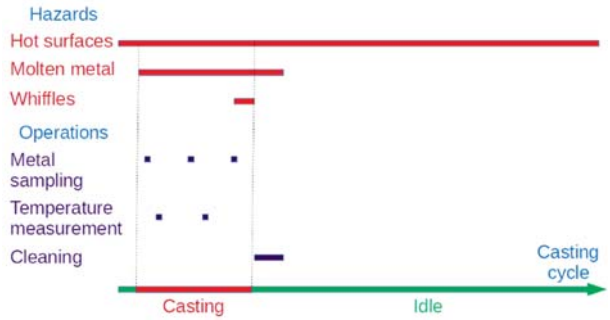


Figure 1. Hazards during casting cycle

It is important to note that the hazards for people and new autonomous systems are essentially similar: both human operators and robotic solutions could be damaged by molten metal, strong irradiation, gases and dust. Indeed, starting from the beginning of the design phase these risks should be assessed properly in order to identify the optimal configuration of the system as well as to choose proper components (e.g. motors capable to sustain heat). In particular, the authors focused their attention on the problem of overheating of the UGV. To address this issue the authors conducted series of experiments on hot metal production sites in order to identify critical environmental parameters of such plants.

In particular, the simulation determines intensity of thermal radiation received by the vehicle, taking into account relative position of the system respect to molten materials and hot surfaces, current phase of casting process, environmental conditions, etc. Indeed, based on the 3D simulation of operations it is possible to obtain information regarding special state of the vehicle and trajectory of its movement $P(x_g,y_g,z_g,\alpha,\beta,\gamma,t)$ in the plant, combine it with actual state of the environment and use all this data to supply the thermal model.
The parameters x, y and z represent position of the UGV, while angles α, β and γ represent its rotational state at moment t. At the same time, exposure time is used to evaluate heating during different tasks, such as sampling, but performed in matter of seconds or cleaning activities which are done without exposure to molten metal , but requires more time (e.g. a couple of minutes).
In addition the geometry of vehicle is used to obtain overall heat absorbed by the UGV. Another important aspect is related to the internal heating due by the internal system, indeed, many components, especially motors of the propulsion subsystem and robotic arms could produce substantial amount of heat during the operation. Considering this, it is possible to list principal factors which contribute to the thermal balance:

- Heating of UGV by thermal irradiation from molten metal and hot surfaces
- Heating of UGV by internal heat generation
- Cooling (or heating) of UGV by convection due to environmental temperature
- Cooling by UGV by internal cooling unit

All this data allows to obtain complete set of information required to perform thermal analysis.
In particular, since there is no direct contact between the hot surfaces and materials and the UGV, the heat exchange is based mostly on convection (1) and thermal radiation (2). Of cause, there is still present heating by contact with ground, however, its impact is relatively small.

$$Q_c = h_c S(T - T_{env}) \quad (1)$$

With:
h_c: Conduction Coefficient [W/(K m^2)]
S: Surface [m^2]
T: Temperature of the surface [K]
T_{env}: Environment temperature [K]

$$Q_{1,2} = \varepsilon_1 * A_1 * \varepsilon_2 * A_2 * F_{1-2} * \theta * (T_1^4 - T_2^4) \quad (2)$$

With:
ε_x: Emissivity coefficient
A_x: Area of the Emitting Body [m^2]
$F_{(1-2)}$: Shape Factor
Θ: Boltzmann constant 5.669 10^{-8} W/(m^{2}K^4)
T_x: Temperature of emitting body [K]

In this case, Modeling and Simulation are essential for correct design. Indeed, such calculation could be performed manually, however, they would neglect important details of positioning and require big time investments. Meanwhile, the model allows to evaluate different configurations of platform and combinations of environmental parameters in the real time.

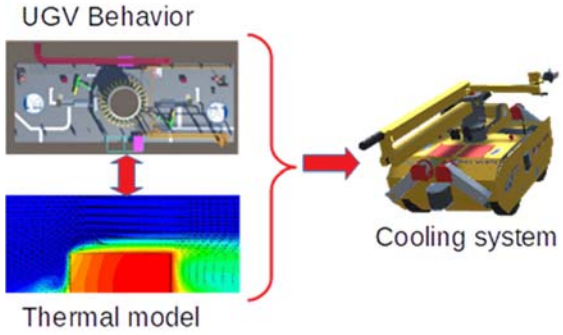


Figure 2: Coupling of 3D simulation and thermal model

In order to obtain precise data about environmental conditions, the authors conducted a series of onsite measurements and experimentations. These activities allowed to measure temperature of surfaces including ground in different points and during different phases of the casting process. This data was obtained from different sources (plant's sensors, pyrometers, thermometers) and summarized in the table below.

