16th INTERNATIONAL INDUSTRIAL SIMULATION CONFERENCE 2018

ISC'2018

EDITED BY

José Machado

António Abelha

Luis Mendes Gomes

Hélia Guerra

June 6- 8, 2018

PONTA DELGADA, PORTUGAL

A Publication of EUROSIS-ETI

16th Industrial Simulation Conference 2018

PONTA DELGADA, PORTUGAL

JUNE 6- 8, 2018

Organised by

ETI - The European Technology Institute

Sponsored by

EUROSIS

University of Skövde

University of Žilina

Co-Sponsored by

Ghent University

Hosted by

University of the Azores

Ponta Delgada, Portugal

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITORS

General Conference Chair
José Machado, Uminho, Braga, Portugal

General Program Chairs
Luis Mendes, Gomes UAzores, Ponta Delgada, Portugal
Hélia Guerra, UAzores, Ponta Delgada, Portugal
AntÓnio Abelha, UMinho, Braga, Portugal

Past Conference Chairs
Jan Owsinski, Polish Academy of Sciences, Warsaw, Poland
Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

Publications Chairs
Peter Lawrence, Swinburne University, Australia
Yan Luo, NIST, Gaithersburg, USA
Wan Ahmad Yusmawiza Wan Yusoff, International Islamic Univ. Malaysia, Kuala Lumpur, Malaysia

INTERNATIONAL PROGRAMME COMMITTEE

Discrete Simulation Methodology
Sang Do No, Sungkyunkwan University, South Korea
Vincent Rodin, European Centre for VR, Brest, France

Discrete Simulation Analysis
Pascal Berruet, Université Bretagne Sud, Lorient Cedex, France

Discrete Simulation Methodology, Languages and Tools
Matthias Becker, University Hannover, Hannover, Germany
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria
Sophie Hennequin, ENIM, Metz Cedex, France
Bjorn Johansson, Chalmers University of Technology, Gothenburg, Sweden
Erik Lindskog, Chalmers University of Technology, Gothenburg, Sweden
Stefano Marrone, Seconda Universita di Napoli. Naples, Italy
Jiri Safarik, University of West Bohemia, Plzen, Czech Republic

Open Source Simulation Languages and Tools
Sang Do No, Sungkyunkwan University, South Korea
Philippe Geril, ETI Bvba, Ostend, Belgium

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

Simulation in Computer Science
Lipika Deka, Tezpur University and IIT Guwahati, India
Ernst Kesseler ,NLR, Amsterdam, The Netherlands

Ambient Intelligence and Simulation
Sang Do No Sungkyunkwan University, South Korea
Selwyn Piramuthu, University of Florida, Gainesville, USA

INTERNATIONAL PROGRAMME COMMITTEE

Artificial Intelligence
Chrissanti Angeli, Technological Institute of Piraeus, Athens, Greece
Paulo Jorge Sequeira Gonçalves, Polytechnic Institute of Castelo Branco, Castelo Branco, Portugal
Sang Do No, Sungkyunkwan University, South Korea
Sanaul Haque, Institute of Technology, Carlow, Ireland

VR and Graphical Simulation in Industry
Track Chair: Guodong Shao, NIST, Gaithersburg, USA
Pascal Berruet, Université Bretagne Sud, Lorient Cedex, France
Sanaul Haque, Institute of Technology, Carlow, Ireland
Fabrizio Lamberti, Politecnico di Torino, Turin, Italy
Christoph Laroque, University of Applied Sciences, Zwickau, Germany
Sudhir Mudur, Concordia University, Montréal, Canada
Paulo Moura Oliveira, UTA University, Vila Real, Portugal
Dorin-Mircea Popovici, Universitatea OVIDIUS din Constanta, Constanta, Romania
Marcos A.Rodrigues Sheffield Hallam University, Sheffield, United Kingdom

Augmented Reality and Pervasive Systems in Factories
Alessandro Genco University of Palermo, Palermo, Italy
Sanaul Haque Institute of Technology, Carlow, Ireland
Eleni Mangina University College Dublin (UCD), Dublin, Ireland

Simulation-based evaluation of interactive systems
Sanaul Haque, Institute of Technology, Carlow, Ireland
Stefan Hellfeld, FZI - House of Living Labs, Karlsruhe, Germany
Benjamin Herd, King's College, London, United Kingdom

Simulation and Training
Manuel Alfonseca, Universidad Autonoma de Madrid, Madrid, Spain
Wenji Mao, Chinese Academy of Sciences, Beijing, P.R. China
Eleni Mangina, University College Dublin (UCD), Dublin, Ireland
Dorin-Mircea Popovici, Universitatea OVIDIUS din Constanta, Constanta, Romania

Industrial Internet of Things
Norbert Jastroch, MET Communications GmbH,Bad Homburg, Germany
Sanaul Haque, Institute of Technology, Carlow, Ireland

e-Business in Collaborative Engineering
Heiko Duin, BIBA, Bremen, Germany
Sanaul Haque, Institute of Technology, Carlow, Ireland
Yo-Ping Huang, Tatung University, Taipei, Taiwan
Norbert Jastroch, MET Communications GmbH, Bad Homburg, Germany

Simulation in Nanotechnology
Clemens Heitzinger, Cambridge University, Cambridge, United Kingdom
Yong K. Kim, University of Massachusetts Dartmouth, Dartmouth, USA
Javier Marin. ETSI, University of Malaga, Malaga, Spain

Complex Systems Modelling
Track Chair: Igor N Litvine, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
Frantisek Capkovic, Slovak Academy of Sciences, Bratislava, Slovak Republic
Alexandre Nketsa, LAAS, Toulouse, France
Miguel Rocha, University do Minho, Braga, Portugal
Alfonso Urquia, UNED, Madrid, Spain

Simulation of Complex Multiprocessor Systems
Peter Kvasnica, Alexander Dubcek University of Trencin, Trencin, Slovakia
Orhan Gemikonakli, Middlesex University, London, United Kingdom

Simulation in Electronics, Computer and Telecommunications
Teresa Alvarez, University of Valladolid, Valladolid Spain
Christos Bouras, University of Patras, Patras, Greece
Adnane Latif, Cadi Ayyad University, Marrakech, Morocco
Silvia Mirri, University of Bologna, Bologna, Italy
Maurizio Palesi, Universita di Catania, Catania, Italy

INTERNATIONAL PROGRAMME COMMITTEE

Marco Roccetti, University of Bologna, Bologna, Italy
Fernando Boronat Segui, UPV, Valencia, Spain
Renate Sitte, Griffith University, Gold Coast, Australia
Pedro Sousa, University of Minho, Portugal

Simulation in Aerospace
Reza Azadegan, Urmia University, Urmia, Iran
Wolfgang Kuehn, University of Wuppertal, Wuppertal, Germany

Marine Simulation
Sergeij Kalashnkikow, DANFOSS, Austria

Simulation in Military and Defense
Roberto de Beauclair Seixas, IMPA, Rio de Janeiro, Brazil
Carlos Palau, UPV, Valencia, Spain
Matthias Reuter, CUTEC Gmbh, TU-Clausthal, Clausthal, Germany
Joseph M. Saur, Regent University, USA

Civil and Building Engineering
Alistair Borthwick, Oxford University, Oxford, United Kingdom
Graham Saunders, Loughborough University, Loughborough, United Kingdom

Simulation in Chemical, Petroleum and Mining Engineering
Pouria Homayonifar Acona Flow Technology, Skien, Norway
Mohamad R. Riazi, Kuwait University, Kuwait

Simulation in Energy and Power Systems
Sergeij Kalashnkikow, DANFOSS, Austria
Janos-Sebestyen Janosy, KFKI Atomic Energy Research Institute, Budapest, Hungary

Servitization
Paulo Jorge Sequeira Gonçalves Polytechnic Institute of Castelo Branco, Castelo Branco, Portugal
Sang Do No Sungkyunkwan University, South Korea

Automation/Robotics
Leopoldo Arnesto Angel, UPV, Valencia, Spain
A. Chatzinikolaou, University of Athens, Athens, Greece
Andrzej Dzielinski, Warsaw University of Technology, Warsaw, Poland
Houcine Hassan, Universitat Politècnica de València, Valencia, Spain
Markus Koch, Orga Systems GmbH, Paderborn, Germany
George L. Kovacs, Hungarian Academy of Sciences, Budapest, Hungary
Martin Mellado, UPV, Valencia, Spain
Paulo Moura Oliveira UTA University, Vila Real, Portugal
Bogdan Raducanu, Computer Vision Centre, UAB, Barcelona, Spain
Sergiu-Dan Stan, Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Machine to Machine
Paulo Jorge Sequeira Gonçalves, Polytechnic Institute of Castelo Branco, Castelo Branco, Portugal
Sang Do No, Sungkyunkwan University, South Korea

Intelligent Connected Systems
Paulo Jorge Sequeira Gonçalves Polytechnic Institute of Castelo Branco, Castelo Branco, Portugal
Sang Do No Sungkyunkwan University, South Korea

Cyber Physical Systems Simulation
Track Chair: Ioan Dumitrache, The Romanian Academy, Bucharest, Romania
Adina Magda Florea, University POLITEHNICA of Bucharest, Bucharest, Romania
Pascal Berruet Université Bretagne Sud, Lorient Cedex, France
Cornel Burileanu, University POLITEHNICA of Bucharest, Bucharest, Romania
Corneliu Bilan, University POLITEHNICA of Bucharest, Bucharest, Romania
Eleni Mangina University College Dublin (UCD), Dublin, Ireland

Fog computing for Robotics and Industrial Automation
Paulo Jorge Sequeira Gonçalves, Polytechnic Institute of Castelo Branco, Castelo Branco, Portugal
Sang Do No Sungkyunkwan University, South Korea

INTERNATIONAL PROGRAMME COMMITTEE

Simulation in Engineering Processes
Chrissanti Angeli, Technological Institute of Piraeus, Athens, Greece
Blazej Balasz Technical University of Koszalin, Koszalin, Poland
Alejandra Gomez Padilla, University of Guadalajara, Mexico
Bernhard Heinzl TU Wien, Vienna, Austria
Jan Studzinski, Polish Academy of Sciences, Warsaw, Poland
Joao Tavares, University of Porto, Porto, Portugal
Henk Versteeg, Loughborough University, Loughborough, United Kingdom

Model Driven Engineering
Norbert Jastroch, MET Communications GmbH,Bad Homburg, Germany
Paulo Jorge Sequeira Gonçalves, Polytechnic Institute of Castelo Branco, Castelo Branco, Portugal

Simulation in Industrial Design and Product Design
Chiara Catalano, IMATI-CNR, Genoa, Italy
Yan Luo, NIST, Gaithersburg, USA
Catarina Rizzi, University of Bergamo, Bergamo, Italy

Choice Modelling Methods and Applications
Karim Labadi, ECAM-EPMI, QUARTZ-Lab, France

Simulation in Multibody Systems
Ignacio Garcia-Fernandez, University of Valencia, Valencia, Spain

Organization and Management in Collaborative Engineering
Eduardo Castellano, IK4-IKERLAN, Mondragon-Arrasate, Spain
Florina Covaci, ”Babes-Bolyai” University, Cluj-Napoca, Romania
Iris Graessler, Cologne University of Applied Sciences, Cologne, Germany

Supporting Technologies
Ricardo Gonçalves, Univ. Nova de Lisboa, UNINOVA, Caparica, Portugal
Sarka Kvetonova, Brno University of Technology, Brno, Czech Republic

Formal Methods, Techniques and Design
Yan Luo NIST, Gaithersburg, USA
Stefano Marrone, Seconda Universita di Napoli, Caserta, Italy

Engineering of Embedded Systems
Petr Hanacek, Brno University of Technology, Brno, Czech Republic
Jose Machado, University do Minho, Braga, Portugal
Stefano Marrone, Second University of Naples, Caserta, Italy

Collaborative Engineering data management and information modeling
Uwe Baake, Daimler Buses, Evobus GmbH, Neu-Ulm, Germany
Florina Covaci, ”Babes-Bolyai” University, Cluj-Napoca, Romania
Joachim Herbst, Daimler AG, Research and Technology, Ulm, Germany
Jiri Vondrich, Czech Technical University, Prague, Czech Republic

Collaborative Engineering process management and simulation
Antonio Carvalho, Brito FEUP, University of Porto, Porto, Portugal
Florina Covaci, ”Babes-Bolyai” University, Cluj-Napoca, Romania
Imed Kacem, University of Technology of Troyes, Troyes, France
Carlo Meloni, Politecnico di Bari, Bari, Italy
Roberto Montemanni, IDSIA, Manno, Switzerland
Maurizio Palesi, University of Catania, Catania, Italy
Christian Andreas Schumann, University of Zwickau, Zwickau, Germany

Collaborative environments for virtual teams
Veronique Baudin, LAAS-CNRS, Université de Toulouse, Toulouse Cedex, France
Khalil Drira, LAAS-CNRS, Université de Toulouse, Toulouse Cedex, France
Paul Stratil, SAP Deutschland SE & Co. KG, Walldorf, Germany

Collaborative Enhanced Lean Manufacturing Simulation
Jean-Claude Hochon, AIRBUS, Blagnac, France
Jose Antonio V. Oliveira, University of Minho, Braga, Portugal
Rik Van Landeghem, Ghent University, Zwijnaarde, Belgium

INTERNATIONAL PROGRAMME COMMITTEE

Practical applications and experiences
James Gao, University of Greenwich, Chatham Maritime, Kent, United Kingdom
Sanaul Haque, Institute of Technology, Carlow, Ireland
Irfan Manarvi, Prince Sultan University, Riyadh, Kingdom of Saudi Arabia

Factory Planning and Control for SME's
Sílvio do Carmo Silva, University of Minho, Portugal
Pascal Berruet, Université Bretagne Sud, Lorient Cedex, France
Paulo J.S. Gonçalves, Polytechnic Institute of Castelo Branco, Portugal
Nuno O. Fernandes, Polytechnic Institute of Castelo Branco, Portugal
Veronique Limere, Ghent University, Ghent, Belgium

Data-Driven Tradespace Exploration and Analysis
Philippe Geril, ETI, Ostend, Belgium
Sanaul Haque, Institute of Technology, Carlow, Ireland
Peter K. K. Loh, Nanyang Technological University, Singapore

Smart Factory 1.0 and Industry 4.0
Florina Covaci, "Babes-Bolyai" University, Cluj-Napoca, Romania
Norbert Jastroch, MET Communications GmbH,Bad Homburg, Germany
Philippe Geril, ETI, Ostend, Belgium

Simulation in Steel Manufacturing and New Materials Manufacturing
Brian Hollocks, Bournemouth University, Bournemouth, United Kingdom

Simulation in Electronics Manufacturing
Gerald Weigert, Dresden University of Technology, Germany

Simulation-based Optimization in Industry
Bernhard Heinzl TU Wien, Vienna, Austria
Karim Labadi ECAM-EPMI, QUARTZ-Lab, France
Amos H.C. Ng, University of Skovde, Skovde, Sweden
Anna Syberfeldt, University of Skovde, Skovde, Sweden

Apparel and Textile Simulation
Jocelyn Bellemare, University of Quebec in Montreal (UQAM), Montréal (Québec) Canada

Simulation in Logistics, Transport and Harbour Simulation
El-Houssaine Aghezzaf Ghent University, Ghent, Belgium
Maria Sameiro Carvalho, University of Minho, Guimaraes, Portugal
Remy Dupas, Université de Bordeaux , Bordeaux, France
Olivier Grunder , UTBM, Belfort, France
Isabel Garcia Guttierez, Univ. Carlos III de Madrid, Madrid, Spain .
Houcine Hassan Universitat Politècnica de València, Valencia, Spain
Dmitry Ivanov, Chemnitz University of Technology, Chemnitz, Germany
Peter Lawrence, Swinburne University, Lilydale, Australia
Marie-Ange Manier, UTBM, Belfort, France
Roberto Montemanni, IDSIA, Manno-Lugano, Switzerland
Rosaldo Rossetti, University of Porto, Porto, Portugal
Hans Veeke, TU Delft, Delft, The Netherlands
Pengjun Zheng, University of Southampton, Southampton, United Kingdom

Simulation in Supply Chains
Florina Covaci, "Babes-Bolyai" University, Cluj-Napoca, Romania
Matthias Klumpp, ild Institute for Logistics and Service Management at FOM, Essen, Germany
Eleni Mangina, University College Dublin (UCD), Dublin, Ireland
Sang Do No, Sungkyunkwan University, South Korea

INTERNATIONAL PROGRAMME COMMITTEE

Hospital Logistics Simulation
Track Chair: Giorgio Romanin-Jacur, University of Padova, Vicenza, Italy
Antonio Abelha, Universidade do Minho, Braga, Portugal
Jose Machado, University of Minho, Braga, Portugal
Jose Antonio V. Oliveira, University of Minho, Braga, Portugal
Peter Summons, University of Newcastle, Australia

Transport Simulation
Jose Antonio V. Oliveira University of Minho, Braga, Portugal

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

Urban Simulation
Hans Degrauwe, Degrauwe Consulting, Belgium

Statistical Analysis and Data Mining of Business Processes
Paulo Alexandre Ribeiro Cortez, University of Minho, Guimaraes, Portugal
Sanaul Haque, Institute of Technology, Carlow, Ireland
Selwyn Piramithu, University of Florida, Gainesville, USA
Inna Pivkina, New Mexico State University, USA
Danny Van Welden, KBC Bank, Brussels, Belgium

Applications using Data Mining
Marco Furini, Università di Modena e Reggio Emilia, Reggio Emilia, Italy
Sanaul Haque, Institute of Technology, Carlow, Ireland

Applications using data mining and open source text information mining and analysis
Marco Furini, Università di Modena e Reggio Emilia, Reggio Emilia, Italy
Sang Do No, Sungkyunkwan University, South Korea

Patient Centered Health Care with Meaningful Use of IT
Sanaul Haque, Institute of Technology, Carlow, Ireland
Gurdev Singh, State University of New York at Buffalo, Buffalo NY, USA

Geosimulation and Big Data
Marwan Al-Akaidi, Arab Open University, Kuwait

Simulation Applied to Economics
J. Manuel Feliz Teixeiro, FEUP-University of Porto, Porto, Portugal
Nuno Melo, Instituto Politécnico de Viseu, Viseu, Portugal
Leonid Novickis, IDC Information Technologies, Riga, Latvia
Kamelia Stefanova, University of National and World Economy, Sofia, Bulgaria

Agents in Business Automation and Economics
Florina Covaci, ”Babes-Bolyai” University, Cluj-Napoca, Romania
Martin Fredriksson, Blekinge institute of technology, Karlskrona, Sweden
Markus Koch, Orga Systems GmbH, Paderborn, Germany
Isabel Praca, Polytechnic of Porto, Porto, Portugal

Simulation in Business and Economics
Eduardo Castellano, IK4-IKERLAN, Mondragon-Arrasate, Spain
Sudhir Mudur, Concordia University, Montréal, Canada
Yan Luo NIST, Gaitherburg, USA

Simulation in Business Games
Marco Furini, Università di Modena e Reggio Emilia, Reggio Emilia, Italy
Renate Sitte, Griffith University, Gold Coast, Australia
David Wortley, Coventry University, Coventry, United Kingdom

INTERNATIONAL PROGRAMME COMMITTEE

Simulation and Analysis of Environmental effects of business
Dietmar Boenke, Hochschule Reutlingen, Reutlingen, Germany
Elias A. Hadzilias, Universite Catholique de Lille, Lille, France
Roberto Razzoli, University of Genova, Genova, Italy

Simulation Tools for Business Management & Business Intelligence
Carlo Meloni, Politecnico di Bari, Bari, Italy

Simulation in OR and Knowledge Management
Chrissanthi Angeli, Technological Institute of Piraeus, Athens, Greece
Imed Kacem, Université Paul Verlaine Metz, Metz, France
Christophe Roche, Universite de Savoie, Le Bourget du Lac Cedex, France

Emergency Management and Risk Analysis Management
Ismail Bouassida, LAAS-CNRS, Université de Toulouse, Toulouse, France
Khalil Drira, LAAS-CNRS, Université de Toulouse, Toulouse, France
Konstantinos Kirytopoulos, University of the Aegean, Greece
Gurdev Singh, UB Patient Safety Research Center, State University of New York at Buffalo, Buffalo NY USA
Lode Vermeersch, Delcredere, Brussels, Belgium

Policy Modelling and Simulation in E-Management, E-Government, E-Commerce and E-Trade
Vrassidas Leopoulos, Nat. technical University of Athens, Athens, Greece

Political and Government Simulation
Joseph M. Saur, Regent University, USA

Sustainable Development Management
Philippe Geril, ETI Bvba, Ostend, Belgium

HORIZON 2020 Workshop
Valerian Croitorescu, University of Bucharest, Bucharest, Romania

Simulation in Automotive Systems
Track Chair: José Machado, Uminho, Braga, Portugal
Antonio Abelha, UMinho, Braga,Portugal
Naoufel Cheikhrouhou, EPFL, Lausanne, Switzerland
Valerian Croitorescu, University of Bucharest, Bucharest, Romania
José António Oliveira, University of Minho in Portugal
Julien Richert, Daimler AG, GR/PAA, Sindelfingen, Böblingen, Germany

INDUSTRIAL SIMULATION 2018

EUROSIS-ETI Publications are ISI-Thomson, IET, SCOPUS and Elsevier Engineering Village referenced
Legal Repository: Koninklijke Bibliotheek van België, Keizerslaan 4, 1000 Brussels, Belgium
CIP 12.620 D/2011/12.620/1

Selected papers of this conference are published in scientific journals.

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

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

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

EUROSIS-ETI Publication
ISBN: 978-94-92859-03-7
EAN: **978-94-92859-03-7**

Preface

The aim of the Industrial Simulation Conference (ISC) is to be the annual international industrial simulation conference based in Europe, which aims at giving a complete overview of industrial simulation related research and at providing an annual status report on present day industrial simulation research within the European Community and the rest of the world. With the integration of artificial intelligence, agents and other modelling techniques, simulation has become an effective and appropriate decision support tool in industry. The exchange of techniques and ideas among universities and industry, which support the integration of simulation in the everyday workplace, is the basic premise at the heart of the ISC conference.

This volume brings together various methodologies and theories, dealing with a complete set of scientific studies and research teams. The selected papers are contributions such as research papers, case studies and demonstrations that present original scientific results, methodological aspects, concepts and approaches in the multidisciplinary field of industrial simulation. Applications were largely encouraged. Several topics such as Simulation Optimization, Prediction Data Management, Agricultural Data Management, Collaborative Systems Management, Trajectory Optimization and Inventory Tracking were welcome.

Important issues are also addressed in the area of integration of optimization techniques into simulation analysis. The use of multilevel models is a major challenge. Management, planning and decision are based increasingly on knowledge and computer-based simulation provides information about the behavior of systems in the context of multi-scale or multi-stage optimization. The collections of data that are available today (e.g. from monitoring devices) allow a more efficient support for these activities. However, their size, nature and complexity require intelligent data structures in order to create storage, processing and analytical discovery of more effective ways of knowledge as to its diffusion technologies. Data mining and data warehousing fields can contribute significantly for empowering and automate the process of decision-making. Important contributions can be made in critical areas. Real time and online intelligent decision support systems are of most importance to supply people with important information in useful time and data quality assessment is fundamental for detecting critical events.

ISC 2018 is organized in the Portuguese autonomous region of Azores. Azores is an archipelago composed of nine volcanic islands in the North Atlantic Ocean. The ISC 2018 is held at the University of the Azores, on the island of São Miguel, which is the main settlement of the Azores. Researchers are warmly welcomed to the Azores and Ponta Delgada, where they can enjoy the natural beauties of the region, undoubtedly one of the most beautiful in the world.

We thank all the authors and EUROSIS.

José Machado
António Abelha
Luis Mendes Gome
Hélia Guerra

The Editors

CONTENTS

CONTENTS

TRAJECTORY OPTIMIZATION

INVENTORY TRACKING

SCIENTIFIC PROGRAMME

SIMULATION OPTIMIZATION

APPLICATION OF FORECASTING AND CONTROL OF ACTIVATED VOLUME INDEX TO IMPROVE THE OPERATION OF SEWAGE TREATMENT PLANT

Bartosz Szeląg[1], Jan Studziński[2]
[1] Kielce University of Technology, Faculty of Environmental, Geomatic and Energy Engineering
Al. Tysiąclecia Państwa Polskiego 7, 25 – 314 Kielce, Poland
E-mail: bszelag@tu.kielce.pl
[2] Polish Academy of Sciences, Systems Research Institute (IBS PAN)
Newelska 6, 01-447 Warsaw, Poland
E-mail: studzins@ibspan.waw.pl

KEYWORDS
Support Vector Machines, Mixed liquor suspended solid, Wastewater treatment, Substrate load, Classification Trees.

ABSTRACT
In the paper a method for prediction and control of activated sludge sedimentation events in wastewater treatment plants is proposed. To simulate the sludge settle ability logit regression models have been applied in which operational parameters of the plant biologic reactor as well as quality indicators of wastewater inflow to the object were used. A logit model is an explicit regression relations that allows to assess clearly an impact of particular process variables on the sludge sedimentation. Because of that such the models can be used efficiently by technologists to control the whole wastewater treatment process. In view of the high costs of measuring the quality indicators of the wastewater inflow in the paper an option of their predicting by means of neural network models in form of multilayer perceptron and cascade networks has been considered. The calculation results confirmed the usefulness of the approach for the neural models obtained showed very satisfied prediction abilities regarding those indicators. Also the logit regression models are effective by simulation and control of sludge settle ability and they can be used successfully to improve operating efficiency of wastewater treatment plants.

INTRODUCTION
An efficient method of communal sewage clearing is system of biological treatment by means of activated sludge. Achievement of targeted level of pollutants reduction is then conditioned not only by correct designing of treatment technological process and its correct technical realization but also by suitable treatment plant operation for which an exact knowledge of operational parameters of Activated Sludge Chambers (ASC) of a Sewage Treatment Plant (STP) is needed. One of these parameters is Sludge Volume Index (SVI) that has got critical impact on the operational effectivity of STP. SVI index makes possible a resultful prediction of unstable states of the treatment process and these are emerging when the filamentous bacteria in the sludge are breeding exceedingly. That occurance leads to so called sludge swelling and is signaled by rising of SVI value above 150 ml/g (Martins et al. 2004, Cortés et al. 2006). The presence of filamentous bacteria in biocenosis of the activated sludge generates problems especially by the sewage clearing occurring in the Secondary Clarifiers (SC) of STP. Appearing of the sludge swelling inhibits the process of sludge separation from the sewage that terminates the whole process of sewage treatment. Consequently some overshooting of permissible concentrations of pollutant indicators like BOD_5 (Biological Oxygen Demand), COD (Chemical Oxygen Demand), total suspended solids or total phosphorus can appear in the sewage leaving the plant as well as problems by sludge condensing and dehydration can occur by the following process of excessive sludge processing. To improve efficiency of plant operation and above all to improve the sludge sedimentation in SC the operational parameters of bioreactor are changed by correction of the oxygen concentration in nitrification chambers, the recirculation ratio, sludge concentration, the sludge load determined by the pollutants load as well as sometimes by the correction of carbon amount dosed externally. If the corrections mentioned above do not cause a satisfied improvement of sludge sedimentation ability then coagulants in form of iron or aluminum compounds are put in bioreactors.
Currently in the majority of plants the relevant decisions are undertaken by the plant operator intuitively in the view of his experience and his observations of the treatment process running. Such process conducting is wrong from neither economical nor ecological point of view for it can cause high costs of sewage treatment and in parallel can require the application of complicated methods of sludge disposal. Such acting does not correspond also with the existent trends of plant operation when efforts are done to make use of large sets of measurements data that are usually collected in STP; with these data mathematical models of treatment processes occurring in STP objects can be developed and used to support in a formalized way the plant operation. From an overview of the available literature concerning that problematics (e.g. Capodaglio et al. 1991, Han et al. 2013, Bagheri et al., 2015, Szeląg & Gawdzik 2016 % 2017, Szeląg & Studziński 2017) results that for modeling the activated sludge settle ability commonly neuronal networks (hierarchic and probabilistic), fuzzy sets (Alsina et al. 2009) or hybrid models being connection between classification and regression model are used. Unfortunately in these works only quantitative compatibility of the measurements with calculation results has been verified but a qualitative influence of different sewage variable (like amount and quality of

sewage inflow or bioreactor parameters) on sludge settle ability has not been examined while these relations are of cardinal importance for the STP operator by his decisions making. Furthermore by developing their models the authors quoted applied numerically complex algorithms what restricts their use in daily plants exploitation. By the models calculation also the problems were overlooked that concern an access to data describing the sewage quality which are very often incorrect and unreliable due to breakdowns of the related measuring probes or to their depreciation.

The majority of models developed are of classical regression type but sometimes also logit regression models are calculated while doing STP mathematical modelling. Then those models are not used to forecast SVI index but to calculate the probability of excessing the accessible index value equal to 150 cm^3/g. Even though a logit model forms a simple and explicit regression relation between related variables, the models presented in the literature (Szeląg 2016, Szeląg & Bąk 2016, Bayo et al. 2006, Bezak-Mazur et al. 2016, Bartkiewicz et al. 2016) did not make possible neither to control parameter values of the bioreactor nor to monitor on-line the sludge sedimentation features. A control and sustained correction of bioreactor operation in order to maintain the SVI value under 150 cm^3/g is a complex and costly task for it requires to measure continuously sewage quality indicators on the sewage inflow to STP as well as the parameters inside of bioreactor. Currently the parameters of ASC are monitored on-line while the sewage quality measurements are collected periodically according to the obligatory regulations. Such procedures lead to the development of models that are not in state to control the bioreactor parameters with an appropriate long time delay. At the same time a relatively long time of determination of variables influencing SVI index, like e.g. BOD_5, complicates an on-line control of ASC parameters.

Therefore it is necessary to elaborate models for prediction and control of sludge settle ability that could consider technical problems connected with the measurement and determination of sewage quality indicators what means the troubles with their collection. Such models shall create a possibility to control and correct continuously the parameters concerning bioreactor operation and dosage of chemical compounds into the sewage considering the fact that the quality of sewage inflowing the plant is changing in very broad range.

In the paper the models of logit regression have been developed to predict the process of activated sludge sedimentation; they are described with quite simple mathematical relations and can be used easily to undertake technological decisions by the plant operator.

OBJECT OF INVESTIGATION

The modeling object is the communal Sewage Treatment Plant of the nominal capacity of 72.000 m^3/d that is located in the community Sitkowka-Nowiny and into which the wastewaters from Kielce city, from the community Sitkowka-Nowiny and partially from the community Masłow inflow. The mixed wastewater is pretreated mechanically in the grid chambers and in aerated sand traps with integrated fat separators and then it is directed to preliminary clarifiers from which it flows into the biological part of the plant (Fig. 1). The biologic reactors there consist of separated chambers of dephosphatation, denitrification and of nitrification where organic pollutants as well as nitrogen and phosphor compounds are removed. After the biologic sewage treatment is performed the wastewater flows to secondary clarifiers where the excessive sludge is set apart and the sewage clarified flows to the end receiver which is Bobrza river.

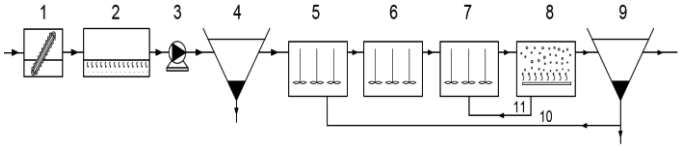


Figure 1. Diagram of the assembly line of a STP: 1 – grids chamber, 2 – sand trap, 3 – pumps room, 4 – primary clarifier, 5 – pre-denitrification chamber, 6 – denitrification chamber, 7 – dephosphatation chamber, 8 – nitrification chamber, 9 – secondary clarifier, 10 – external recirculation, 11 – internal recirculation.

In the summer time phosphor is removed from the wastewater only by biological method without chemical support of the process. Whereas in the winter time the phosphor removal from the sewage is chemically supported by dosing coagulant in form of iron compounds what is caused by decrease of air temperature and succeeding slowdown of biologic reactions in the wastewater. Likewise in that time the problems with the sludge sedimentation occur: the SVI exceeds the value of 150 cm^3/g what points out on the process of activated sludge swelling. Addition of coagulant to the sewage in the summer is intended also by desire to improve the sludge dropping ability and especially in wet periods when some disturbances in operation of activated sludge chambers can appear. In those time periods the operational parameters of ASC shall be amended appropriately and oftentimes to avoid the sludge swelling. In the light of above remarks one can see that there is an essential need to develop some reliable mathematical models with which the STP operator would be able to monitor exactly and correct effectively sludge settle ability.

METHODOLOGY

At the first step of investigation a mathematical model for prediction of sludge settle ability features has been calculated by means of logit regression method. To calculate the model measurements data concerning the raw sewage inflow (Q), sewage quality indicators (biochemical and chemical oxygen demands (BOD_5 and (COD), ammoniacal nitrogen ($N\text{-}NH_4$), total nitrogen (TN), total phosphor (TP)), total suspended solids (TSS) parameters of activated sludge chambers (pH, sludge temperature (T_{sl}), oxygen concentration (DO), sludge concentration (MLSS), sludge age (WO), food to mass ratio (F/M)), as well as amount of PIX coagulant dosed (m_{PIX}) and amount of excessive sludge removed from STP (WAS) have been used. The data come from the time period of four years 2012–2017 and they refer to 200 measurement sessions. As a criterion of evaluation of influence of sewage inflow quality and of parameters of bioreactor on the quality of its operation was the SVI limit value, i.e. SVI_{lim}=150 cm^3/g (Bayo et al. 2006). When SVI value measured was bigger than SVI_{lim} then an established linguistic variable took the value 1 and otherwise it was 0. In order to select appropriate explanatory variables as inputs of the model assuring its satisfied

prediction ability, the matching norms have been applied which are used in classification models.

In view of stability and rightness of sewage treatment process realized by means of activated sludge method, in the model investigated the parameters describing the food to mass ratio ($F/M=Q\cdot BOD_5\cdot MLSS^{-1}\cdot V_k^{-1}$, where V_k – combined volume of aeration chambers) and sludge age ($WO=WAS\cdot MLSS^{-1}\cdot V_k^{-1}$) have been concerned, whereas the food to mass ratio limits the sewage clearing ability and the sludge age describes the time period in which micro-organisms from the sludge reside in the bioreactor. By the short sludge age some nitrification problems can appear in the biomass while sludge age overly long can lead to sewage decay what will essentially worsen the quality of clarified sewage on the outlet form STP. Therefore there is necessary to take the parameters mentioned into consideration by development the respective models and also because the variables appearing in the following relation (1) can determine importantly the F/M and WO values.

On account of the fact that sewage quality measurements on the inflow to STP are gathered only temporarily and not every day and they values influence substantially the sludge sedimentation features (Lou & Zhao 2012, Han et al. 2013, Alsina et al. 2009), so to produce persistent forecasts of SVI index, especial models to simulate sewage quality indicators have been prepared. To do it measurements data concerning the sewage inflow intensity and the sewage inflow temperature have been used (Szeląg & Studziński 2017). Such approach makes possible to produce continuously predictions of sludge sedimentation features even if measurements concerning sewage quality indicators will be temporarily missing. To identify variables describing quality indicators the method of classification trees has been used while for forecasting their values ANN methods of multilayer perceptron (MLP) and cascade network (CNN) types were applied. The model predicting selected sewage quality indicators for which the smallest errors of matching the measurements with the prediction values have been got was then used to calculate sewage quality on the inflow to STP. That calculation was done based on the measurements of inflow rate (Q) and temperature (T) from the time period of 8 months (from January till August). In the time period from March till August the total rainfall was equal to 552 mm and the number of rainy days amounted to 50. Then the results of calculation of the selected sewage quality indicators and the measured values of parameters of the biological reactor were substituted into the logit model and the probability of overshooting the limit value of SVI index (i.e. of SVI_{lim}) has been determined. Doing it also the current SVI value was calculated and if it was bigger than SVI_{lim} then attemps to improve the effectivity of STP operation by elimination of sludge swelling occurrence in the relevant time period were undertaken; those improvement approaches relied on limitation of chemical reagent amounts added to the sewage.

The logit regression method called also as binomial logit modeling is commonly used by the analyses of data of either/or kind as well as by the calculation of probability of events occurrence or no occurrence. The logit models are widely applied in economic and medical sciences and while solving numerous problems of environmental engineering (Szeląg & Bąk 2016, Szeląg 2017, Łomotowski & Dańczuk 2012, Heyer & Stamm 2013, Bezak–Mazur et al. 2016, Bartkiewicz et al. 2016).

A logit regression model is described by the following relation:

$$p(X) = \frac{\exp(X)}{1+\exp(X)} \quad (1)$$

where: p – probability of excessing the limit value SVI_{lim} for which p=0,50, X – vector being a linear combination of explanatory variables described by the formula:

$$X = \sum_{i=1}^{j} \beta_i \cdot x_i + \beta_0 \quad (2)$$

with: x_i – variables denoting SVI which are among others: activated sludge concentration, temperature, pH value of the sludge, amount of coagulant PIX dosed, amounts of ammoniacal and total nitrogen and of organic substances in the sewage inflow to STP, β_0, β_i – coefficients determined by means of the maximum likelihood method.

To assess prediction abilities of models calculated the coefficients of correlation of McFadden, Neglekerke and Cox-Snell, i.e. $R^2_{MCFadden}$, $R^2_{Neglekerke}$ and $R^2_{Cox\text{-}Snell}$ respectively, as well as matching measures like specificity (SPEC), sensitivity (SENS) and reckoning error (Harell 2001) have been used; the latter coefficients are commonly used by the assessment of classification models.

In the paper to model the indicators of sewage quality the neuronal networks of the types of multilayers perceptron (MLP) and of cascade networks (CNN) having additional connections between the weighting factors laying on the input layer and on all other network layers. The optimal structure of neuronal net modelled was determined every time while for the number of neurons admitted on the hidden layer and for the taken activation functions the matching errors MAE (mean absolute error) and MAPE (mean absolute percentage error) have been calculated and compared; the neurons number changed from 3 up to 30 and as the activation function could be taken the following ones: linear (lin), exponential (exp), sinusoidal (sin), sigmoidal (sigm) and hyperbolic tangent (tanh). In the ouput layer linear activation function has been applied. A model structue assumed was recognized as optimal when the matching measures calculated had the minimal values among all models considered. Based on the experiences of (Szeląg 2017, Capizzi et al. 2015), Al-Batah et al. 2015) three connections between the weighting factors on the input layer and the following layers of CNN network were determined.

To estimate the weighting factors in neuronal nets on their teaching stage the method of Broyden–Fletcher–Goldfarb–Shanno (Rutkowski 2006) was applied. To determine the parameter values in logit models and to develop MLP and CNN models the MATLAB software has been used.

RESULTS

Based on the measurement data gained in the sewage treatment plant their variability intervals have been calculated (Tab. 1). The values of daily sewage inflows and of sewage quality indicators have been changed in broad ranges what influenced essentially the parameters of activated sludge chambers. The sludge concentration changed from 1,19 up to 5,89 kg/m^3, the oxygen concentration in the nitrification chamber changed from 0,55 up to 5,78 mg/dm^3 and the mean

value of DO was equal to 2,56 mg/dm^3. In the time period concerned the sludge sedimentation ability was worsening as evidenced by the mean value of SVI which equals to 186 ml/g, as well as by SVI maximal value equal to 320 cm^3/g. Because of that the treatment plant investigated can be a good example of a object to perform a test that would show in what extent an implementation of a model for predicting the sludge sedimentation could improve the effectivity of plant operation. By means of the measurements data concerning the amount and quality of the sewage inflow to STP and the bioreactor parameters and while using the MATLAB software a logit regression model has been developed to predict the SVI values; in the modeling the vector being a linear combination of the variables considered is described by the formula:

$$X = \beta_1 \cdot \frac{BOD_5}{TN} + \beta_2 \cdot \frac{BOD_5}{TP} + \beta_3 \cdot L_{N-NH_4} + \beta_4 \cdot MLSS + \\ + \beta_5 \cdot T_{sl} + \beta_6 \cdot m_{PIX} + \beta_7 \cdot DO + \beta_0 \quad (3)$$

The calculated values of β_i coefficients in equation (3) and of parameters decided of fitting the modelling results with the measurements data are listed in Table 2. The logit regression model obtained shows a satisfied prediction ability. This is evidenced by the calculated values of SPEC=0,900 and of SENS=0,867; the SPEC value determined means that in case of 80 events for which SVI>SVI$_{lim}$, 72 of them were classified rightly, and the SENS value determined means that in case of 110 events for which SVI<150 cm^3/g, 95 of them were identified properly.

Table 1. Variability intervals of parameters describing the sewage amount and quality and of operational parameters of activated sludge chambers.

Variable	Minimum	Mean	Maximum
Q, m^3/d	32564	40698	86592
T_{sl}, ^{O}C	10.0	15.9	23.0
pH	7.2	7.7	7.8
MLSS, kg/m^3	1.98	4.26	6.59
PIX, m^3/d	0.00	0.80	1.93
methanol, m^3/d	0.00	1.35	4.56
F/M, gBOD$_5$/g MLSS·d	0.03	0.07	0.15
WO, d	10.0	16.25	22.35
DO, mg/dm^3	0.55	2.56	5.78
SVI, cm^3/g	95	186	320
BOD$_5$, mg/dm^3	127	309	557
TSS, mg/dm^3	126	329	572
NH_4^+-N, mg/dm^3	24.4	49.4	65.9
TN, mg/dm^3	39.9	77.7	124.1
NO_2, mg/dm^3	0.14	0.002	1.55
NO_3, mg/dm^3	0.17	0.05	1.20
TP, mg/dm^3	4.30	7.80	12.6

From the above analysis results that the model obtained has classified correctly 167 events from 190 cases investigated (88%). The subsequently performed model validation shown that for 10 events considered in 8 cases the events qualification was right (80%). This lets to find that our model has got better classification and prediction features than the models developed by Belanche et al. (2000) or by Bayo et al. (2006), who in their models used as explanatory variables the bioreactor operational parameters (RAS), bacterial microflora parameters, the temperature and pH measurements in the activated sludge chambers. Moreover it should be noted that the functional relationships calculated p=f(x_i) have been confirmed by investigations carried out by Luo and Zhao (2012) and by Flores–Alsina et al. (2009) in communal sewage treatment plants.

Table 2. Calculated values of coefficients β_i and of parameters matching the measurements and modelling results.

Variable	β_i	Standard deviation (σ)
BOD$_5$/TN	0.019	0.003
BOD/TP	0.32	0.032
L_{N-NH4}	0.0009	0.00005
Tsl	-0.368	0.233
MLSS	-1.365	0.39
DO	-1.183	0.272
PIX	-1.756	0.239
Constant term	-14.385	1.25
$R^2_{Mc\ Fadden}$	0.756	SPEC = 0.900
$R^2_{Negelekerke}$	0.821	SENS = 0.867
$R^2_{Cox\ -\ Snell}$	0.756	R_z^2 = 0.880

Based on the results obtained the odds ratios (OR) have been calculated and listed in Table 3. Those indicators show that the quality of sewage inflowing STP as well as operational parameters of activated sludge chambers have got an essential impact on the probability of overshooting the limit value of SVI. For example, if the amount of ammoniacal nitrogen will be increased by 100 kg/d then the probability of overshooting SVI limit value will be grown by 9%. On the second part, an increase of oxygen concentration in the nitrification chamber by 0,1 mg/dm^3 will lead to the decrease of SVI limit value overshooting by 11%.

Table 3. Odds ratios (OR) calculated for respective variables.