Table 1: Temperature measurements for determining boundary conditions.

Description	*Acquisition Mode*	*Temperature*
Hot metal temperature during casting	SCADA System plus Pyrometer	1500C°
External Air Temperature, depends on the location of the plant	Thermometer/ Historical Data Set	-20 C°<T_{env}<45 C°
Temperature of the slag residual. Time dependent due to natural cooling of the hot metal	SCADA System plus Laser Pyrometer	$T_{env} < T_{Slag}(t) < 1500\ C$
Tap hole zone temperature	SCADA System plus Laser Pyrometer	$T_{env} < T_{VHS}$ <400C°
Ground temperature	Laser Pyrometer	$T_{env} < T_{ground} <$ 120C°

6 THERMAL DESIGN AND SIMULATION

As mentioned, simulation is essential for design of autonomous system for operation in harsh environments, such as high temperature in hot metal production; interestingly, similar requirements are applied even to satellites (Humphries and Griggs 1977; Tsai 2004, Faghri 1995; Maydanik 2005, Baturkin 2005).

Indeed, necessity to introduce thermal regulation is present in various fields, which often address such issues in different ways, such as focusing attention on thermal shields, reflective materials and active cooling systems (Fortescue and Stark, 1995; Swanson and Birur 2003, Osiander et.al 2004). In general, a thermal control system can be distinguished into following main categories:
Passive Systems. This category does not require any source of external power supply to influence the internal temperature. Indeed, they utilize special materials and paints, in order to guarantee required emissivity and reflection to protect from external thermal irradiation. Such systems include

- Surface Coating and paints
- Insulation
- Thermal straps and radiators

Contrary to the passive systems, Active Systems require external power to operate, but often several important benefits, such as precise thermal regulation as well as possibility to handle much more intensive heat flows. Typical devices of this type are:

- Heat exchangers
- Heat pumps

In the case of interest, the inside temperature must be maintained in the range between -5…40C°, while the system could be subjected to 1-2 kW heat flow. In such case active cooling system is required. In the following is presented equation of heat balance of the system of interest.

$$\frac{\partial E_{in}}{\partial t} + \frac{\partial E_{out}}{\partial t} + \frac{\partial E_{gen}}{\partial t} = \frac{dE_{Stored}}{dt} \qquad (3)$$

With
E_{in}: Total incoming energy [W]
E_{out}: Total outgoing energy [W]
E_{gen}: Total generated energy [W]
E_{stored}: Total stored energy [W]

This equation is the heart of the thermal model. Indeed, combining equations (1-3) it is possible to take into account such factors as natural convection, heating due by irradiation as well as impact of different active and passive cooling systems.
The 3D virtual simulation has been developed mostly in C# (operational models, industrial processes and procedures) and Unity engine, while thermal model is implemented in Matlab; the models could be integrated by means of a HLA gateway. This approach could allowed to obtain a set of overall Key Performance Indicators in different configurations of the system within distinct boundary conditions. The resulting simulation confirmed a strong correlation between operation type (sampling, cleaning, etc) and heating. At the same time, proper passive heat protection (e.g. reflective material) allows to save power of on-board cooling unit significantly reducing consumptions. Finally, using the digital twin and the virtual experimentation, it was calculated that in order to operate at high temperature for a required periods of time an on-board cooling system with a heat capacity around 500W is necessary.

CONCLUSIONS

The paper addresses important safety issues in the field of industrial plants dealing with hot metal production. In this context, the autonomous systems are a very promising solution and could be develop in optimal way by using innovative simulation models. Indeed, different models have been wrapped into a digital twin able to support engineering of an ad hoc autonomous system for this kinds of plants. Indeed, it was validated the fact that this approach allows to assess the efficiency of alternative proposed configurations and operational models, hence, to perform optimization of the configuration.
Considering this, it is evident that virtual prototyping applied to autonomous vehicles is useful at all stages of development, during the production as well as in the future exploitation.
The severe working conditions in the sector of Iron & Steel are quite typical also for other industries. Indeed, the authors expect to extend the field of application of the proposed solution and methodologies to other field in order to improve safety and efficiency of operations.