ΔMLSS	OR	ΔT_{sl}	OR	ΔPIX	OR	ΔDO	OR
kg/m^3	-	^{O}C	-	m^3/d	-	mg/dm^3	-
0.25	0.71	0.5	0.82	0.25	0.64	0.1	0.89
0.50	0.51	1.0	0.68	0.50	0.41	0.2	0.79
0.75	0.36	1.5	0.56	0.75	0.27	0.3	0.70
1.00	0.26	2.0	0.46	1.00	0.17	0.4	0.62
ΔL_{N-NH4}	OR	BOD$_5$/TN	OR	BOD$_5$/TP	OR		
kg/d	-	-	-	-	-		
100	1.09	2.5	1.04	1.0	1.38		
150	1.14	5	1.09	1.1	1.42		
200	1.20	7.5	1.14	1.2	1.47		
250	1.25	10	1.19	1.3	1.52		

Because in the calculated logit regression model both the values of quotients BOD_5/TN and BOD_5/TP and of N-NH_4 parameter influence the sludge volume index, then statistical models to predict selected sewage quality indicators have been determined. By means of the method of classification trees the explanatory variables for modelling the sewage quality indicators (BOD_5, TN, TP i N-NH_4) and the sewage inflow (Q) have been determined; that determination was done by using the factors of predictors importance (IMP) (Table 4). In the following analyses only the variables have been accounted for which IMP values were bigger than 0,9 (Verma et al. 2013).

Based on the results obtained one can state that to predict the intensity of sewage inflow to STP only the values of measurements Q(t-1) up to Q(t-3) can be used for the IMP values calculated for them are not less than 0,90. From Table 4 results also that both the amount and temperature of the sewage inflow have got an influence on the values of sewage quality indicators (BOD_5, TN, N-NH_4^+). Those conclusions can be performed by the calculation results received for communal sewage treatment plants by Lubos et al. 2017 and by Szeląg 2016 and 2017. Similar confirmations refer also to the delay values regarding the variables Q and T_{in}.

Table 4. Predictors importance IMP calculated for variables explaining selected indicators of sewage quality.

Variable	IMP				Variable (Q)	IMP
	BOD_5	TN	TP	N-NH_4		
Q(t)	1.00	1.00	1.00	1.00	Q(t-1)	1.00
Q(t-1)	0.95	0.93	0.93	0.97	Q(t-2)	0.99
Q(t-2)	0.78	0.91	0.92	0.92	Q(t-3)	0.93
Q(t-3)	0.74	0.82	0.9	0.85	Q(t-4)	0.88
Q(t-4)	0.87	0.62	0.9	0.9	Q(t-5)	0.85
Q(t-5)	0.92	0.6	0.9	0.9	Q(t-6)	0.76
Q(t-6)	0.9	0.9	0.75	0.8	Q(t-7)	0.73
Q(t-7)	0.75	0.88	0.79	0.72	T_{in}(t-5)	0.28
T_{in}(t)	0.9	0.91	0.9	0.8	T_{in}(t-6)	0.26
T_{in}(t-1)	0.9	0.92	0.9	0.92	T_{in}(t-7)	0.25
T_{in}(t-2)	0.67	0.9	0.9	0.91	T_{in}(t-2)	0.24
T_{in}(t-3)	0.92	0.9	0.91	0.9	T_{in}(t-3)	0.23
T_{in}(t-4)	0.7	0.61	0.85	0.84	T_{in}(t-4)	0.22
T_{in}(t-5)	0.9	0.6	0.82	0.76	T_{in}(t-1)	0.21
T_{in}(t-6)	0.8	0.68	0.76	0.73	T_{in}(t)	0.2
T_{in}(t-7)	0.9	0.57	0.9	0.72		

Based on the above calculation results in the following statistical models to predict the sewage quality indicators by using the CNN and MLP methods have been elaborated.

In Table 5 the parameters showing the exactness of fitting of prediction results to the measurements and the variables describing the structures of the models developed are listed. From the Table goes out that the neurons number changed in the models calculated from 5 up to 7 for CNN method and from 8 up to 9 for MLP method and that efficiency of hyperbolic tangent, exponential or sigmoidal functions applied as activation functions on the hidden layers of the networks depended on the indicator type modelled. A following conclusion is that the smallest errors by prediction of the sewage quality indicators have been got while using CNN and not MLP method.

The results obtained in the paper are confirmed by Abyaneh (2014), Lubos (2017) and Szeląg (2016) who showed in their research that the sewage inflow Q has got an essential impact on the sewage quality. The correlation value R calculated by CNN method while modelling variable BOD_5 (R=0.89) is bigger than R value determined by Abyaneh (2014) while using an ANN method in form of multilayer perceptron (R=0.83); to predict that indicator Abyaneh applied as the predictors the temperature and pH values of the sewage inflow and TSS concentration. A better matching of the measurements with the simulation results concerning BOD_5 as Abyaneh has been got also by Dogan et al. (2008) who used in their modeling as the input variables the measurements data of TN, TP, TSS and Q.

Table 5. Fitting of modelling results to the measurements of sewage quality indicators obtained by CNN and MLP methods; R – coefficient of Pearson correlation.

Indicator	CNN				
	MAE mg/dm^3	MAPE %	R	Neurons number	Activation function
BOD_5	32.35	10.21	0.89	7	tanh
TP	0.79	10.50	0.83	5	exp
TN	4.66	5.46	0.85	6	tanh
N-NH_4	2.74	5.19	0.87	6	tanh
Indicator	MLP				
	MAE mg/dm^3	MAPE %	R	Neurons number	Activation function
BOD_5	49.2	18.6	0.69	8	exp
TP	1.15	16.5	0.38	9	sigm
TN	6.25	8.22	0.59	8	tanh
N-NH_4	4.08	8.33	0.68	8	exp

By the models predicting the sewage inflow Q to the plant the number of neurons on the network hidden layer was equal to 5 for CNN method and it was equal to 6 by MLP method; the activation function on the hidden layer was hyperbolic tangent in both cases. The calculation errors obtained by Q prediction were MAE=1373 m^3/d, MAPE=3,26% and R=0,96 in case of CNN method and they were MAE=1900 m^3/d, MAPE=5,35% and R=0,83 in case of MLP method. Those values are smaller than the errors received by Bartkiewicz and al. (2010, 2016) and by Studziński et al. (2013) who used in their modeling neuronal networks also. On the other hand those values are bigger than these once obtained by Wei et al. (2012) and by El–Din & Smith (2002) (R=0,98).

Based on the measurements of the bioreactor parameters and on the models calculated for prediction of the sewage quality indicators (BOD_5, TN, TP, N-NH_4) and of the sludge sedimentation, the probability of overshooting SVI_{lim} value in the time period 01.01.2014 – 31.08.2014 has been determined

(see Fig. 2). From Fig. 2 results that in the winter time phenomena of activated sludge swelling arisen. Also in that time the predicted values of food to mass ratio and of sludge age amounted to F/M=0,040÷0,075 $gBOD_5/gMLSS \cdot d$ and WO=15÷21 d, respectively. Up to February 2016 (i.e. by 60 days) the F/M variable modelled did not exceed the value of 0,05 $gBOD_5/gMLSS \cdot d$. In that time period as well as in the spring and summer time there was found that a rapid rise of sewage inflow to STP and the resulted sewage dilution, drop of sludge temperature and increase of food to mass ratio lead to the increase of the sludge food to mass ratio and to the sludge swelling; that is indicated by the value of probability of overshooting SVI_{lim} calculated from formula (3). The F/M values obtained from the simulation amounted in those time periods even up to 0,098 $gBOD_5/gMLSS \cdot d$.

However in some days of the spring and summer periods p values determined were significantly lower than 0,50 what indicated on the lack of sludge swelling events. In those days the sludge age equaled to WO=12÷14 d and the food to mass ratio in most cases changed in the values range of 0,055÷0,070 $gBOD_5/gMLSS \cdot d$.

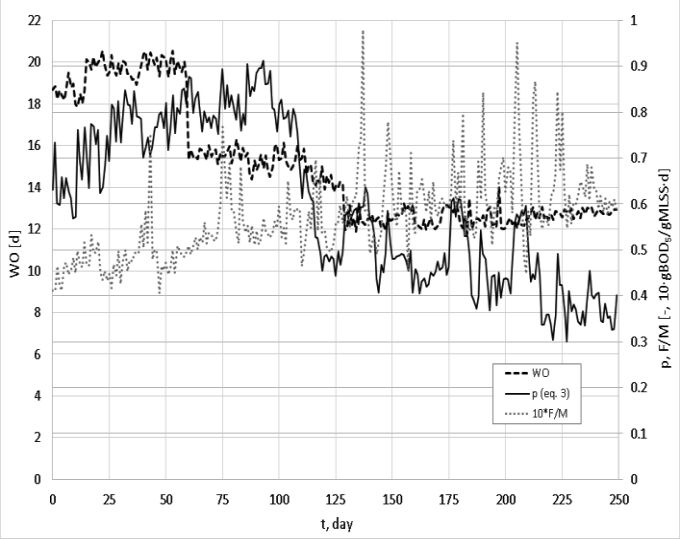


Figure 2. Presentation of measurements data concerning the sludge age (WO) and of calculation results regarding the probability of overshooting SVI_{lim} value and the food to mass ratio (F/M) for the time period investigated.

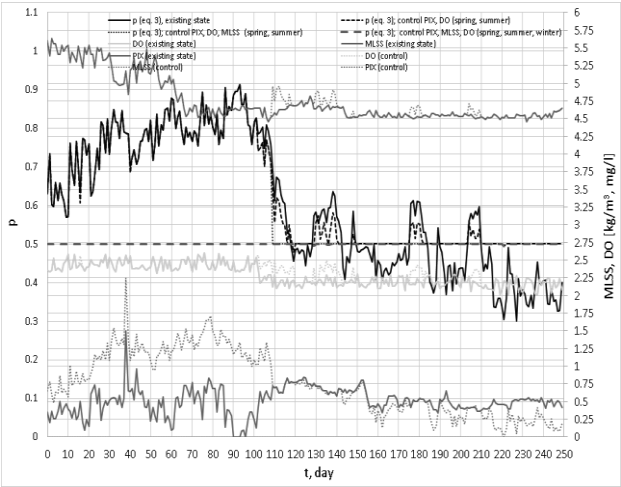


Figure 3. Results of calculation of probability of overshooting SVI_{lim} value for different variants of simulation.

In order to improve the efficiency of STP operation an approach has been taken to apply the calculation results obtained and the model developed for making a correction of values of the bioreactor parameters with the aim of elimination of the sludge swelling.

While doing it on the object investigated a limitation of the amount of coagulant dosed in cases when SVI<150 cm^3/g has been examined. In the analyses done a reduction of the daily intake of related chemicals was tested so that the p value calculated then amounted to 0,50 what corresponded with the SVI value equal to 150 cm^3/g. When in the summer time the measured SVI value exceeded the limit of 150 cm^3/g then concurrent corrections of oxygen concentration (that was raised by 0,20 mg/dm^3) and of sludge concentration have been taken so that the p value calculated amounted again to 0,50. If despite the corrections of the oxygen and sludge concentration (to achieving maximally the level of MLSS=5,0 kg/m^3) the p value calculated was bigger than 0,50 then the amount of coagulant dosed was raised to the value determined by formula (3).

In the winter period the values of DO and MLSS were all the time exalted and the improvement of the sludge sedimentation ability was then realized by increasing the coagulant dosage.

The results of calculations performed presented in Fig. 3 show that the amount of PIX dosed could be reduced in the summer and spring times. The application of the mathematical model prepared shown the possibility to reduce the coagulant amount from the baseline of 71.1 m^3 to 50.3 m^3 in the time period 21.03.2014 – 09.07.2014. The maximal achieved reduction of the coagulant dosage within that period was above five times, i.e. from the level of 0,48 m^3/d to 0,09 m^3/d.

Moreover in the rainfall periods momentary sludge swelling occurances could be eliminated by raising the oxygen concentration by 0,2 mg/dm^3 and by exalting the sludge concentration according to formula (3) but at most to the level of 5,0 kg/m^3 (see Fig. 3). In the winter time the daily amount of coagulant shall be increased from the value of 1,38 m^3/d to 2,35 m^3/d in order to exclude the sludge swelling events. In that time period which means in our calculations the time frame from 01.01.2014 to 09.04.2014 (i.e. slightly more than 3 months) the total dosage of PIX shall be increased from the used level of 45,5 m^3 to 87,6 m^3 to eliminate the sludge swelling.

It will lead in the practice to the improvement of ability of sewage clarification occuring in secondary clarifiers, to improve the sewage quality on the outlet of STP as well as to restricting the problems of sludge dewatering.

CONCLUSIONS

In the paper a methodology of mathematical modelling of activated sludge sedimentation in the bioreactor of a communal sewage treatment plant by means of logit regression models has been described and tested. Those sedimentation models are developed by using parameter measurements of the bioreactor for which the sewage quality indicators have been predicted with the aid of mathematical modelling. Several simulation runs done confirmed that the sewage quality prediction is possible while using only measurements of the amount and temperature of sewage inflow to the treatment plant; by that modelling better

prediction results have been got by means of the method of cascade neuronal networks than of multilayer perceptron ones.

An analysis of calculation results obtained shows that on the treatment plant investigated in the summer time the amount of coagulant dosed could be essentially reduced without worsening the sludge settle ability what was indicated by the value of sludge volume index not exceeding 150 ml/g. Based on those results one can say also that in the winter periods the sludge swelling events can be eliminated by increasing the daily coagulant dosages. In cases of rainfall episodes leading to disturbances of treatment plant operation the improvement of sludge dropping ability was possible by appropriate increasing of oxygen and sludge concentration in the sewage.

The calculations of modelling and simulation performed indicate that operational activities carried out on the plant and consisting in appropriate controlling of changes concerning coagulant amounts added to the sewage or concentrations of oxygen and activated sludge residing in the biologic reactor will lead to the improvement of operational efficiency of the object and to increase of its operational reliability.

REFERENCES

Abyaneh H.Z.: Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering* 12(40), 1–8 (2014).

Al–Batah S.B., Alkhasawneh M.S., Tay L.T., Ngah, U.K., Lateh H.H., Isa M.T.A.: Landslize occurrence prediction using trainable cascade forward network and multilayer perceptron. *Mathematical Problems in Engineering* (20)15, 1–9 (2015).

Bartkiewicz L., Studziński J.: Mathematical modeling of the hydraulic load of communal wastewater networks. In: *Modeling and Simulation* (G.K. Janssens, K. Ramakers, A. Caris, Eds.), EUROSIS-ETI, Hasselt (Belgium), pp. 156–160 (2010).

Bartkiewicz L., Szeląg B., Studziński J.: Ocena wpływu zmiennych wejściowych oraz struktury modelu sztucznej sieci neuronowej na prognozowanie dopływu ścieków komunalnych do oczyszczalni. *Ochrona Środowiska*, Vol. 38, No. 2, 29-36 (2016).

Bagheri M., Mirbagheri S.A., Bagheri Z., Kamarkhani A. M.: Modelling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach. *Process Safety and Environmental Protection*, Vol. 26, pp. 12–25 (2015).

Bayo J., Angosto J. M., Serrano-Aniorte J.: Evaluation of physicochemical parameters influencing bulking episodes in a municipal wastewater treatment plant. *Water Pollution* VIII: Modelling, Monitoring and Management. 95, 531–542 (2006).

Belanche L., Valdes J., Comas J., Roda I., Poch, M.: Prediction of the bulking phenomenon in wastewater treatment plants. *Artificial Intelligence in Engineering* 14(4), 307-317 (2000).

Bezak-Mazur E., Stoińska R., Szeląg B.: Ocena wpływu parametrów operacyjnych i występowania bakterii nitkowatych na objętościowy indeks osadu czynnego – studium przypadku. *Rocznik Ochrona Środowiska*, 18, 487–498 (2016).

Capizzi G., Sciutto G.L., Monforte P., Napoli C.: Cascade feed forward neural network based model for air pollutants evaluation of single monitoring stations in urban areas. *International Journal of Electronics and Telecommunications* (61)4, 327–332 (2015).

Capodaglio A., Jones H., Novotny V., and Feng X.: Sludge bulking analysis and forecasting: Application of system identification and artificial neural computing technologies. *Water Research*, 25(10), pp. 1217–1224 (1991).

Cortés U., Martínez J., Comas M., Sànchez-Marrèa M., Rodríguez I.: A conceptual model to facilitate knowledge sharing for bulking solving in wastewater treatment plant. *AI Communications*, 16, pp. 279–289 (2006).

Dańczuk M., Łomotowski J.: Application of the microwave energy to the hygienization of sewage sludge. *Environment Protection Engineering*, Vol. 4, No. 36, pp. 77–85 (2010).

Dogan E., Ates A., Yilmaz E.C., Eren B.: Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand. *Environmental progress* 27(4), 439–446 (2008).

El-Din A.G., Smith D.W.: Modelling approach for high flow rate in wastewater treatment operation. *Journal of Environmental Engineering and Science*, Vol. 1, No. 4, pp. 275–291 (2002).

Flores-Alsina J., Comas I.R., Roda M., Poch K., Gernaey V., Jeppsson, U.: Evaluation of plant-wide WWTP control strategies including the effects of filamentous bulking sludge. *Water Science and Technology*, Vol. 60, No. 8, pp. 2093–2103 (2009).

Flores-Alsina X., Comas J., Rodriguez-Roda I., Gernaey K.V., Rosen Ch.: Including the effects of filamentous bulking sludge during the simulation of wastewater treatment plants using a risk assessment model. *Water Research*, 43(18), 4527–4538 (2009).

Han H., Li Y., Guo Y., Qiao, J.: A soft computing method to predict sludge volume index based on a recurrent self-organizing neural network. *Applied Soft Computing* (38) 477–486 (2016).

Han H.G., Qiao J.F.: Hierarchical neural network modelling approach to predict Sludge Volume Index of wastewater treatment process. *Control Systems Technology*, IEEE Transactions on 21, pp. 2423–2431 (2013).

Harrell F.: Regression modeling strategies with application to linear models, logistic regression and survival analysis. Springer Verlag, New York (2001).

Heyer T., Stamm J.: Levee reliability analysis using logistic regression models – abilities, limitations and practical considerations. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, Vol. 7, No. 2, pp. 77–87, (2013).

Jurik L., Kaletova T., Sedmakova M., Balazova P., Cervenanska A. Comparison of service characteristics of two town's WWTP. Journal of Ecological Engineering. Vol. 18, No. 3, 61– 67.

Kusiak A., Verma A., Wei X.: A data–mining approach to predict influent quality. *Environmental Monitoring and Assessment* (185), 2197– 2210 (2013).

Lou I., Zhao Y.: Sludge bulking prediction using principle component regression and artificial neural network. *Mathematical Problems in Engineering*, pp. 1–17 (2012).

Martins A.M.P., Heijnen J.J., van Loosdrecht M.C.M.: Bulking sludge in biological nutrient removal systems. *Biotech. Bioeng.*, 86 (2), pp. 125–135 (2004).

Minsoo K., Yejin K., Hyosoo K., Wenhua P., Changwon K.: Evaluation of the k–nearest neighbour method for forecasting the influent characteristics of wastewater treatment plant. *Frontiers of Environmental Science & Engineering* 10(2), 299–310 (2016).

Szeląg B., Gawdzik J.: Application of selected methods of artificial intelligence to activated sludge settle ability Predictions. *Polish Journal of Environment Studies* 25(4), 1709–1714 (2016).

Rutkowski L.: Flexible neuro-fuzzy systems: structures, learning and performance evaluation. *Springer Science & Business Media*, Kluwer Academic Publisher (2004).

Studzinski J., Bartkiewicz L., Stachura M.: Development of mathematical models for forecasting hydraulic loads of water and wastewater networks. EnviroInfo 2013: *Environmental Informatics and Renewable Energies*, Shaker Verlag, Aachen (2013).

Szeląg B. (2017). Stochastic model for estimating the annual number of storm overflow discharges. *Proceeding of the fifth National Congress of Environmental Engineering*, Lublin, 29.05-01.06 (2016).

Szeląg B., Bąk Ł.: Probabilistic model for the annual number of storm overflow discharges in a storm water drainage system. *Urban Water Journal*, Vol. XX, No. XX, 1–8 (2016).

Szeląg B., Gawdzik J.: Assessment of the effect of wastewater quantity and quality, and sludge parameters on predictive abilities of non-linear models for activated sludge settle ability predictions. *Polish Journal of Environmental Studies*, Vol. 26, No. 1, pp. 315–322 (2017).

Verma A., Wei X., Kusiak A.: Predicting the total suspended solids in wastewater: data-mining approach. *Engineering Applications of Artificial Intelligence*, Vol. 26, No. 4, pp. 1366–1372 (2013).

Wei X., Kusiak A., Sadat H.R.: Prediction of influent flow rate: data-mining approach. *Journal of Energy Engineering*, Vol. 19, No. 2, pp. 118- 123 (2012).

MULTISCALE SIMULATION AND OPTIMIZATION: NATURAL GAS TO PRODUCTS

Abhinav Verma
Stanislav Jaso
Alejandro Lopez
Jiaqi Chen
Maarten Bracht

Shell Global Solutions International B.V.
Grasweg 31
1013HW Amsterdam, Netherlands
E-mail: Abhinav.A.Verma@Shell.com

KEYWORDS
Multiscale, Optimization, Natural Gas, Gas To Liquids, Shell, Modelling, GTL

ABSTRACT

Shell Gas to Liquids (GTL) technology converts natural gas into high quality liquid products that would otherwise be made from crude oil. Shell began developing GTL technology in the 1970s and operates commercial GTL plants in Bintulu, Malaysia (startup 1993) and the world's largest GTL plant, Pearl GTL, in Qatar (startup 2011). At the heart of our GTL plants is our proprietary Shell GTL technology with over 500 patent assets in developing the GTL process. We continue improving our GTL technology by enhancing our designs, improving operations, adding new GTL products and modelling plays an essential role in the different levels of GTL development and operation. For instance, models help us making strategic decisions and lowering the cost development and design by using master-planning/networking optimization tools and flow-sheeting, design and process synthesis models. Molecular models help understanding and tuning fundamental chemistry and catalysis. Integrated design and economic models help to optimize the value and prioritize investment decisions. Operating, monitoring and optimization models help to drive effectiveness and efficiency. Models also allow us to explore the boundaries and test constraints and assumptions.

Modelling and optimization of an integrated process at a plant scale is a complex task with various elements and models representing the different time and length scales. Different methods and techniques are available for varying scales/disciplines for representing the subsystem with sufficient accuracy (see Figure 1). For example, the internal chemistry of the reactor itself is as important as the energy/heat balance of the plant. For a unified representation of a plant itself, it is therefore important to combine the different length and time scales in the models for a suitable representation. Further operational constraints become an important part of the whole model that combines details from physics/chemistry to engineering and operational complexity

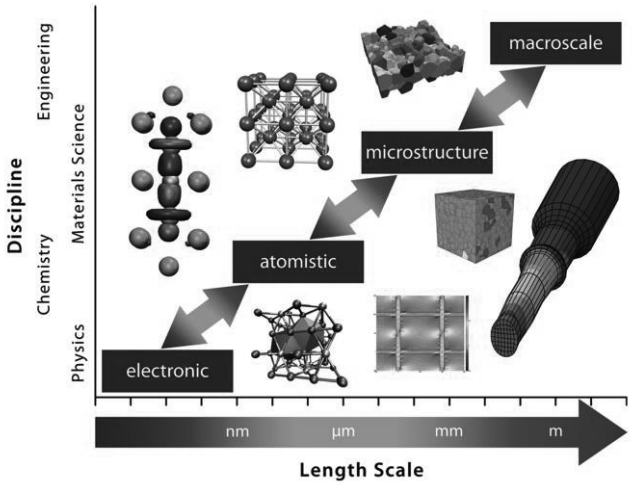


Figure 1: Length/Time Scales and disciplines involved in a complex optimization problem

In the GTL process, natural gas (mainly CH4) is first converted to syn-gas (CO and H2), which is further combined in a Fischer-Tropsch process to make long chain hydrocarbons and final products as schematically shown in Figure 2.

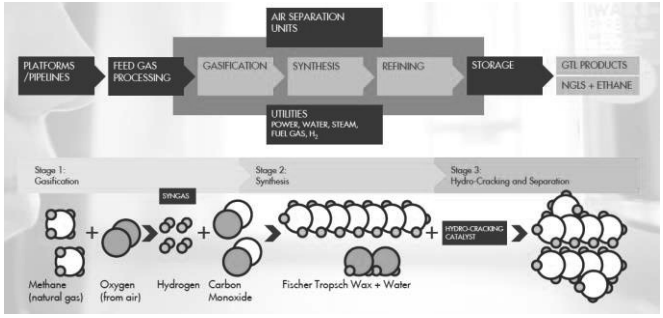


Figure 2: Schematic of a GTL process

To effectively model the whole plant, it is important to understand the process from chemical catalysis to the operational reactor constraints and stream flows. We have combined the different models based on the understanding of the processes from different levels of operation and fitting the data to physical/chemical or empirical relations. Such a

high-level model assists in a daily optimal operation, scheduling and value optimizations.

To obtain level of detail necessary for the operational optimization or design optimization, it is necessary that the key trends, that are rooted in fundamental physics and chemistry are translated to the plant optimization level. We achieve this through various length scales simulations and appropriate stitching of the scales.
First step is electronic and atomistic understanding, that we obtain by performing in-house DFT (density functional theory) and MD (molecular dynamics) simulations (Car and Parinello, 1985). Typically, here we identify the morphology of the ideal FT catalyst (Agarwal et. al. 2018) and possible reactions steps that lead to the formation of how desired and undesired products. In addition, we perform MD simulations to obtain transport parameters like diffusion coefficients of different species through different mediums, that are later used in the reactor simulations. MD simulations are also used to predict what facets are most likely to be present on the catalyst surface, and most likely contribute to the overall reaction rate that is observed (see Figure 3).

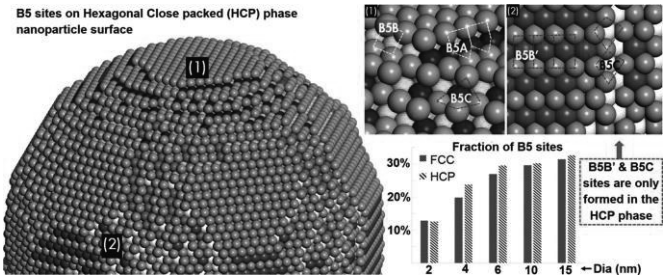


Figure 3: Results of the MD simulations, showing which facets are present at which Co nano-particle diameter (Agarwal et. al. 2018)

Next step is to come up with a "skeleton" of the kinetic model that can be used to predict the performance of an ideal catalyst. We typically perform KMC (kinetic Monte-Carlo) simulations with inputs from DFT models to construct how a reaction would proceed on the surface of the active material and predict rates of different reaction steps. Typically rate determining reaction steps from KMC models are used to build micro-kinetic model that can be later used to explain experimental measurements.

Measurements of "intrinsic" kinetics is performed in ideal "perfectly stirred" Berty-type reactors, and screen all conditions that a catalyst pellet in industrial conditions may encounter (see Figure 4). We perform fitting exercise to determine rate equation parameters. Rate equation themselves are either derived or inspired from the KMC model, to preserve the underlying physical and chemical phenomena.
After the intrinsic kinetics are known, the following step is to perform modeling of the entire reactor. Modeling of the entire reactor, with a distributed catalyst pellet is represented by DAE (differential algebraic equation) systems describing a combination of the kinetics and the mass transfer phenomena. A platform like gPROMS is typically used to solve the equation set. Model validation is performed using ~ 250 data points from experiments, and model results shows an excellent match to what is observed experimentally (see Figure 5).

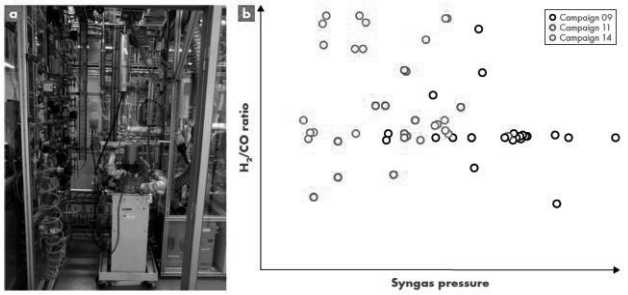


Figure 4: Apparatus for measuring intrinsic kinetics (left) and scatter plot of the data collected for constructing the intrinsic kinetics

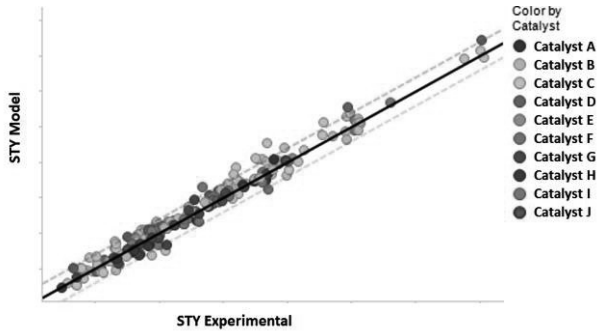


Figure 5: Parity plot – modeled vs experimental STY (Space Time yield) for different types of catalysts. Data are obtained from the pilot plant experiments.

With a validated reactor model, one can perform the plant-wide simulation and optimization – for operational and design purposes. In the past we have already published a work (Ellepola et. al. 2012) that outlines a main approach, that has been further improved with high fidelity reactor models. Namely the main non-linearities come from the Fisher-Tropsch (FT) section of the plant, which we have modeled and validated from molecular to the reactor scale. Exact replica of the reactor model that is used for pilot plant validation is used in the optimization step. Other units such as syngas manufacturing, utilities and hydrocracking can typically be represented with smaller number of algebraic equations, typically representing mass, component and energy balances. A block diagram of a generic GTL complex is represented in Figure 6.

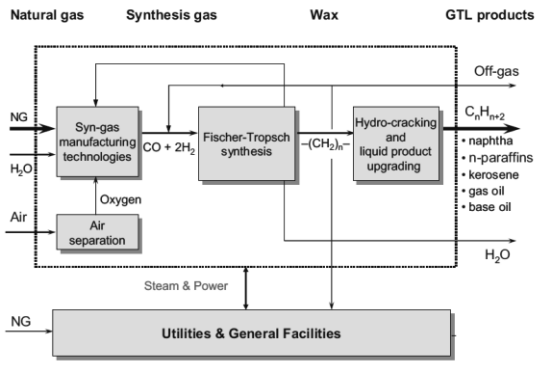


Figure 6: Block diagram of a generic GTL complex (Ellepola et. al. 2012)

To avoid convergence issues and provide robust execution, we typically avoid the use of rigorous thermodynamics, and we use fitted expressions that represent physical properties in the domains we perform optimization. In this way we can obtain optimal designs or optimal operation points within minutes (using AIMMS and/or MATLAB to solve our optimization problem). In view of the size of the problem, our degree of freedom is typically 100-200 (independent process variables) and we normally have 10-20 integer variables related to the choice of equipment. However, the total number of variables and constraints that are required to be satisfied are ~ 15000 – 30000, and most of them can be traced back to discretized DAE system of the FT reactor model. Utilities, and other equipment do not have more than few 100s of constraints.

Having such an optimization model available, we are capable of rapidly screening a large number of variations to the existing GTL line-ups, variations in operating modes, as well as support our existing assets with optimal operation and with de-bottlenecking. The latter can be achieved by determining the most critical parameters in the optimization that are "stuck" at their upper or lower boundaries, as well as by looking at the information of the gradient and Jacobian information for the converged case (i.e. changing which variable at the boundary/constraint would lead to most value).

ACKNOWLEDGEMENTS

We would like to thank Johan Grievink and Nort Thijssen on discussions relating to simulation and optimization.

REFERENCES

Agrawal R., Phatak, P. and L. Spanu. 2018. "Effect of phase and size on surface sites in cobalt nanoparticles",*Catalysis Today*, (in press)

Car, R. and Parrinello, M. 1985. "Unified Approach for Molecular Dynamics and Density-Functional Theory". *Physical Review Letters*, 55, 2471–2474 (and references therein).

Ellepola, J., Thijssen, N., Grievink, J., Baak, G., Avhale, A. and J.van Schijndel. 2012. "Development of a synthesis tool for Gas-To-Liquid complexes" *Computers & Chemical Engineering*, 42, 2-14

WEB REFERENCES

AIMMS Optimization, 2016, https://aimms.com

MATLAB 2016a, The MathWorks, Natick, 2014 http://www..mathworks.com/products/matlab.html

Process Systems Enterprise, 1997-2018, gPROMS, http://www.psenterprise.com/gproms

ABHINAV VERMA was born in Chandigarh, India and went to the Indian Institute of Technology Bombay where he studied physics. He completed his Ph.D. from TU Dortmund, Germany in 2007 in applied and computational physics and worked as a researcher in universities and national labs in US, Spain and Germany. He joined Shell in 2013 as a computational researcher and has been working in industrial modelling and simulation ever since.

MULTI STAGE MONTE CARLO OPTIMIZATION WITH AVERAGING ON A 150 NONLINEAR EQUATION SYSTEM

William Conley
Austin Cofrin School of Business
University of Wisconsin at Green Bay
480B L. G. Wood Hall
Green Bay, Wisconsin 54311-7001
U.S.A.
E-mail: Conleyw@uwgb.edu

KEYWORDS
Complex systems, statistical optimization, survey sampling, estimation, exact solution.

ABSTRACT

Multi stage Monte Carlo optimization (MSMCO) simulation is used here to find the exact solution to a 150 variable 150 nonlinear equation system. MSMCO makes repeated "random" samples of the feasible solution space looking for the minimum total error of the 150 equations. MSMCO follows a trail of the better and better answers (less error). However, after each stage of the simulation the search regions bounds are reduce in size so as it travels across the feasible solution sampling distribution it begins to funnel into the region where it's highly probably that the true optimal or a useful approximate solution may be located. After many of these MSMCO solution tries (all in one program) if the answer is still not found then these approximate solutions are statistically averaged and the process repeated over and over again (all in one simulation program) until the true optimal is found. The overall goal here is to try to develop a general purpose solution technique for nonlinear systems of equations that are so prevalent in business, science and many fields of engineering.

INTRODUCTION

The linear system of equations:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 43 \\ 2x_1 + 3x_2 + 5x_3 + x_4 + 6x_5 &= 108 \\ 3x_1 + x_2 + x_3 + 8x_4 + 9x_5 &= 163 \\ 7x_1 + 5x_2 + x_3 + x_4 + x_5 &= 163 \\ 6x_1 + x_2 + x_3 + x_4 + 9x_5 &= 121 \end{aligned}$$

could be solved with substitution and/or cancellation techniques after courses in algebra and linear algebra. However, it could also be solved by using MSMCO after transferring this system to minimize:

$$\begin{aligned} f(x_1,x_2,x_3,x_4,x_5) = \\ +|x_1 + x_2 + x_3 + x_4 + x_5 - 43| \\ + |2x_1 + 3x_2 + 5x_3 + x_4 + 6x_5 - 108| \\ + |3x_1 + x_2 + x_3 + 8x_4 + 9x_5 - 163| \\ + |7x_1 + 5x_2 + x_3 + x_4 + x_5 - 163| \\ + |6x_1 + x_2 + x_3 + x_4 + 9x_5 - 121| \end{aligned}$$

down to zero error with a multi stage Monte Carlo optimization (MSMCO) simulation to reveal the answer $x_1=14$, $x_2=9$, $x_3=7$, $x_4=12$ and $x_5=1$. Please see Figures 1 and 2 for a partial geometric and statistical representation of this process.

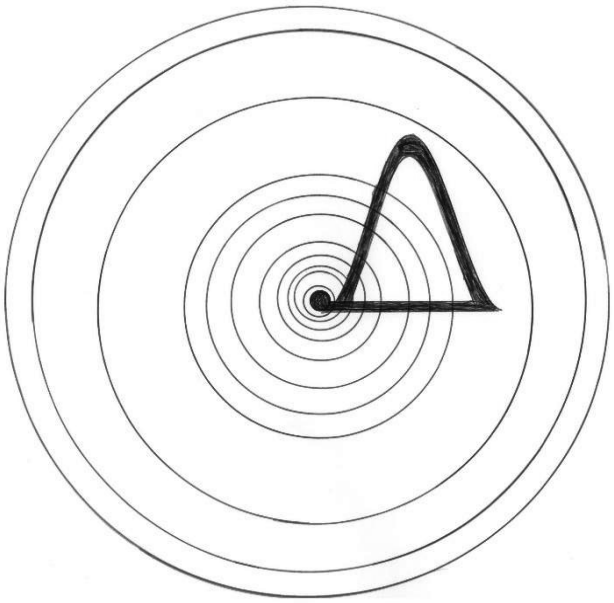


Figure 1: N Dimensional Circles or Spheres Closing in on the Minimum Solution

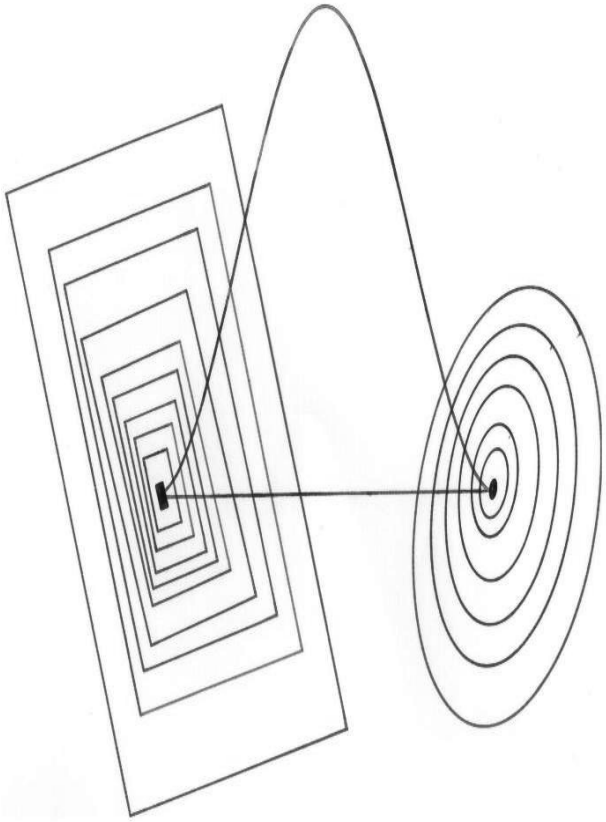


Figure 2: Minimizing with K Dimensional Rectangles and Maximizing Functions with K Dimensional Spheres

Now let us consider a larger nonlinear problem and try a similar approach.

A 150 EQUATION 150 VARIABLE NONLINEAR SYSTEM

It is important to the organization to find an all integer solution (whole numbers in this case) to the following system of equations:

$x_i^3 + x_{151-i} \cdot x_{150-i} + x_{149-i} + x_{148-i} = C_i$

for i = 1, 2, 3, . . . 147

and

$x_{148}^3 + x_{70}\, x_{66} + x_{131} = C_{148} = 6010$ (148)

$x_{149}^3 + x_{56}\, x_{111} + x_{88} = C_{149} = 4860$ (149)

$x_{150}^3 + x_{25}\, x_{95} + x_{142} = C_{150} = 47259$ (150)

for $1 \le x_i \le 50$ and whole numbers (Note that there are about 1×10^{254} or 1×10^{255} feasible solutions and the correct one is required.) for i=1,2,3, . . . 150 all of the constants (C_is) for the right hand sides of the equations follow here in the pattern:

12772= C_1 C_2 C_3 C_4 C_5
C_6 C_7 C_8 C_9 C_{10}
" " " " "
" " " " "
" " " " "
" " " " "
" " " " "
C_{146} C_{147} C_{148} C_{149} C_{150}

Note that the four written out constants above (C_1, C_{148}, C_{149} and C_{150}) are underlined for ease of reading.

Table 1: Right Hand Side of Equations Constants

12772	17921	786	14363	13572
15752	125123	27387	51315	1669
64307	74493	703	433	5013
60084	28880	61392	10830	51987
28169	858	43489	74576	74222
27151	5529	79591	117704	5536
17700	1842	81342	125584	11234
79952	1767	111194	27352	2924
8253	167	138	64755	22592
2973	97521	104469	20846	730
850	85649	10219	125939	1768
27623	16956	69484	13898	36008
47074	27342	1564	179	30146
889	778	126437	117908	50805
54971	237	731	33105	125886
6522	3546	64095	284	12153
1994	106283	27335	4971	85322
477	2296	3143	13304	6166
866	5952	33834	69732	3145
18191	40407	20624	3806	210
85493	18905	14371	837	47089
6036	74149	42	125081	20008
16074	2825	125593	1057	111631
2486	112780	51212	379	59853
4268	2155	27803	8594	1553
4550	37444	6144	17844	80677
27846	125888	65226	92319	4048
165	27103	10824	6640	3842
74538	4560	12196	1324	27607
97922	1572	6010	4860	47259

Therefore, the 150x150 nonlinear system is transformed to

minimize $f(x_1,x_2,x_3 \ldots x_{150}) = \sum_{i=1}^{150} |L_i - C_i|$

subject to that $1 \le x_i \le 50$ and all whole numbers with the L_i and C_i being the left and right hand sides of equation i for i = 1, 2, . . . 150.

THE SOLUTION APPROACH OR ALGORITHM

Note that real arithmetic will be used on each x_i until the modal averaging starts to work. Then by the end of the program simulation an all whole number answer will appear.

The computer simulation starts with a 31 stage multi stage Monte Carlo optimization (MSMCO) solution attempt drawing 70,000 feasible solutions in the $1 \leq x_i \leq 50$ for I = 1, 2, . . . 150 range using real values for x_i's. Then a second 31 stage MSMCO solution try is attempted and these two solution try answers are compared coordinate wise (the x_i's). If any of the two values for the same variable are within .015 of each other, their values are averaged and rounded to the nearest whole number. An example would be if in the first two MSMCO simulation attempts when these two best answers found are compared coordinate-wise if the two x_i values are 23.010 and 22.998. Then 23.010-22.998 = .012 which is less than our allowable width bound of .015 (a measure of deviation). Therefore, the average of 23.010 and 22.998 rounded to the nearest whole number is 23 and x_i is fixed at 23 for the remainder of the simulation.

These first two solution tries of 70,000x31=2,170,000 feasible solutions had total errors of approximately 200,000 and resulted in about two dozen of the 150 x_i values being pinned down for the next two solution attempts.

Then two more MSMCO solution tries of 70,000x31 stages=2,170,000 sample answers are tried and the total error drops to about 150,000 and more x_i's are pinned down for the remainder of the simulation program algorithm.

These two pairs of 70,000x31=2,170,000 sample answers with subsequent pinning down of x_i's that are less than .015 apart are done seven more times.

Their total errors are

86,290	60,638
20,327	21,634
7,835	6,552
2,479	2,078
291	925
82	70
.406	.267

Then in stage 19 the total error is reduced to 0.000 and the following solution emerges and is printed. It checks exactly in all 150 nonlinear equations. The solution is printed in the pattern:

X_1	X_2	X_3	X_4	X_5
X_6	X_7	X_8	X_9	X_{10}
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"
X_{146}	X_{147}	X_{148}	X_{149}	X_{150}

THE SOLUTION

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

The potential of this type of solution approach on a wide variety of multivariate nonlinear optimization problems seems to be enormous.

However, there are risks" involved in this approach. Therefore, let us think about this a bit.

DISCUSSION

One criticism of the Monte Carlo or multi stage Monte Carlo optimization (MSMCO) was always that no one ever knew if there was a better answer (higher profit, lower cost or lower pollution, etc.) than the one that was produced in the simulations. However, when one uses it to solve a system of equations (as was done here) the 0.000 total error term is self-checking that the answer is exactly correct with no error because

$$f(x_1,x_2 \ldots x_{150}) = \sum_{i=1}^{150} |L_i - C_i| = 0.000$$

Still there are the risks of thousands or even trillions of local optimals in the k dimensional space as k (the number of variables) increases to large value. Also, it is important not to pin down an x_i value incorrectly at any stage of the simulation. Drawing large samples helps to overcome these problems.

However, just as the insurance industry uses applied statistics (actuarial mathematics) and computers to manage risks, the potential with the computing power available in the 21st century seems to be there to "manage" the risks mentioned while using MSMCO simulation.