REFERENCES

Baturkin, V. (2005). Micro-satellites thermal control — concepts and components. Acta Astronautica, 56(1-2) pp.161-170.

Boschert, S. & Rosen, R. (2016). Digital twin—the simulation aspect. In Mechatronic Futures pp. 59-74. Springer, Cham.

Brimacombe, J. K. (1999). The challenge of quality in continuous casting processes. Metallurgical and Materials Transactions A, 30(8), 1899-1912.

Bruzzone, A. G. (2018). MS2G as Pillar for Developing Strategic Engineering as a New Discipline for Complex Problem Solving. Keynote Speech at I3M, Budapest.

Bruzzone, A.G., Massei, M., Agresta, M., Di Matteo, R., Sinelshchikov, K., Longo, F., Nicoletti, L., Di Donato, L., Tommasini, L., Console, C., Ferraro, A., Pirozzi, M., Puri, D., Vita, L., Cassara, F., Mennuti, C., Augugliaro, G., Delle Site, C., Di Palo, F. & Bragatto, P. (2017). Autonomous systems & safety issues: the roadmap to enable new advances in Industrial Application. In Proceedings of EMSS, edited by Longo F., Affenzeller M.,Piera M.A.,Bruzzone A.G. and Jimenez E., pp. 565-571. Rende, Italy, CAL-TEK S.r.l.

Bruzzone, A.G., Massei, M., Maglione, G.L., Di Matteo, R. & Franzinetti, G. (2016) Simulation of Manned & Autonomous Systems for Critical Infrastructure Protection. In Proceedings of DHSS, edited by Bruzzone A.G., Sottilare R.A., pp. 82-88. Genova, Italy, Dime University of Genoa.

Faghri, A. (1995). Heat pipe science and technology. Global Digital Press.

Fortescue, P. & Stark, J. (1995). Spacecraft System Engineering, third ed., ISBN 0 471 95220 6, Copyright John Wiley, 1995.

Garro, A. & 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 (TMS), SpringSim, edited by Wang M.H.,Barros F.,D'Ambrogio A.,Zacharewicz G. Volume 47, Issue 8, 2015, pp. 9-16. San Diego, CA, USA, The Society for Modeling and Simulation International.

Geerdes, M., Chaigneau, R., & Kurunov, I. (2015). Modern blast furnace ironmaking: an introduction (2015). Ios Press.

Humphries, W. R. & Griggs, E.I. (1977). A design handbook for phase change thermal control and energy storage devices. NASA STI/recon technical report N.78.

Mailliet, P. & Metz, J. (1993). U.S. Patent No. 5,246,208. Washington, DC: U.S. Patent and Trademark Office.

Massei, M., Poggi, S., Agresta, M. & Ferrando, A. (2014). Development planning based on interoperable agent driven simulation. Journal of Computational Science, 5(3), pp. 395-407.

Maydanik, Y. F. (2005). Loop heat pipes. Applied thermal engineering, 25(5-6), pp. 635-657.

Nelson, L.R. & Hundermark, R.J. (2014). The tap-hole – Key to Furnace Performance. Journal of the Southern African Institute of Mining and Metallurgy, Volume 116, Issue 5, May 2016, pp. 465-490.

Osiander, R., Firebaugh, S. L., Champion, J. L., Farrar, D. & Darrin, M.G. (2004). Microelectromechanical devices for satellite thermal control. IEEE Sensors Journal, 4(4), 525-531.

Pardo, N. & Moya, J. A. (2013). Prospective scenarios on energy efficiency and CO2 emissions in the European Iron & Steel industry. Energy, 54, 113-128.

Sarcar, A. S., Ghosh, B., Rao, B. & Swaminathan, K. S. (1982). Repair and Maintenance Practices of Blast Furnaces and Hot Blast Stoves. Transactions of the Indian Ceramic Society, Vol 4 (1) pp. 17-21, Taylor & Francis

Swanson, T. D. & Birur, G.C. (2003). NASA thermal control technologies for robotic spacecraft. Applied thermal engineering, 23(9), 1055-1065.

Shikdar, A. A. & Sawaqed, N. M. (2003). Worker productivity, and occupational health and safety issues in selected industries. Computers & industrial engineering, 45(4), 563-572.

Tsai, J. R. (2004). Overview of satellite thermal analytical model. Journal of spacecraft and rockets, Vol. 41(I), pp.120-125.

Author's Index