Last century (1980's and 1990's particularly) (Conley 1989) used modal averaging of up to a mode of 7 (only fixing a value after it occurred seven times) because relatively small samples were being used (and also due to the complexity of the famous test problems (Conley 1991a) and (Conley 1991b) being studied along with new ones at the time.

Instead of the mode would it be acceptable (or even better) to use the arithmetic mean or a weighted mean (higher weights for better approximate solutions)? Also, instead of a somewhat arbitrary measure of deviation of .015, would it be better to use some function of the standard deviation for closing in to improve the overall answer? The author experimented with these successfully in the 20th century on various problems using slower computers.

However, the increases in speed and capacity of computers this century makes all of these approaches more viable today. Also, local optimals tend to pile up around the true optimal in large nonlinear systems. That can be exploited statistically in the simulation to improve answers.

There is another risk on the other side of this matter. The elegance and certainty with which linear algebra can solve linear systems of equations and linear programming problems makes practitioners susceptible to linearizing essentially nonlinear optimization problems. Then one gets the exact solution to the possibly "wrong question", versus MSMCO simulation getting possibly approximate solutions to a more accurate question.

Both approaches appear to be useful as our computer age moves forward.

CONCLUSION

Presented here was the featured example of a 150 variable and 150 nonlinear equation system solved with multi stage Monte Carlo optimization (MSMCO) with averaging to fine tune the simulation over such a gigantic feasible solution space.

Figures 1 and 2 are a partial representation of the closing in of the subsequent stages as the program proceeds across the feasible solution space. (Anderson and Sweeney 1999), (Anderson 2003), (Black 2014), (Hayter 2002), (Keller and Warrack 2003) and (McClave, Benson and Sincich 2001) provide good review of statistics.

(Conley 2017) and (Conley 2016) present more nonlinear systems solved with MSMCO.

The linear programming and linear algebra techniques are still viable and useful where appropriate. However, MSMCO with averaging allows engineers, scientists and business people to take on the difficult nonlinear problems that are so prevalent in today's world.

REFERENCES

Anderson, D. R., Sweeney, D. J., Williams, T. A. 1999. *Statistics for Business and Economics*. 7th Edition. Southwestern College Publishing, Cincinnati, Ohio.

Anderson, T. W. 2003. *Multivariate Statistical Analysis*. 3rd Edition. Wiley and Sons, New York.

Black, K. 2014. *Business Statistics for Contemporary Decision Making*. 8th Edition. John Wiley and Sons, New York.

Conley, W. C. 1988. "Controlling Costs by Finding the Shortest Route." *Cost Engineering Journal*. November, Vol 30, No. 11, EUROSIS, 20-24.

Conley, W. C. 1989. "Finding the Shortest Routes in Ohio."

Proceedings of the Twentieth Annual Meeting of the Midwest Decision Science Institute. Miami University, Ohio, 2219-221.

Conley, W. C. 1991. "Multi Stage Monte Carlo Optimization Applied to a Two Hundred Point Traveling Salesman Problem." *Proceedings of the 1991 Summer Computer Simulation Conference* (Baltimore). The Society for Computer Simulation, SCS, San Diego, 145-151.

Conley, W. C. 1991. "Programming an Automated Punch or Drill." *International Journal of Systems Sciences*, Vol. 22, No. 11, 2039-2056.

Conley, W. C. 2016. "Statistical Optimization Applied to Chemical Yield Equations." *Proceedings of the 2016 Industrial Simulation Conference*. ISC 2016, Universites Politechnica, Bucharest, Rumania. EUROSIS, Ostend, Belgium.

Conley, W. C. 2017. "Statistical Optimization Applied to a Variety of Nonlinear Systems of Equations." In *Proceedings of the 2017 European Simulation and Modelling Conference.* Ostend, Belgium. ESM207-ETI, 54-59.

Hayter, A. J. 2002. *Probability and Statistics for Engineers and Scientists*. 2nd Edition. Duxbury Press, Pacific Grove, CA.

Keller, G. and Warrack, B. 2003. *Statistics for Management and Economics*. 6h Edition. Thompson Brooks/Cole, Pacific Grove, CA.

McClave, J. T., Benson, P.G., Sincich, T. 2001. *Statistics for Business and Economics*. 8th Edition. Prentice Hall, Inc. Upper Saddle River, New Jersey.

BIOGRAPHY

WILLIAM CONLEY received a B.A. in mathematics (with honors) from Albion College in 1970, an M.A. in mathematics from Western Michigan University in 1971, a M.Sc. in statistics in 1973 and a Ph.D. in mathematics-computer statistics from the University of Windsor in 1976. He has taught mathematics, statistics, and computer programming in universities for over 30 years. He is currently a professor emeritus of Business Administration and Statistics at the University of Wisconsin at Green Bay. The developer of multi stage Monte Carlo optimization and the CTSP multivariate correlation statistics, he is the author five books and more than 200 publications world-wide. He is a member of the American Chemical Society, a fellow in the Institution of Electronic and Telecommunication Engineers and a senior member of the Society for Computer Simulation. He was also named to KME, the National Mathematics Honorary.

PREDICTION DATA MANAGEMENT

Modeling Football Match Scoring Outcomes using Multilevel Models

Liberato Camilleri, Naomi Farrugia
Department of Statistics and Operations Research
University of Malta
Msida (MSD 06) Malta
E-mail: liberato.camilleri@um.edu.mt

KEYWORDS
Hierarchical nested data, random coefficient model, intra class correlation, multilevel model.

ABSTRACT

Multilevel modelling technique recognizes the existence of hierarchal structures in the data by allowing for random effects at each level in the hierarchy, thus assessing the variation in the dependent variable at several hierarchical levels simultaneously. Multilevel modelling is becoming an increasingly popular technique for analysing nested data with such popularity accredited to the computational advances in the last two decades. In many sports, including football, the game fixtures are nested within seasons, which in turn are nested within country leagues invoking a multilevel structure in the data. Many gaming companies engage in sport data analysis in a bid to understand the dynamics and patterns of the game. This will assist the gaming company in developing fantasy sport games that will enhance gamer engagement and augment revenue to the company.

This paper presents a comprehensive description of two and three level models, which are applied to a real football data set accessed from an online free football betting portal. The aim is to examine the relationship between the number of goals scored during a football match and several game-related predictors. These multilevel models, which assume a Poisson distribution and a logarithmic function, are implemented using the facilities of GLLAMM (Generalized Linear Latent and Mixed Models), which is a subroutine of STATA.

1. Introduction

The concept of Generalized Linear Models (GLMs) was first introduced by Nelder and Wedderburn (1972) where several widely used distributions, including the Normal, Poisson, Binomial, Gamma, Geometric, Multinomial and Inverse Gaussian distribution were combined together as members of the exponential family. The iteratively reweighted least squares algorithm was used for maximum likelihood estimation. A fundamental assumption of GLMs is that the responses are independent making these models inappropriate for longitudinal data, repeated measures and multilevel data with a nesting structure. To overcome this limitation, Liang and Zeger (1986) developed the concept of Generalized Estimating Equations (GEE) by removing the independence assumption. This development gave rise to GEE models that accommodate highly correlated data by specifying a structure for the working correlation matrix. To accommodate nested hierarchical structured data, Bryk and Raudenbush and (1992) introduced the concept of multilevel models. In contrast with the GLM and the GEE, these models take into consideration the hierarchical nature of the nested data by accommodating the error term and random effects at each hierarchical level of nesting. The development of software packages and the introduction of supercomputers alleviated the implementation of multilevel models to large data sets, particularly when the hierarchical structure exceeds two levels of nesting and the number of random effects is considerable.

2. Theory

In a generalized linear model framework, the expected value of the response $\mathbb{E}(y_{ij}) = \mu_{ij}$ is related to linear predictor η_{ij} through a non-linear invertible link function $g(\cdot)$ given by:

$$\mathbb{E}(y_{ij}) = g(\eta_{ij})$$

In this generalized linear model, the response mechanism is fully described by the conditional probability density function of the response y_i given the linear predictor η_{ij}. The model is completed by specifying a distribution for the observed response $y_{ij}|\mu_{ij}$, which in the case of count data is the Poisson distribution with parameter μ_{ij}.

$$\mathbb{P}(y_{ij}|\mu_{ij}) = \frac{{\mu_{ij}}^{y_{ij}} e^{-\mu_{ij}}}{(y_{ij})!}$$

where

$$\mathbb{E}(y_{ij}) = Var(y_{ij}) = \mu_{ij}$$

The link function $g(\cdot)$ for count data is the logarithm link specified in the following way:

$$g^{-1}(\mu_{ij}) = \log \mu_{ij} = \eta_{ij}$$

A Poisson model assumes that the duration of the observation period is fixed in advance (constant exposure); however, this is not always the case. The model can be extended further by including a varying exposure rate m_{ij}. As a result the Poisson regression model can be written in the form:

$$\log \mu_{ij} = \log m_{ij} + \beta_{0j} + \beta_{1j} x_{1ij} + \cdots + \beta_{pj} x_{pij}$$

This implies that the relationship between μ_i and the linear predictor η_i is offset by the amount $\log m_{ij}$. This term is a fixed part offset and if required, it is centred on the mean so as to avoid numerical instabilities. Yet, we do not always require an offset, or where the offset is a constant. A two-level random intercept model with one explanatory variable x_{1ij} can be provided for count data and is given by:

$$\log \mu_{ij} = \log m_{ij} + \beta_{0j} + \beta_{1j} x_{1ij} + U_{0j}$$

An extended two-level random intercept model with several explanatory variables is given by:

$$\log \mu_{ij} = \log m_{ij} + \boldsymbol{x}'\boldsymbol{\beta} + U_{0j}$$

Similarly, the two-level random coefficient, 1-predictor model for count data is given by:

$$\log \mu_{ij} = \log m_{ij} + \beta_{0j} + \beta_{1j} x_{1ij} + U_{0j} + U_{1j} x_{1ij}$$

More generally, we have that:

$$\log \mu_{ij} = \log m_{ij} + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{z}'\boldsymbol{U}_j$$

$\boldsymbol{U}_j$ follows a multivariate normal distribution $\boldsymbol{U}_j \sim \mathcal{N}(\mathbf{0}, \mathbb{T}_j)$ respectively with:

$$\mathbb{T}_j = \begin{bmatrix} var(U_{0j}) & cov(U_{0j}, U_{1j}) \\ cov(U_{1j}, U_{0j}) & var(U_{1j}) \end{bmatrix} = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{bmatrix}$$

A three-level random intercept model with one explanatory variable x_{1ijk} can be provided for count data and is given by:

$$\log \mu_{ijk} = \log m_{ijk} + \beta_{0jk} + \beta_{1jk} x_{1ijk} + U_{0jk} + V_{00k}$$

An extended three-level random intercept model with several explanatory variables is given by:

$$\log \mu_{ijk} = \log m_{ijk} + \boldsymbol{x}'\boldsymbol{\beta} + U_{0jk} + V_{00k}$$

where $U_{0jk} \sim \mathcal{N}(0, \tau_0^2)$ and $V_{00k} \sim \mathcal{N}(0, \theta_0^2)$
Similarly, the three-level random coefficient, 1-predictor model for count data is given by:

$$\log \mu_{ijk} = \log m_{ijk} + \beta_{0jk} + \beta_{1jk} x_{1ijk} + U_{0jk} + U_{1jk} x_{1ijk} + V_{00k} + V_{10k} x_{1ijk}$$

where $U_{1jk} \sim \mathcal{N}(0, \tau_1^2)$ and $V_{10k} \sim \mathcal{N}(0, \theta_1^2)$

More generally, we have that:

$$\log \mu_{ijk} = \log m_{ijk} + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{z}'^{(2)}\boldsymbol{U}_j + \boldsymbol{z}'^{(3)}\boldsymbol{V}_k$$

3. Application

The data set is sourced from www.football-data.co.uk, a free football betting portal that provides historical results and odds. The dataset comprises information about 6,860 football matches, two professional European football leagues and ten football seasons dating from 2005/2006 to 2014/2015. One of the European leagues is the German football league, the Bundesliga, where in every season there are 306 match fixtures. The other football league is the Serie A, an Italian league with 380 match fixtures per season.
Every football match is nested in the season during which it was played and, each season is nested in either one of the two football leagues. This structure invokes the multilevel nature of this data set where the level-1units are the football matches, the seasons are the level-2 units and the football leagues are the level-3 units. The response variable is the number of goals scored per match, and the ultimate scope of the study is to investigate the variability in this response variable induced by observed and unobserved heterogeneity. The following table defines the explanatory variables used in the Poisson multilevel models.

Table 1: Description of predictors

Notation	Predictor
hthg	The total number of goals scored by the home team during the first half
htag	The total number of goals scored by the away team during the first half
sa	The shooting accuracy is the ratio of the total shots on target to the total number of shots
fouls	The total number of fouls committed during the match
cards	Total number of yellow and red cards received during the match
home1h	1 corresponds to a home team win after the first half and 0 corresponds to otherwise
away1h	1 corresponds to an away team win after the first half and 0 corresponds to otherwise
corners	The total number of corners awarded during the match
dhtg	The absolute difference between the total home and away goals after the first half

In the two-level random intercept Poisson model given by:

$$\log \mu_{ij} = \log m_{ij} + \boldsymbol{x}'\boldsymbol{\beta} + U_{0j}$$

$\boldsymbol{x}'$ is a row vector including the values of the explanatory variables, $\boldsymbol{\beta}$ is a column vector of regression parameters in the fixed component of the multilevel model and U_{0j} is the random intercept with distribution $U_{0j} \sim \mathcal{N}(0, \tau_0^2)$. In this section, a parsimonious two-level random intercept Poisson model is fitted using seven predictors $x_{1ij}, \ldots, x_{7ij}$, where $\beta_1, \ldots, \beta_7$ are the corresponding parameters and β_0 is the intercept parameter. The model is implemented using the facilities of GLLAMM.

The adaptive quadrature converged after two iterations and another five iterations were needed to update the parameters using the Newton-Raphson algorithm. The log-likelihood of the parsimonious two-level random intercept Poisson model is -11255.4. The explanatory variables *dhtg* and *corners* were not significant and so were removed from the model fit. The estimated parameters $\beta_0, \beta_1, \ldots, \beta_7$ and estimated variance τ_0^2 are displayed in Table 2.

Since the mean and variance of the Poisson distribution are equal then the variance to mean ratio is 1. Thus a value of 1 is used for level 1 variance. The fractions of residual variability that are attributed to level 1 and level 2 are 0.975 and 0.025

respectively. This implies that 97.5% of the total variance is accounted for by level-1 variations between matches and 2.5% is accounted for by level 2 variations between seasons.

Table 2: Parameter estimates, standard errors and p-values

Parameter	Coef.	S.E.	Z	$P>\|z\|$
Constant	-2.285	0.050	-45.6	0.000
hthg	0.273	0.011	25.6	0.000
htag	0.288	0.012	24.3	0.000
sa	1.455	0.074	19.7	0.000
fouls	-0.006	0.001	-5.52	0.000
cards	0.008	0.004	2.09	0.037
home1h	0.100	0.021	4.68	0.000
away1h	0.116	0.023	4.98	0.000
offset	2.398			
Level-1 var.	1			
Level-2 var. (int.)	0.025	0.007		

Figure 1 displays the path diagram to present the structure of the implemented 2-level random intercept model.

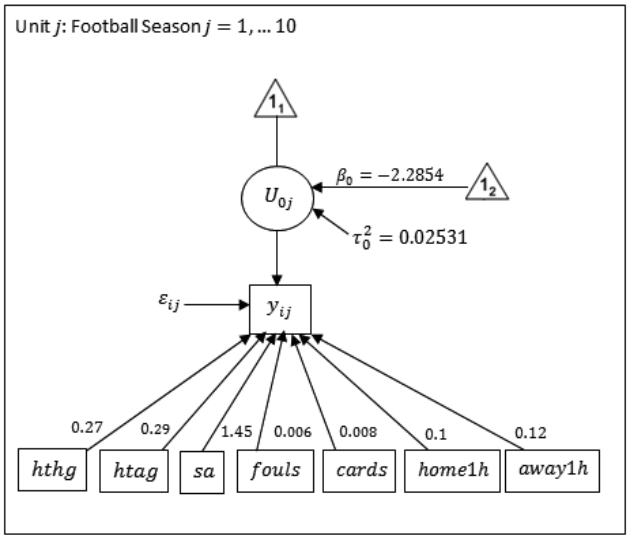


Figure 1: Path diagram for 2-level random intercept model

The gllapred poster directive is used to estimate the posterior means and posterior standard deviations using empirical Bayes prediction for random effects. For this 2-level random intercept model, different posterior means and posterior standard deviations are estimated for each of the 10 seasons. In order to predict the level-2 units specific regression lines with varying intercepts, the parameter estimates and the empirical Bayes estimates of the random intercept are plugged into the model. The gllapred predict, linpred directive is used to compute the linear predictor of the fixed component and adds it to the posterior mean.

The posterior standard deviations are the conditional standard deviations of the prediction errors given the observed responses and treating the parameters as known in a Bayesian context. Taking the square root of these standard deviations, one gets the conditional mean squared error of prediction conditional on the observed responses. The empirical Bayes estimates of the random intercept $U_{0j} \sim \mathcal{N}(0, 0.025)$ for each level-2 unit $j = 1,2, \dots ,10$ along with the posterior standard deviations are provided in Table 3.

Table 3: Posterior means and posterior standard deviations

Season	Posterior Mean	Posterior St. Deviation
1	0.0322	0.0288
2	0.0286	0.0228
3	0.0741	0.0223
4	0.0399	0.0229
5	0.0474	0.0231
6	0.0617	0.0228
7	0.6658	0.0230
8	0.0990	0.0229
9	-0.0219	0.0231
10	-0.1704	0.0231

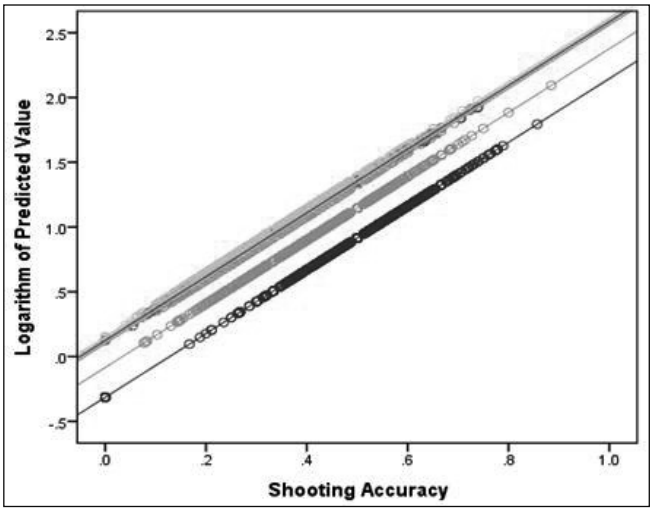


Figure 2: Log of predicted values against shooting accuracy

Figure 2 displays the logarithm of the predicted values against the shooting accuracy of the football match. The parameter estimate of *sa* is 1.4554 which implies that for every 1 unit increase in shooting accuracy the logarithm of the predicted value is expected to increase by 1.455, given that the other effects are kept fixed. The ten seasons trajectories displayed in Figure 2 have different intercepts but the same gradient as conditioned by two-level random intercept model. The ten trajectories have positive gradients implying that the number of goals per match increases with the shooting accuracy. It can be noted that the trajectories for the 2014/2015 and 2013/2014 seasons are below the other seasons which implies that in the last two football seasons the number of goals scored per match was less compared to the other seasons.

In the two-level random coefficient Poisson model given by:

$$\log \mu_{ij} = \log m_{ij} + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{z}'\boldsymbol{U}_j$$

$\boldsymbol{U}_j$ includes the random intercept U_{0j} and the random slope U_{1j}. The row vector $\boldsymbol{z}'$ holds a single explanatory variable, x_{3ij}, which is the shooting accuracy during the match. So the random slope U_{1j} allows the linear relationship between the logarithm of predicted values and shooting accuracy to have a

different slope for each of the ten seasons. $\boldsymbol{x}'$ includes the values of the predictors and $\boldsymbol{\beta}$ holds the regression parameters. The log-likelihood of the parsimonious two-level random coefficient model is -11253.9. The parameter estimates $\beta_0, \beta_1, \ldots, \beta_7$ and the estimated variances τ_0^2, τ_1^2 and τ_{10} are displayed in Table 4. Figure 3 displays the path diagram to present the structure of the implemented 2-level random coefficient model.

Table 4: Parameter estimates, standard errors and p-values

Parameter	Coef.	S.E.	Z	$P>\|z\|$
Constant	-2.296	0.048	-48.1	0.000
hthg	0.271	0.011	25.4	0.000
htag	0.288	0.012	24.3	0.000
sa	1.592	0.101	15.8	0.000
fouls	-0.005	0.001	-5.13	0.000
cards	0.007	0.004	1.96	0.049
home1h	0.100	0.021	4.67	0.000
away1h	0.113	0.023	4.86	0.000
offset	2.398			
Level-1 var.	1			
Level-2 var. (int.)	0.021	0.034		
Level-2 var. (slope)	1.158	0.537		
Level-2 covariance	-0.152	0.156		

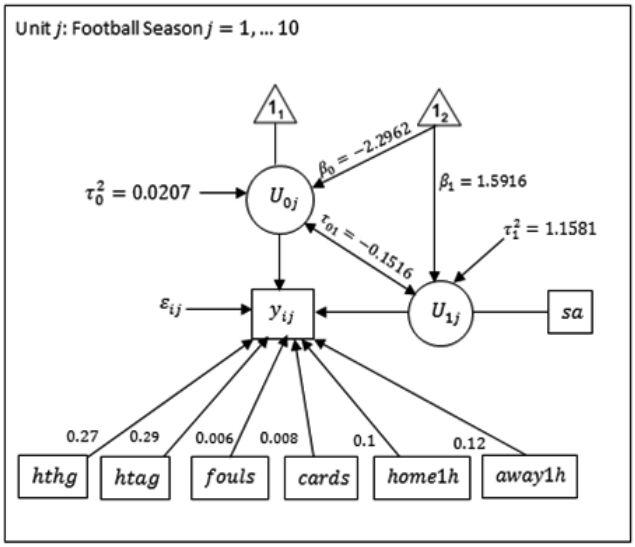


Figure 3: Path diagram for 2-level random coefficient model

Comparing these estimates to the random intercept model, the fixed effects estimates have not changed substantially but the estimates of the covariance matrix are quite different. The fraction of total residual variance attributed to the level-1 units is 0.459 and the fraction of total variance attributed to the level-2 random intercept and random slope are 0.01 and 0.531 respectively. This implies that 45.9% of the total variance is accounted for by level-1 variations between matches, 1% of the variance is accounted for variations between season intercepts and 53.1% of the variance is accounted for variations between season slopes. The empirical Bayes predictions for the random intercepts and the random slopes of the ten seasons are provided in Table 5 along with the posterior standard deviation.

$$U_j = \begin{bmatrix} U_{0j} \\ U_{1j} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.0207 & -0.1516 \\ -0.1516 & 1.1581 \end{bmatrix} \right)$$

Table 5: Posterior means and posterior standard deviations

Season	P.M. 1	S.D. 1	P.M. 2	S.D. 2
1	-0.0044	0.0399	-0.0223	0.1254
2	0.0005	0.3989	-0.0477	0.1260
3	-0.0063	0.0392	0.0762	0.1215
4	-0.0010	0.0391	-0.0193	0.1216
5	-0.0081	0.0404	0.0249	0.1299
6	-0.0253	0.0401	0.1127	0.1263
7	-0.0008	0.0398	0.0341	0.1246
8	-0.0151	0.0409	0.1674	0.1346
9	0.0585	0.0374	-0.3433	0.1009
10	0.0827	0.0364	-0.6210	0.0831

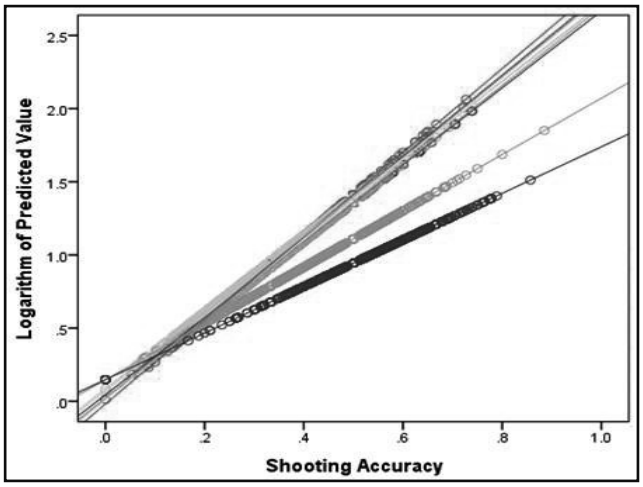


Figure 4: Log of predicted values against shooting accuracy

Table 6: Parameter estimates, standard errors and p-values

Parameter	Coef.	S.E.	Z	$P>\|z\|$
Constant	-2.295	0.053	-43.2	0.000
hthg	0.271	0.011	25.1	0.000
htag	0.288	0.012	24.1	0.000
sa	1.388	0.130	10.6	0.000
fouls	-0.006	0.001	-5.48	0.000
cards	0.009	0.004	2.29	0.022
home1h	0.099	0.021	4.64	0.000
away1h	0.111	0.023	4.78	0.000
offset	2.398			
Level-1 var.	1			
Level-2 var. (int.)	0.011	0.056		
Level-3 var. (int.)	0.019	0.037		
Level-3 var. (slope)	0.334	0.117		
Level-3 covariance	0.074	0.051		

The three-level random coefficient Poisson model given by:

$$\log \mu_{ijk} = \log m_{ijk} + \boldsymbol{x}'\boldsymbol{\beta} + U_{0jk} + \boldsymbol{z}'\boldsymbol{V}_k$$

$\boldsymbol{V}_k$ holds the random intercept V_{00k} and slope V_{10k} at level-3 and U_{0jk} is the random intercept at level-2. $\boldsymbol{z}'$ holds the

predictor, x_{3ij} and $\boldsymbol{x}'$ and $\boldsymbol{\beta}$ are the same as in previous models. The log-likelihood of the parsimonious three-level random coefficient model is -11243.8. The parameter estimates $\beta_0, \beta_1, \ldots, \beta_7$ and the estimated variances τ_0^2, θ_0^2, θ_1^2 and θ_{10} are displayed in Table 6.

This implies that 73.3% of the total variance is accounted for by level-1 variations between matches, 0.8% of the variance is accounted for by level 2 variations between seasons, 1.4% is accounted for by level 3 variations between football league intercepts and 24.5% is accounted for by level 3 variations between football league slopes.

Figure 5 displays the path diagram to present the structure of the implemented 3-level random coefficient model.

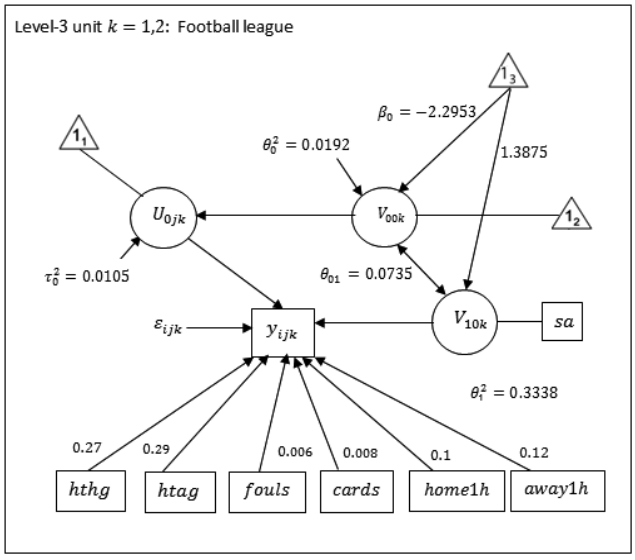


Figure 5: Path diagram for 3-level random coefficient model

$$U_{0jk} \sim \mathcal{N}(0, 0.011)$$

$$\boldsymbol{V}_k = \begin{bmatrix} V_{00k} \\ V_{10k} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.019 & 0.074 \\ 0.074 & 0.334 \end{bmatrix} \right)$$

The empirical Bayes estimates for the random intercept U_{0jk} for $j = 1,2, \ldots, 10$ along with the posterior standard deviations are given in Tables 7, along with the Bayes estimates for the random effects V_{00k} and V_{10k} for $k = 1,2$ and their posterior standard deviations given in Table 8.

Table 7: Posterior means and posterior standard deviations

Season	Posterior Mean	Posterior St. Deviation
1	-0.1370	0.0428
2	-0.0089	0.0419
3	0.0460	0.0419
4	-0.0040	0.0426
5	0.0080	0.0427
6	0.0527	0.0425
7	0.0097	0.0428
8	0.0758	0.0422
9	0.0664	0.0425
10	-0.2573	0.0475

Table 8: Posterior means and posterior standard deviations

League	P.M. 1	S.D. 1	P.M. 2	S.D. 2
1	0.0096	0.0328	0.1002	0.0705
2	0.0259	0.0336	0.1508	0.0729

The 3-level random coefficient model provides the best fit because it has the lowest AIC value.

4. Conclusion

Football is a game that has matured over the years, where football players run faster, they shoot harder, they dribble quicker and, they pass the ball more accurately. As a result, game practices including offside traps, pressing and triangular passing have evolved over time. Changes in these techniques are the main reason why goal scoring has gradually decreased from an average of approximately 4.5 goals per game in 1900 to an approximate average of 2.6 goals more than 100 years later. Goal scoring has remained essentially stable in the last two decades. Results in chapter 4 confirm the latter statement, since goal scoring is not affected much by the football season during which the game was played.

This paper presents a proper methodology to model count data in the presence of nested data. The three level random coefficient model which included shooting accuracy both as a main effect and as a random effect showed that 73.3% of the total variation is accounted for by variation at level-1, 0.8% is accounted for by variation at level-2, 1.4% is accounted for by variation at level-3 intercepts and 24.5% is accounted by variation in level-2 slopes. Moreover, shooting accuracy, number of fouls, number of red and yellow cards booked by referees, number of goals scored during the first half by the home team and by the away team and whether the home team is winning/losing at half time where all found to be significant predictors of the number of goals scored per match.

References

Bryk, A.S., and Raudenbush, S.W. (1992) Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized Linear Models. Journal of the Royal Statistical Society, Series A, 135, 370-384

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. Biometrika 73, 13-22.

LIBERATO CAMILLERI studied Mathematics and Statistics at the University of Malta. He received his PhD degree in Applied Statistics from Lancaster University. His research specialization areas are related to statistical models, which include Generalized Linear models, Latent Class models, Multilevel models and Mixture models. He is an associate professor and Head of the Statistics department at the University of Malta.

Data Mining in Urgency Department: Medical Specialty Discharge Prediction

Marco Prata
Department of Informatics
University of Minho
Campus Gualtar, Braga
Portugal
a68516@alunos.uminho.pt

Hugo Peixoto, José Machado and António Abelha
Algoritmi Research Center
University of Minho
Campus Gualtar, Braga
Portugal
{hpeixoto, jmac, abelha}@di.uminho.pt

KEYWORDS

Data Mining, Machine Learning, Urgency Department, Feature Selection, Prediction of Medical Specialty Discharge.

ABSTRACT

The aim of this paper is to analyze and process a dataset to predict the Medical Specialty (MS) discharge in a hospital Urgency Department (UD). MS discharge is the medical specialty in which a patient gets discharged from the UD. This predictive analysis would improve medical and staff performance, not to mention, less time consuming to the patient and less expensive to both patient and the hospital. However, it is a challenging task due to the quality of data retrieved from UDs that's usually non-treated, with a lot of irrelevant information, sparse and, sometimes, incomplete. This predictive analysis is obtained through Data Mining techniques and machine learning algorithms in Weka environment. It was concluded that feature selection and structured modelling are important factors that affect classification accuracy. It was also concluded as that randomly decreasing or increasing dataset information by varying patient values does not assist directly in increasing accuracy for the prediction of MS discharge. The best results were achieved using the Bagging algorithm with a REPTree classifier and a ten-fold cross-validation, achieving 91.96 % of accuracy and 0.85 F1-score.

INTRODUCTION

Being able to predict the Medical Specialty discharge of a patient may have an enormous monetary impact, increase of efficiency by medical and clinical staff and reduce of patient waiting time in the Urgency Department. Nowadays, when a patient goes to the Urgency Department it may have to go through three stages: the triage, observation and discharge. At the first stage, the patient is sent to a MS to be observed. After observation, the patient is sent to another MS to perform clinical exams, surgery or drug prescription. Finally, the patient will be sent to another medical specialty or will have medical discharge. You see, sometimes it happens that the triage is incorrectly made, or, for some reason, the patient is not well attended at the observation stage which causes unpredictable scenarios for the patients and unnecessary resources spending.

Though, detection of correctly MS discharge is a very challenging task, data mining and machine learning techniques using Electronic Health Records (EHR) have proved to be very efficient in successfully building clinical decision support systems that present satisfactorily predictive results in a large population of patients.

However, finding significant statistical patterns from UD data is a challenging problem due to its high dimensional nature. UD data frequently contains more than 50 variables, which makes it extremely difficult to process. In addition, this data is incomplete, which would make most of the statistical and empirical analysis complex. It is thus important to detect the factors that enhance or degrade the performance of a clinical decision support system due to the high-dimensional nature of UD data.

In this paper, the process behind the results obtained focus, mainly, on three steps:

- Dataset properties, which include knowing all features included and, possibly, some type of correlation between them;
- Feature selection, along with the dataset properties we also vary the number of patient variables based on a technique; namely, feature selection (Oliveira 2015). We perform classification with as well as without feature selection while varying the lab values. We investigate the classification performance of the dataset with selected features obtained from a feature selection algorithm based on rank aggregation (Sarkar 2012);
- Model Structured information, which consist in the model making before machine learning;

Machine Learning

Machine learning is the science of getting machines/computers to act without being explicitly programmed using algorithms that iteratively learn from data allowing computers to find hidden insights (Rodrigues 2017). The core idea behind machine learning stands on having general systems that can adapt to new circumstances, rather than explicitly writing a different program for each special circumstance (Alpaydin 2010). Nevertheless, the role of computer sciences still plays a major role in this field, as if so, in the training phase, we need

efficient algorithms to solve the optimization problem, as well as to store and process the massive amount of data we generally have. Once a model is learned, its representation and algorithmic solution for inference needs to be efficient as well. In certain applications, the efficiency of the learning or inference algorithm, namely, its space and time complexity, may be as important as its predictive accuracy (Alpaydin 2010).
Having this said, we can classify machine learning tasks into three broad categories (Alpaydin 2010; Russel 2003):

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal. The program is provided feedback in terms of rewards and punishments as it navigates its problem space.

Application of machine learning methods to large databases is called data mining.

Data Mining

Data mining has attracted more and more attention in recent years, probably because of the popularity of the ‘‘big data’’ concept (Xu 2014).
Data mining, aka “Knowledge mining” (‘Knowledge mining from data’), is an interdisciplinary subfield of computer science which involves extracting, discovering, or mining, interesting patterns and knowledge from a large amount of data (Xu 2014; Reddy 2016).
As a highly application-driven discipline, data mining has been successfully applied to many domains, such as business intelligence, artificial intelligence, Web search, scientific discovery, digital libraries. It bridges the gap from applied statistics and artificial intelligence (which usually provide the mathematical background) to database management by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery algorithms more efficiently, allowing such methods to be applied to even larger datasets (Reddy 2016).
Data mining involves six common classes of tasks (Morais 2017):

- Anomaly detection (outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors that require further investigation;
- Association rule learning (dependency modelling) – Searches for relationships between variables. For example, a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis;
- Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data;
- Classification – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam";
- Regression – attempts to find a function which models the data with the least error that is, for estimating the relationships among data or datasets;
- Summarization – providing a more compact representation of the dataset, including visualization and report generation;

BACKGROUND

It is quite accepted in the engineer community that Machine Learning and Data Mining have come to revolutionize several computing areas, from artificial intelligence to stocks value predictions. Nowadays, there is a wide variety of recent works that relate Electronic Health Records, machine learning and UD data. More so, it is necessary to understand the methods and results of the current State of the Art so that it can be possible to structure our approach to the ML problem stated in this article. To do so, it was selected some articles that have ML approaches to a UD. Some of these approaches are: prediction of cardiac arrest in critically ill patients presenting to the UD incorporating heart rate variability (Ong 2012); big data– driven, machine learning approach is compared to existing clinical decision rules (CDRs) and traditional analytic methods using the prediction of sepsis in-hospital mortality as the use case (Taylor 2016); Machine-Learning-Based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index (ESI) (Levin 2017).
In the first study presented, the authors aimed to validate a novel machine learning score incorporating heart rate variability (HRV) for risk stratification of critically ill patients presenting to the UD by comparing sensitivity and specificity for prediction of cardiac arrest with the modified early warning score (MEWS). The MEWS is based on physiological parameters: systolic blood pressure, pulse rate, respiratory rate, temperature and other correlating factors (Ong 2012). A ML-based prediction model utilizing age, HRV parameters, and vital signs was proposed to compute risk score on patient’s hospital outcome. The authors found ML scores to be more accurate than the MEWS in predicting cardiac arrest within 72 hours (Ong 2012).
In the second study, Sepsis was defined as meeting criteria for systemic inflammatory response syndrome with an infectious admitting diagnosis in the UD. A random forest model (machine learning approach) was constructed using over 500 clinical variables from data available within the EHRs of four hospitals to predict in-hospital mortality. The machine

learning prediction model was then compared to a classification and regression tree (CART) model, logistic regression model, and previously developed prediction tools on the validation dataset using area under the receiver operating characteristic curve (AUC) and chi-square statistics (Taylor 2016). The objective in this study was to provide a proof-of-concept example of a local, big data–driven, machine learning approach to predictive analytics in emergency care. The authors chose prediction of sepsis mortality as the use case because of its clinical importance (half of hospital deaths in the United States are related to sepsis) and because early detection and treatment has been shown to improve outcomes (Taylor 2016). As a result, the machine learning approach outperformed existing clinical decision rules as well as traditional analytic techniques for predicting in-hospital mortality of UD patients with sepsis (Taylor 2016).

In the third article, the authors sought to evaluate an electronic triage system (e-triage) based on machine learning that predicts likelihood of acute outcomes enabling improved patient differentiation. A multisite, retrospective, cross-sectional study of 172,726 ED visits from urban and community UDs was conducted. E-triage is composed of a random forest model applied to triage data (vital signs, chief complaint, and active medical history) that predicts the need for critical care, an emergency procedure, and inpatient hospitalization in parallel and translates risk to triage level designations. The outcomes were evaluated and compared with the Emergency Severity Index (ESI). E-triage more accurately classifies ESI level 3 patients and highlights opportunities to use predictive analytics to support triage decision-making (Levin 2017).

Upon analyzing these articles, we found ways to build and evaluate our case study following some of these past approaches, such as seeing how sensitivity and specificity vary (Ong 2012; Taylor 2016) and try to obtain correlation with other metrics that we seem fit. There is a common agreement between the authors that all these matters have potential to be more developed and to be placed in practice, outperforming old practices currently in taking place in UDs.

METHODOLOGY

Materials

1. **Dataset**: The dataset used for training was prepared by using data from an Urgency Department that contained actual labels for MS discharge preventable events for approximately 13000 patients. This data was extracted from a data warehouse previously built for the calendar years of 2016 and 2017. As it was mentioned before, UD data may contain more than 50 features. Of this 50, we removed 22 variables due to sparsity and incomplete data. Of the final 28 variables, it was extracted 9 features that the authors concluded that had high values of correlation between them, including patient demographic data such as provenience, waiting time and age, injury or decease, MS of triage and observation.

Methods

In this section, the methods used to build and evaluate the model performance are briefly explained.

Metric evaluation with accuracy, precision, specificity, sensitivity, and F1-score:

2. **Classification accuracy**: Classification accuracy refers to the number of MS that are correctly identified to belong to a class by the classification algorithm (Fayyad 1996).
3. **Precision**: Also called positive predictive value, is the fraction of relevant instances among the retrieved instances. Precision is the probability that a (randomly selected) retrieved value is relevant (Olson 2008)
4. **Specificity:** Also called the true negative rate, measures the proportion of negatives that are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition). Specificity relates to the test's ability to correctly reject healthy patients without a condition. Consider the example of a medical test for diagnosing a disease. Specificity of a test is the proportion of healthy patients known not to have the disease, who will test negative for it (Altman 1994).
5. **Sensitivity:** Also called the true positive rate, the recall, or probability of detection, measures the proportion of positives that are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition). Sensitivity refers to the test's ability to correctly detect ill patients who do have the condition (Altman 1994). In the example of a medical test used to identify a disease, the sensitivity of the test is the proportion of people who test positive for the disease among those who have the disease.
6. **F1-Score**: In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision and the sensitivity of the test to compute the score. The F1-score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall). F1- measure effectively references the True Positives to the Arithmetic Mean of Predicted Positives and Real Positives, being a constructed rate normalized to an idealized value, and expressed in the form it is known in statistics as a Proportion of Specific Agreement as it is applied to a specific class (Powers 2011).

Prediction of MS discharge depends on intrinsic information of the patients which are aptly represented in the UD data. As a result, prediction performance of a decision tree will have influence from the data quality. Hence, if we vary the number of patients in the dataset, we would be able to capture the trend in increase or decrease in the classification accuracy.

7. **Feature selection**: Feature selection is a very important pre-processing task in data mining. The reason is that not all features/attributes are relevant to a given problem. Some

of the features interfere and reduce prediction accuracy. In addition, dealing with larger number of features is generally considered expensive. Reducing dimensionality is important when dealing with very high-dimensional data such as in electronic health records. Feature selection also helps in providing a better understanding of the process that generated the data (Oliveira 2015). In this paper, we use rank aggregation based feature selection (Sarkar 2012). One advantage of using this feature selection technique is that it has been proved to be more robust and accurate in a previous study (Sarkar 2012) for the prediction of relapse rates in leukemia patients. Rank aggregation is the procedure that takes multiple rankings of a fixed set of alternatives (or candidates) as input, and aggregates them into a single consensus ranking of the candidates. In our case, candidates are the features and the rankers are different feature evaluation techniques such as information gain, symmetric uncertainty and correlation attribute evaluation (Hall 2009).

8. **Classification accuracy with feature selection**: We perform classification for the prediction of PPE after selecting subsets of variables or features that are statistically deemed as significant. For rank aggregation, we use a position based scoring mechanism to calculate the final score of a feature. Next, we sort each feature based on their final score and rank them from highest (being rank 1) to lowest (being rank n, for a total of n features) according to their final scores generated as given in equation (1) (Sarkar 2013).

$$score_{final} = \sum_{i=1}^{n} score_{pos(i,j)} \quad (1)$$

Where n is the total number of features selection techniques (or ranker) used. *pos (i, j)* is the j^{th} position of a feature ranked by the ranker *i*. $score_{pos}$ *(i,j)* is the score of a feature in list I generated by ranker *i* at j^{th} position. *scorefinal* is the sum of all the positional score from all the lists. Here, we consider the feature selection techniques (information gain, symmetrical uncertainty and correlation attribute as the rankers and the candidates as the features (Fonseca 2017). We used the following feature evaluation techniques for ranking the features according to their distinct statistical properties of data, namely:

- Information gain attribute evaluation - This algorithm evaluates features individually by measuring their information gain with respect to the classes.
- Symmetrical uncertainty attribute evaluation – This algorithm evaluates features individually by measuring their symmetrical uncertainty with respect to the class.
- Correlation Attribute Evaluation- Evaluates the worth of an attribute by measuring the correlation (Pearson's) between it and the class.
- The output of this rank aggregation based feature selection is a global rank for each feature.

9. **Model Evaluation**: The model evaluation phase consists in comparing results of classification algorithms on both models and with a baseline classification algorithm. This process consists in obtaining and analyzing the best model performance and compare both models and their respective confusion matrix, error in classification and attribute selection, specificity, sensitivity precision and F1-score.

DATA MINING PROCESS

Business Understanding

We document in this section the different observations obtained through the results:

- Variation of classification accuracy (in percentage) of predicting MS discharge with and without using feature selection with feature selection technique (Sarkar 2012) and varying number of patients;
- Baseline and model Evaluation with Weka (ZeroR, Bagging and J48);

We could define, as a matter of perspective that the model built will perform well if certain evaluations metrics values, mostly F1-score and accuracy, are up to 85%. This will mean that the model is capable of predicting MS discharge with reasonable consistency, having quality to be a matter of study in the upcoming future, perhaps.

Data Understanding

There were previously data treatment before loading the data warehouse. Even though, it was necessary to preprocess data upon inquiring the model in data mining techniques. This process mainly included removal of sparse and incomplete data, normalization and make clear distinction of what would be nominal and numeric in the dataset.

Data Preparation

We divided the processed dataset in four. One dataset with 1000 records, other with 5000, other with 1000 and, finally, one containing all data. This was made in order to evaluate model performance. Meaning, the idea would be dividing in two models, both containing the four previous datasets analysis but one with attribute selection and other without, to evaluate model performance with two different algorithms. This method was chosen due to the lack of quality information compared to the number of attributes/features in the dataset. Dividing in four datasets in each model, will clarify if the domain information is of relevant quality in order to predict MS discharge.

Modeling

As it was mentioned before, in this project Weka was used to test the models. Nevertheless, it is necessary to present the

machine learning algorithms used: *meta.*Bagging, *trees.*J48 and *rules.*ZeroR.

- *meta.*Bagging - Recent developments in computational learning theory have led to methods that enhance the performance or extend the capabilities of these basic learning schemes. Those learning schemes have been called "meta-learning schemes" or "meta-classifiers" because they operate on the output of other learners. Instead of using a single classifier to make predictions, why not arrange a committee of classifiers to vote on the classification? This is the basic idea behind combining multiple models to form an ensemble or meta classifier (Fonseca 2017).
- *trees.*J48 - This is an implementation of C4.5 release 8 (Neto 2017), a standard algorithm that is widely used for practical machine learning. This implementation produces decision tree models.
- *rules.*ZeroR - This is the most primitive learning scheme in Weka. It models the dataset with a single rule. Given a new data item for classification, ZeroR always predicts the most frequent category value in the training data for problems with a nominal class value, or the average class value for numeric prediction problems. Although it seems to make little sense to use this scheme for classification, it can be useful for generating a baseline performance that other learning schemes are compared to. In some datasets, it is possible for other learning schemes to induce models that perform worse on new data than ZeroR which is a clear indicator of serious overfitting (Fonseca 2017).

Evaluation

Next, we describe the evaluation results. Initially, the values presented on Figure 1 refer to detailed accuracy values obtained after applying a classification algorithm, in this case, *meta.*Bagging. As we can see, the classification accuracy values tend to be higher with the increase of MS number. This tell us that this model tends to evolve to higher classification accuracy values with the increase of input information. Even though, it would be expected that values of the classification accuracy with feature extraction to be higher than values with no feature selection, which is not what is verified in this case. With the increase number of patients, we can see that the classification accuracy with feature selection tends to increase exceptions for the last value. This can be explained by the classification algorithm that is being used.

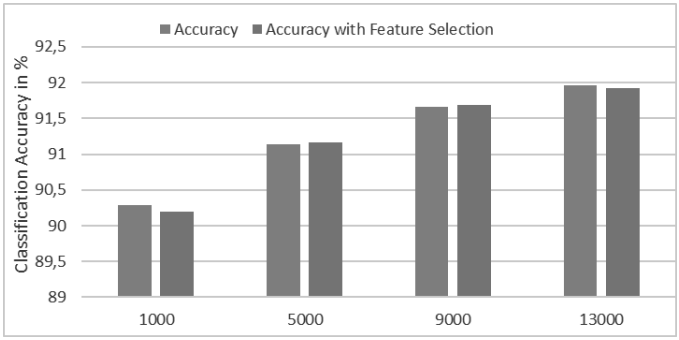


Figure 1: Variation in Classification accuracy with increase number of patients with Bagging algorithm

On a more general analysis of this model, seen of Figure 2, it is possible to observe that the general classification accuracy increases with the increase of data, even though, at the maximum value of data, the accuracy with attribute selection not being the desired. The main objective presented by Figure 2 is to correlate attribute selection with increase, or not, in classification accuracy. Having this said, it was divided in 3 feature selection types: F1, representing removal of one undesired feature from the dataset; F2, removal of two undesired features of the dataset; F3, removal of three undesired features from the dataset. It is important to refer that when using feature selection, not always the same attributes are selected, or removed, from the dataset. For example, when doing feature selection on the dataset with 1000 patients, the attributes selected with less "weight" differ from the ones obtained for the dataset that contains 5000 patients. This might explain why in Figure 1 we achieved an undesired classification accuracy value with feature selection.

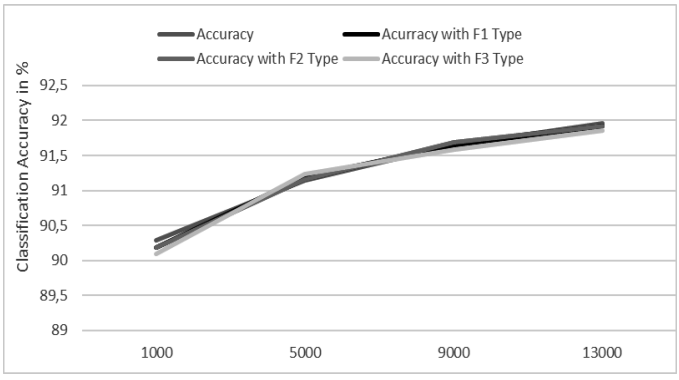


Figure 2: Variation in Classification accuracy with extended feature selection and increase number of patients with Bagging algorithm

Table 1: Detailed values obtained with Bagging algorithm and with feature selection

Instances	Attributes	Accuracy (%)
1000	7	90.1902
5000	7	91.1582
10000	7	91.688
13000	7	91.9219

Table 2: Detailed values obtained with Bagging algorithm and without feature selection

Instances	Attributes	Accuracy (%)
1000	9	90.2903
5000	9	91.1382
10000	9	91.6546
13000	9	91.9605

Next, we describe the evaluation results for the same model but with a different classification algorithm. Initially, the values presented on Figure 3 refer to detailed accuracy values obtained after applying a classification algorithm, in this case, *trees.*J48.

As it happened with the previous model, the accuracy tends to be higher with the increase number of MS. As it also happened with the previous model, this tell us that it tends to evolve into higher accuracy values, with or without feature selection. Similar, the first value containing 1000 MS has a higher value of classification accuracy instead of classification accuracy made with feature selection. This could be explained by the lack of data, different attributes obtained with feature selection for the model or the type of classification algorithm used. However, we can also observe from figure 3 that classification accuracy with feature selection method is consistently higher than without feature selection irrespective of the variation in MS number values. This result, in turn, hints us about the fact that, random variation in MS number values does not contribute to the enhancement of classification accuracy for predicting MS discharge. This consistent increase in the classification accuracy shows that this feature selection technique (Olson 2008) is robust with varying MS quantity.

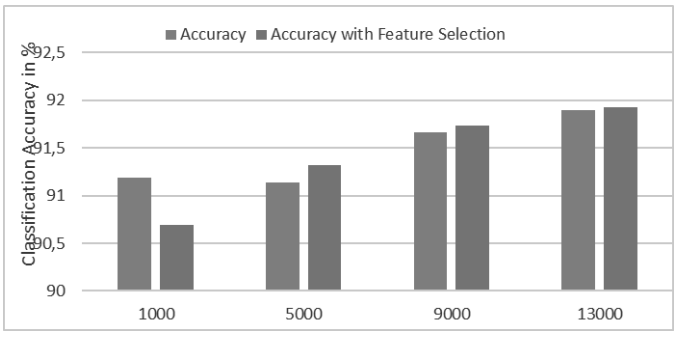


Figure 3: Variation in Classification accuracy with increase number of patients with J48 algorithm

Analyzing Figure 4 and comparing it to Figure 2, since the method to obtain them was the same, we can detect clear differences, not just on values of accuracy, but also on accuracy behavior prediction. In Figure 2, the classification accuracy followed a more linear behavior. In this figure, we can see that the classification accuracy value drops with the increase of number of patients from 1000 to 5000 and then increases with the increase of patient number from 5000 to 13000 patients. We can also identify that the F2 type of feature selection is the best one since its behavior tends to be the best with the increase of data, following a more linear behavior. Analyzing this model in terms of predictions, we may infer that with the increase of data, the classification accuracy, with and without feature selection, would stabilize around 92/93%.

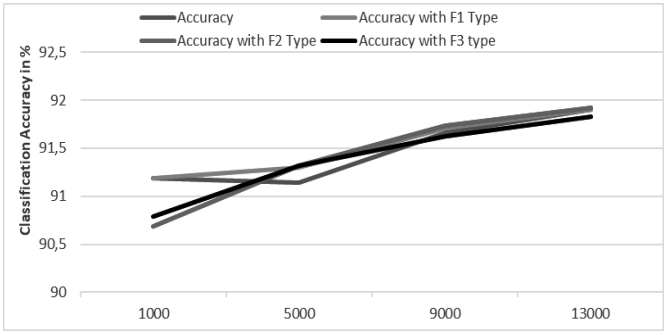


Figure 4: Variation in Classification accuracy with extended feature selection and increase number of patients with J48 algorithm

Table 3: Detailed values obtained with J48 algorithm and with feature selection

Instances	Attributes	Accuracy (%)
1000	7	90.6907
5000	7	91.3183
10000	7	91.7324
13000	7	91.9219

Table 4: Detailed values obtained with J48 algorithm and without feature selection

Instances	Attributes	Accuracy (%)
1000	9	91.1912
5000	9	91.1382
10000	9	91.6657
13000	9	91.8988

These results presented in Tables 1, 2, 3 and 4 show us interesting facts about the classification accuracy and feature selection in the data that are easily overlooked. On one hand, it indicates that eradicating features that have no ponderation/weight on the outcome may not help in increasing classification accuracy. Thus, more information (in terms of number of patient's records) is useful in prediction of MS discharge (we perform *k*-fold cross validation, thus reducing the chance of overfitting of data).

We also observe that feature selection techniques, in some cases, can indeed help in a more accurate prediction of MS discharge. This firstly indicates the fact that more information in terms of feature variables is not very useful. Thus, we conclude that extraction of significant features is an important criterion for a more accurate prediction of MS discharge.

Baseline and Model Evaluation

As it was mentioned before, ZeroR is the most primitive learning algorithm in Weka. It models the dataset with a single rule. Given a new data item for classification, ZeroR always predicts the most frequent category value in the training data for problems with a nominal class value, or the average class value for numeric prediction problems. Although it seems to make little sense to use this algorithm for classification, it can be useful for generating a baseline performance that other learning schemes are compared to. In some datasets, it is possible for other learning schemes to induce models that perform worse on new data than ZeroR which is a clear indicator of serious overfitting (Fonseca 2017).
It was selected from each model the best classification accuracy obtained with feature selection and without. Analyzing figure X, we can see that the baseline classification algorithm presents a rather low classification accuracy. Meaning, no algorithm will perform worse than ZeroR unless. Comparing Bagging and J48 classification, we can denote that the Bagging algorithm performs a little better than the J48. This difference can be explained by the nature of the algorithm, since meta-classifiers are, nowadays, widely known to outperform other classifiers types (Fonseca 2017).

Table 5: Best classification accuracy from each model with and without feature selection. Relative absolute Error*

Algorithm	Instances	Attributes	Accuracy (%)	Error (%)*
ZeroR	13000	9	25.2604	100
ZeroR	13000	7	25.2604	100
Bagging	13000	9	91.9605	16.542
Bagging	13000	7	91.9219	16.7872
J48	13000	9	91.9142	17.1164
J48	13000	7	91.9219	17.1679

Analyzing confusion matrix of both J48 and Bagging classification algorithms, we can see that the incorrectly classified instances are, in some cases, practically the same and with the same values. Why? As it was previously mentioned before, meta-classifiers operate on the output of other learners, and the learner that is being used by Bagging, in this case, is the classification algorithm *trees*.REPTree. Meaning, Bagging is operating on the output of a tree classification algorithm, much like J48, resulting in confusion matrixes with great similarity but, in most cases, with better results. Also, it was possible to calculate and interpolate specificity, sensitivity, precision and f1-score. These values were calculated based on the confusion matrixes and only 9 classes, the ones with most data, were used to interpolate these values and, subsequently, average performance of each model.

Table 6: Metric Evaluation of J48 and Bagging algorithms

Average Values	**J48**	**Bagging**
Precision	0.9199	0.9227
Sensitivity	0.8200	0.8219
Specificity	0.9882	0.9882
F1-Score	0.8539	0.8540

As we can observe, the number of correctly classified classes was practically the same. Nevertheless, it is empirical to notice that the Bagging algorithm performed slightly better than the J48. Even though, looking at this table we can denote that the number of false negatives was way superior to false positive values (difference between precision and sensitivity), it is still considered reasonable results. Looking to the F1-score, the weighted average of precision and sensitivity, both values meet the expected score stipulated as acceptable and of what was expected of the model performance.
Some discrepancy in the results (e.g. difference between sensitivity and the other metrics) may be caused by the lack, or excess, of data in some classes leading to nearly inexistent classification of some and increased number of false negative values justifying the presented lower sensitivity values.

CONCLUSION

In this paper, it was performed an investigation of some important factors that would impact MS discharge prediction. From pre-processing detailed data of a dataset extracted from a data warehouse to comparing different models with two machine learning algorithms, leading to rather conclusive results on the classification outcome.
First and most important, it is possible to predict with and without feature selection, even though with a slight error margin, the MS discharge. It is not linear that the model behavior will present better results with feature selection than without it. Possibly because, with the feature selection method used, some relevant features may be discarded, thus the underperformance in some cases. As it was shown previously, it only happened with one classification algorithm, the J48. Both models, the one tested with Bagging classification algorithm and the one tested with J48, presented an improvement in classification accuracy with the increase number of patients showing a positive prediction for bigger models with, possibly, better results. Comparing both models, the one tested with J48 and the one tested with Bagging algorithm, it is possible to conclude that the Bagging classification it is slightly better, even though showing worst results with feature selection, (around 91.96 % of accuracy and 0.85 F1-score). Future work would rest more upon different feature selection techniques and detailed analysis of data quality to improve the outcome prediction of MS discharge.

ACKNOWLEDGEMENTS

This work has been supported by Compete: POCI-01-0145-FEDER-007043 and FCT within the Project Scope UID/CEC/00319/2013.

REFERENCES

Alpaydın, E. 2010. Introduction to Machine Learning Second Edition Ethem Alpaydın The MIT Press Cambridge, Massachusetts London, England.

Altman D.G., Bland J.M. 1994. "Diagnostic tests. 1: Sensitivity and specificity".

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. 1996. "Advances in knowledge discovery and data mining".

Fonseca, F., Peixoto, H., Miranda, F., Machado, J. and Abelha, A. 2017. "Step Towards Prediction of Perineal Tear", in Proceedings of 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017), Procedia Computer Science, Volume 113, Elsevier.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. 2009. "The weka data mining software: an update". ACM SIGKDD Explorations Newsletter, 11(1):10–18.

Levin, S., Toerper, M., Hamrock, E., Hinson, J. S., Barnes, S., Gardner, H., and Kelen, G. 2017. "Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index". Annals of Emergency Medicine.

Morais, A., Peixoto, H., Coimbra, C., Abelha, A. and Machado, J. 2017. "Predicting the need of Neonatal Resuscitation using Data Mining", in Proceedings of 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017), Procedia Computer Science, Volume 113, Elsevier.

Neto, C., Peixoto, H., Abelha, V., Abelha, A. and Machado, J. 2017. "Knowledge Discovery from Surgical Waiting Lists", in Proceedings of CENTERIS 2017, Procedia Computer Science, Volume 121.

Oliveira, S., Portela, F., Santos, M. F., Machado, J., Abelha, A., Silva,A., & Rua, F. 2015. "Intelligent decision supports to predict patient barotrauma risk in intensive care units". Procedia Computer Science, 64, 626-634.

Olson, D. L., & Delen, D. 2008. "Advanced data mining techniques". Springer Science & Business Media.

Ong, M., Goh, K., Liu, N., Koh, Z. X., Shahidah, N. and Lin, Z. 2012. "Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score". *Critical Care*, *16*(3), R108.

Powers, D. M. 2011. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation".

Reddy, K. S., Sasanka, K., Prasanna, S., & Venkatesan, R. 2016. "GATE: Classification and clustering of text for semi-vowel/j/-morphophonemic approach". In Circuit, Power and Computing Technologies (ICCPCT), International Conference on (pp. 1-7). IEEE.

Rodrigues, M., Peixoto, H., Esteves, M., Machado, J. and Abelha, A.. 2017. "Understanding Stroke in Dialysis and Chronic Kidney Disease", in Proceedings of 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017), Procedia Computer Science, Volume 113, Elsevier.

Russell, S., Norvig, P. 1995. "Artificial Intelligence: A Modern Approach (2nd ed.)". Prentice Hall. ISBN 978-0137903955.

Sarkar, C. and Srivastava, J. 2013. "Impact of density of lab data in ehr for prediction of potentially preventable events."Healthcare Informatics (ICHI), IEEE International Conference on. IEEE.

Sarkar, C., Cooley, S. and Srivastava, J. 2012. "Improved feature selection for hematopoietic cell transplantation outcome prediction using rank aggregation". In Computer Science and Information Systems (FedCSIS), Federated Conference on, pages 221–226. IEEE.

Taylor, R. A., Pare, J. R., Venkatesh, A. K., Mowafi, H., Melnick, E. R., Fleischman, W. and Hall, M. K. 2016. "Prediction of In hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven", Machine Learning Approach. Academic Emergency Medicine, 23(3), 269-278.

Xu, L., Jiang, C., Wang, J., Yuan, J., & Ren, Y.. 2014. "Information security in big data: privacy and data mining". IEEE Access, 2, 1149-1176.

BIOGRAPHIES

MARCO PRATA was born in Braga, Portugal. Currently studying Biomedical Engineering, master's in Medical Informatics at University of Minho. Intern at Philips Research, Eindhoven, for his master's thesis project in Deep Learning.

HUGO PEIXOTO was born in Braga, Portugal and went to the University of Minho where he studied Biomedical Engineering and obtained is Master Degree in 2008. After joining the IT team at Centro Hospitalar Tâmega e Sousa he concluded his PhD in 2013 in Healthcare Information Systems domain. Invited Auxiliary professor at University of Minho he is also a researcher for Centro ALGORITMI since 2009.

JOSÉ MACHADO is Associate Professor with Habilitation at the Department of Informatics of the University of Minho, in Braga, Portugal, where he has been since 1988. He got his PhD in Informatics, in 2002, and habilitation in 2011. Director of Centro ALGORITMI, he is also former Director of the Computer Science and Technology Center.

ANTÓNIO ABELHA is Auxiliary Professor at the Informatics Department at University of Minho where he started in 1996. He got his phD in Informatics in 2004 and he is a research member of the Computer Science and Technology field for Centro ALGORITMI. He is one of the co-founders of the Knowledge Engineering group in 2016, group that he is still a member of.

THREE AND SEVEN COMMODITY ECONOMIC EQUILIBRIUM EQUATIONS

William Conley
Austin E. Cofrin School of
Business 408 Wood Hall
University of Wisconsin at Green
Bay Green Bay, Wisconsin
54311-7001
U.S.A.
Conleyw@uwgb.edu

KEYWORDS
Statistical optimization, economics, equilibrium, multivariate, nonlinear equations, complex systems.

ABSTRACT

The term equilibrium is used and applied in so many fields and in so many ways. This presentation looks at equilibrium in the economic or business sense. After a brief review of one product or commodity equilibrium, two more sophisticated examples involving three and seven commodities that are each substitutes for one another is presented. Then, the question of what prices in combination will lead to supply equaling demand for all three and all seven commodities is answered for each problem using a statistical optimization simulation solution approach.

INTRODUCTION

Figure 1 shows a downward sloping demand equation intersecting and upward sloping supply equation at the equilibrium point where supply equals demand and X is the commodity price and Y is the amount available.

These two equations are linear. An example would be:

Y=1000 – 10X (demand equation)
and Y=10X (supply equation)

where X is the price in dollars and Y is the supply or demand of the commodity in tons. Solving these two equations by substitution yields the equilibrium price of $50 leading to a balanced supply and demand of 500 tons. However, it is possible for the supply and demand equations to be nonlinear (as in Figure 2).

Let us look at two more sophisticated examples.

A THREE COMMODITY EXAMPLE

The company's statisticians and economists working together have managed to estimate the three supply equations and the three demand equations for three commodities that are substitutes for each other. The s_1, s_2 and s_3 are the supply amounts in tons and the d_1, d_2 and d_3 are the demand amounts in tons also. The x_1, x_2, and x_3 are the commodity prices in dollars.

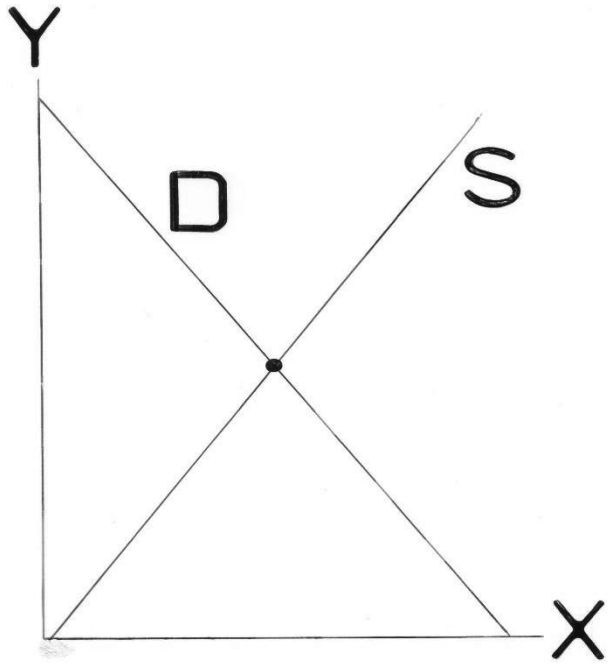


Figure 1: Supply and Demand Equations Intersecting at the Circle or the 50 Dollar Equilibrium Price for X

$$d_1 = 68{,}817 - 9x_1 - 3x_2 - 2x_3 - 6(x_1x_2)^{.5} - (x_1x_3)^{.3} - 6(x_2x_3)^{.5} - 4(x_1x_2x_3)^{.333} \quad (1)$$

$$s_1 = 51{,}183 + 3x_1 + 4x_2 + 7x_3 + 3x_1^{.4}x_2^{.6} + 5x_1^{.3}x_3^{.7} + 8x_1^{.5}x_2^{.55} + 7x_1^{.1}x_2^{.3}x_3^{.4} \quad (2)$$

$$d_2 = 144{,}928 - 11x_1 - 3x_2 - 10x_3 - 8x_1^{.35}x_2^{.4} - 6x_1^{.4}x_3^{.35} - 5x_2^{.4}x_3^{.59} - 7_1^{.2}x_2^{.7}x_3^{.2} \quad (3)$$

$$s_2 = 50{,}072 + 4x_1 + 12x_2 + 2x_3 + 8x(x_1x_2)^{.8} + 3_1^{.6}x_2^{.1} + 7x_2^{.3}x_3^{.7} + 4x_1^{.25}x_2^{.55}x_3^{.2} \quad (4)$$

$$d_3 = 126{,}809 - 2x_1 - 8x_2 - 6x_3 - 4x_1^{.2}x_2^{.9} - 5x_1^{.3}x_3^{.5} - 3x_2^{.7}x_3^{.1} - 5x_1^{.5}x_2^{.3}x_3^{.1} \quad (5)$$

$$s_3 = 113{,}191 + 5x_1 + 4x_2 + 5x_3 + 2x_1^{.6}x_2^{.3} + 5x_1^{.4}x_3^{.5} + 7x_2^{.2}x_3^{.9} + x_1^{.4}x_2^{.1}x_3^{.3} \quad (6)$$

Then try to minimize $f(x_1,x_2,x_3) = |s_1-d_1| + |s_2-d_2| + |s_3-d_3|$ subject to $1 \leq x_1 \leq 1000$ dollars (with accuracy in dollars and cents) using a multi stage Monte Carlo optimization of 50 stages drawing 100,000 feasible solutions at each stage and decreasing the search dimensions in each subsequent stage by a factor of 1.41 (or the square root of 2). A thirty second run on a desk top computer produced the solution of x_1=\$247.11 and x_2 = \$368.80 and x_3 = \$196.21 for the equilibrium of the whole three commodity system. Also, putting these prices into the six equation yields:

$S_1 = D_1 = 60603$ tons
$S_2 = D_2 = 133135$ tons
$S_3 = D_3 = 120560$ tons

The printout below presents the total ($f(x_1,x_2,x_3) = |s_1-d_1| + |s_2-d_2| + |s_3-d_3|$ equation error for the first fifteen stages of the simulation and subsequently the total stage errors for stages 20, 25, 30, 35, 40, 45, and 50.

Table 1: The printout

Stages	Total Error
1	5761.97266
2	3361.78516
3	1916.38672
4	1916.38672
5	712.16016
6	712.16016
7	712.16016
8	635.96875
9	635.96875
10	429.53516
11	396.96484
12	292.97656
13	147.82422
14	139.09766
15	29.06094
..	..
20	8.41797
..	..
25	1.58203
..	..
30	.37813
..	..
35	.07031
..	..
40	.01172
..	..
45	.00000
	..
50	.00000

Notice that the first stage error is 5761.97266 after 100,000 feasible solutions. Then the error terms start decreasing with each subsequent stage until in stages 45 through 50 the total error is 0.00000 or accurate to about five decimals.

DISCUSSION

Even though there are six equations (s_1, s_2, s_3, d_1, d_2, and d_3) that give rise to our f function to be minimized, this is essentially a function f three variables as the problem is to minimize $f(x_1, x_2, x_3)$ down to zero error. Figures 3 and 4 give a partial two-dimensional illustration of the search looking for and finding the minimum.

However, the feasible solution space is three dimensional. Therefore, think of opening a holiday or birthday present given by a friend. It is in a larger rectangular box (the whole feasible solution space). After opening this box, inside it is a second rectangular box with each dimension reduced by a factor of 1.41 (the square root of two). After opening the second smaller box a third, even smaller, rectangular box whose dimensions were also reduced by a factor of 1.41 is found. Opening this box produces another smaller box and so on for 50 concentric smaller boxes. The fiftieth little box contains an emerald ring of great value or the exact solution to our nonlinear system of economic equilibrium equations, so to speak. Any geometric shapes can be used so long as they are decreasing in size and driven by a large enough sample size at each stage to contain some meaningful statistical information of which direction to head in to locate and close in on the true optimal solution or a useful approximation.

Also note that the boxes are re-centered and "free floating" after each better answer is found in the three-dimensional space in the computer simulation. Therefore, in the simulation, the smaller boxes can leap out of the confines of the larger boxes in efficient pursuit of the optimal.

However, in our human reality, the smaller cardboard boxes cannot leap out of the confines of the larger boxes without the gift recipient's help of opening the boxes in order from larger to smallest.

Also, quantum and nuclear physicists such as (Wolfson 2012) and others think that this ability to go through walls "tunneling" may be possible in the future. However, for the present, we will just use statistical optimization simulation where the ever smaller rectangles are free to leap out of any confined space to pursue and find the optimal solution.

THE MORE SOPHISTICATED PROBLEM

The economists and statisticians for a particular industry working together have come up with the following fourteen supply (S_1, S_2, S_3, S_4, S_5, S_6, S_7) and demand (D_1, D_2, D_3, D_4, D_5, D_6, D_7) equations for seven competing commodities, where the X_i's are their prices respectively for i=1, 2, 3, 4, 5, 6 and 7 and the S's and D's are the supplies and demands for the commodities in tons.

$$S_1 = 925766.91 + 12X_1^{.93} + 15X_7^{1.16} + 9(X_1X_2X_3X_4X_5X_6X_7)^{.12}$$
$$S_2 = 942396.22 + 13X_2^{1.14} + 16X_4^{.87} + 11(X_1X_2X_3X_4X_5X_6X_7)^{.11}$$
$$S_3 = 1124124.60 + 9X_3^{1.23} + 16X_5^{.95} + 8(X_1X_2X_3X_4X_5X_6X_7)^{.13}$$
$$S_4 = 861906.47 + 14X_1^{1.08} + 10X_7^{1.06} + 7(X_1X_2X_3X_4X_5X_6X_7)^{.18}$$
$$S_5 = 1147988.82 + 6X_4^{1.09} + 4X_6^{.96} + 11(X_1X_2X_3X_4X_5X_6X_7)^{.16}$$
$$S_6 = 806171.25 + 21X_3^{1.15} + 18X_6^{.86} + 1\ 5(X_1X_2X_3X_4X_5X_6X_7)^{.15}$$
$$S_7 = 741794.6 + 18X_2^{1.10} + 8X_5^{.89} + 17(X_1X_2X_3X_4X_5X_6X_7)^{.14}$$

$$D_1 = 985000 - 6X_1^{1.13} - 11X_3^{.94} - 21(X_1X_2X_3X_4X_5X_6X_7)^{.13}$$
$$D_2 = 1001001 - 17X_3^{1.06} - 13X_5^{1.09} - 12(X_1X_2X_3X_4X_5X_6X_7)^{.15}$$
$$D_3 = 1203458 - 9X_2^{1.12} - 14X_6^{1.06} - 7(X_1X_2X_3X_4X_5X_6X_7)^{.18}$$
$$D_4 = 983499 - 5X_1^{1.11} - 10X_5^{.89} - 14(X_1X_2X_3X_4X_5X_6X_7)^{.19}$$
$$D_5 = 1200325 - 16X_4^{.95} - 8X_7^{1.15} - 11(X_1X_2X_3X_4X_5X_6X_7)^{.10}$$
$$D_6 = 873954 - 4X_2^{1.18} - 5X_4^{.96} - 16(X_1X_2X_3X_4X_5X_6X_7)^{.12}$$
$$D_7 = 909192 - 7X_6^{1.03} - 18X_7^{.85} - 18(X_1X_2X_3X_4X_5X_6X_7)^{.20}$$

Therefore, the problem to solve for total economic equilibrium is:

$S_1=D_1$
$S_2=D_2$
$S_3=D_3$
$S_4=D_4$
$S_5=D_5$
$S_6=D_6$
$S_7=D_7$

With $0 \leq X_i \leq 1000$ dollars and cents (two decimals of accuracy is desired) and no commodities will be priced as high as 1000 dollars or more.

A SEVEN COMMODITY EXAMPLE

Figure 2 shows a supply(s) and demand(s) curve intersecting at its equilibrium point where X is the commodity price in dollars and Y is the supply (and demand) amount in tons. This is fairly easy to see and understand as usually with commodities if the price goes up demand decreases and supply increases. The reverse is true as the price goes down. There are very few times and circumstances when this is not the case.

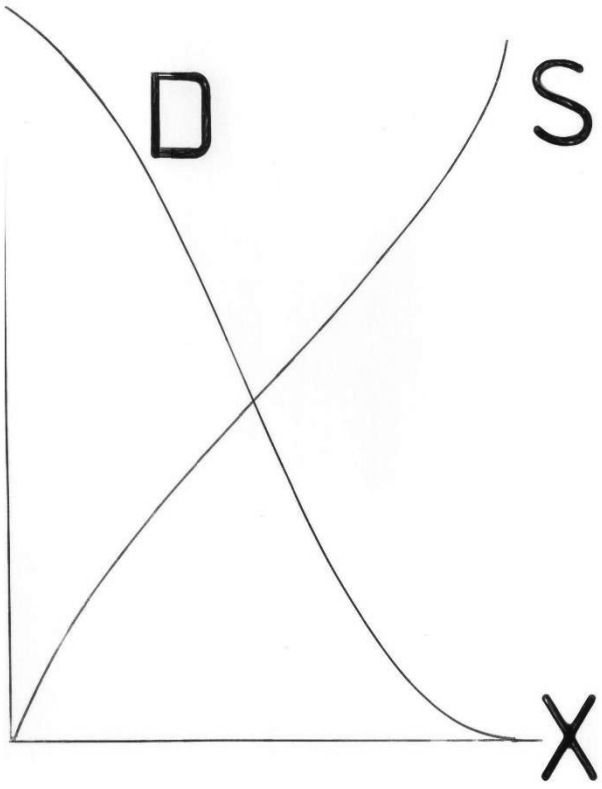


Figure 2: The Nonlinear Forces of Supply and Demand Driven by the Price of the Commodity

The problem is that if a commodity (or several commodities) has substitutes then the supplies and demands are affected not only by their individual prices but also by the other commodity prices. Let us look at a seven commodity example and use MSMCO to solve this more sophisticated problem than the three commodity one was. Figures 3 and 4 give partial geometric and statistical illustration of MSMCO at work.

Therefore, it is transformed to the optimization problem. Minimize $f(x_1,x_2,x_3,x_4,x_5,x_6,x_7) = |S_1-D_1| + |S_2-D_2| + |S_3-D_3| + |S_4-D_4| + |S_5-D_5| + S_6-D_6| + |S_7-D_7|$ subject to $0 \leq X_i \leq 1000$.

Then it is solved with a multi stage Monte Carlo optimization (MSMCO) simulation of 45 ever decreasing in size and repositioning seven dimensional "rectangles" (please see Figures 3 and 4 for partial illustrations of this approach) driven by 100,000 sample feasible solutions at each stage and steered through seven dimensional space by following the better and better (less total error) answers.

The printout of the solution is:

Commodity prices in dollars and cents are X_1=\$505.15, X_2=273.38, X_3=736.41, X_4=824.95, X_5=660.63, X_6=590.86, X_7=783.40 for commodities 1, 2, 3, 4, 5, 6 and 7 respectively.

---

e_1=0.00 e_2=0.00 e_3=0.00 e_4=0.00 e_5=0.00 e_6=0.00 e_7=0.00
individual equation errors

S_1=D_1=965735.94 tons S_2=D_2=957206.38 tons
S_3=D_3=1164688.13 tons S_4=D_4=907042.44 tons
S_5=D_5=1172901.88 tons S_6=D_6=864385.00 tons
S_7=D_7=761872.19 tons
Supplies and demands – all seven are equal

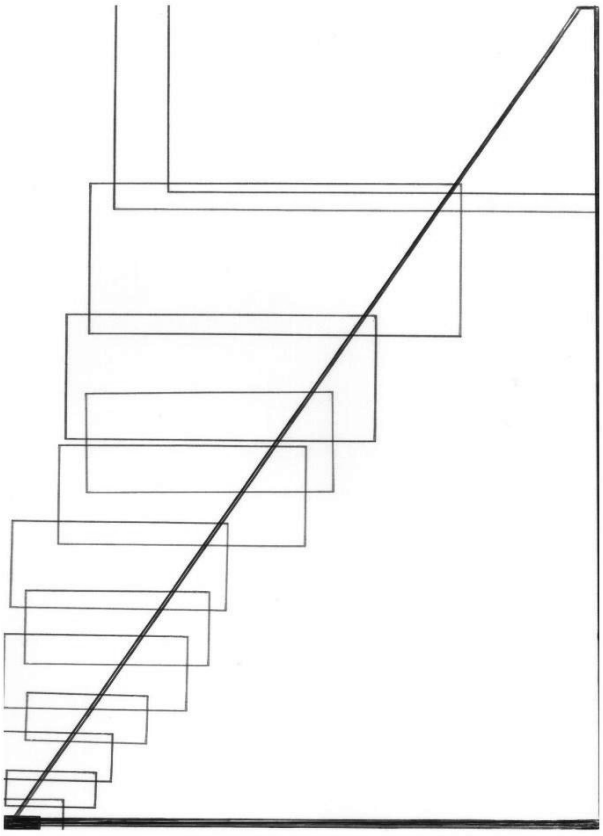


Figure 3. Fourteen Stages of MSMCO Looking for the Minimum Error Solution

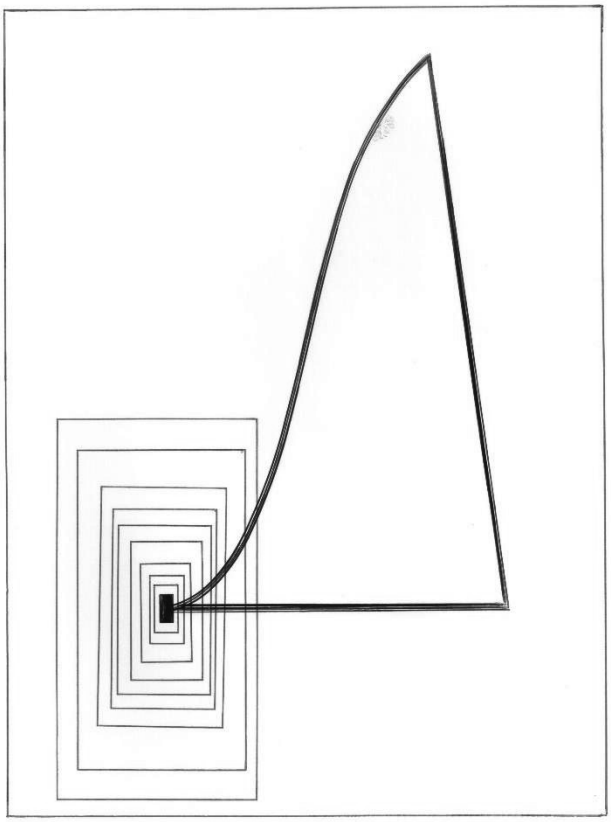


Figure 4: K Dimensional Rectangles Closing in on the Solution in the Lower Left Hand Corner of the Feasible Solution Distribution

DISCUSSION

Therefore, the statisticians, economists and top management know that commodity prices in dollars of \$505.15, \$273.38, \$736.41, \$824.95, \$660.63, \$590.86, and \$783.40 for commodities 1, 2, 3, 4, 5, 6, and 7 respectively will lead to supplies and demands of 965735.94, 957206.38, 1164688.13, 907042.44, 1172901.88, 864385.00 and 761872.19 tons for commodities 1, 2, 3, 4, 5, 6, and 7 respectively. This could be valuable knowledge that managers and decision makers in this industry could act on.

Also, if the government decides to (or is considering) putting price controls on any of the commodities the nonlinear system could be resolved for the new equilibrium (if one exists) or a near approximation given the new temporary, or permanent price controls or constraints. This information could also allow the industry representatives to approach the government and argue against the price controls or constraints. However, sometime governments impost price controls for very good and fair reasons (in the short term anyway.

CONCLUSION

Presented here were hypothetical three and seven commodity economic equilibrium problems where all three and seven commodities are substitutes for each other. The supply and demand equations are all solved simultaneously for the three commodity prices that put the whole system in equilibrium.

The statistical simulation technique multi stage Monte Carlo optimization (MSMCO) was used to solve for the equilibrium prices.

Multi stage Monte Carlo optimization (MSMCO) is a general purpose technique that can be used on a wide variety of problems. Some examples are available in (Conley 1991), (Conley 1993), (Conley 2008) and (Conley 2010). (Hayter 2002) presents statistics for engineers and scientists from a technical view point. Good economic reviews are presented in (Samuelson and Nordhaus 2009) and (Dorfman, Samuelson and Solow 1987). (Samuelson and Nordhaus 2010) and (Keller and Warrack 2003) present statistics from managerial and economic viewpoints.

REFERENCES

Conley, W. C. 1991. "Programming an automated punch or drill." *International Journal of Systems Science*, Vol. 22, No. 11, pp. 2039-2056.

Conley, W. C. 1993. "Multi Stage Monte Carlo Optimization Applied to a Six Hundred Point Traveling Salesman Problem. *International Journal of Systems Science,* Vol. 24, No. 60, pp. 69-626.

Conley, W. C. 2008. "Ecological optimization of pollution control equipment and planning from a simulation perspective," *International Journal of Systems Science*, Vol. 39, No. 1, pp. 1-7.

Conley, W. C. 2010. "We can save the world if worlds collide." Industrial Simulation Conference, IS, Budapest, June 9-12. EUROSIS-ETI, Ghent, Belgium, pp. 45-49.

Dorfman, Robert; Samuelson, Paul, and Solow, Robert. 1987. *Linear Programming and Economic Analysis*. Dover Books.

Hayter, A. J. 2002. *Probability and Statistics for Engineers and Scientists.* 2nd Edition. Duxbury Press, Pacific Grove, CA.

Keller, G. and Warrack, B. 2003 *Statistics for Management and Economics*, 6th Edition. Thompson Books/Cole, Pacific Grove, CA.

Samuelson, Paul and Nordhaus, Robert. 2010. *Macroeconomics*. 19th Edition. McGraw-Hill, New York.

Samuelson, Paul and Nordhaus, Robert. 2009. *Economics*. 19th Edition. International Edition. McGraw-Hill, New York.

Wolfson, Richard. 2012. *Essentials of University Physics*. 2nd Edition. Addison-Wesley, Pearson Education, San Francisco, CA.

BIOGRAPHY

WILLIAM CONLEY received a B.A. in mathematics (with honors) from Albion College in 1970, an M.A. in mathematics from Western Michigan University in 1971, an M.Sc. in statistics in 1973 and a Ph.D. in mathematics-computer statistics from the University of Windsor in 1976. He has taught mathematics, statistics, and computer programming in universities for over 30 years. He is currently a professor emeritus of Business Administration and Statistics at the University of Wisconsin at Green Bay. The developer of multi stage Monte Carlo optimization and the CTSP multivariate correlation statistics, he is the author of five books and more than 200 publications world-wide. He is a member of the American Chemical Society, a fellow in the Institution of Electronic and Telecommunication Engineers and a senior member of the Society for Computer Simulation. He was also elected to Omicron Delta Epsilon (ODE) the national economics honorary.

OLIVE OIL SCREENING

Henrique Vicente
Departamento de Química, Escola de Ciências e Tecnologia
Centro de Química de Évora,
Universidade de Évora, Évora, Portugal
Centro Algoritmi, Universidade do Minho, Braga, Portugal
E-mail: hvicente@uevora.pt

M. Rosário Martins
Departamento de Química, Escola de Ciências e Tecnologia
Laboratório HERCULES
Universidade de Évora
Évora, Portugal
E-mail: mrm@uevora.pt

Inês Barrucho, Miguel Santos, Rafaela Fernandes, João Costa
Departamento de Química, Escola de Ciências e Tecnologia
Universidade de Évora
Évora, Portugal
E-mail: {ines_baruxo, miguel_galego77, rafa_ela.cf}@hotmail.com, joaommrc@gmail.com

Jorge Ribeiro
Escola Superior de Tecnologia e Gestão
ARC4DigiT – Applied Research Center for Digital Transformation – Instituto Politécnico de Viana do Castelo
Viana do Castelo, Portugal
E-mail: jribeiro@estg.ipvc.pt

José Neves
Centro Algoritmi
Universidade do Minho
Braga
Portugal
E-mail: jneves@di.uminho.pt

KEYWORDS
Artificial Intelligence, Olive Oil Screening, Sensorial Analysis, Artificial Neural Networks, Incomplete Information.

ABSTRACT

On the one hand, the olive oils' quality is fixed by the region in which it is produced, the olive variety, the year of production, the degree of maturation, the extraction and the preservation processes, i.e., the olive oils' quality is assessed by two types of analyzes, the organoleptic and the biochemical ones. On the other hand, there are no studies linking analytical parameters to sensory data due to the nonlinearity that exists between these two types of variables. The present study aims to provide an answer to these problems by modeling the causal processes used at different types of olive and olive oil production. To this end, we called at an Artificial Neural Network approach to problem solving once it provides a way to handle the various stages of ripeness of the olives and olive oil production, by allowing one to deal with incomplete, unknown or even self-contradictory information or knowledge.

INTRODUCTION

On the one hand, the World Health Organization points to the relevant nutritional outcomes for the prevention of chronic diseases. On the other hand, several studies lead to a possible correlation between the Mediterranean diet and the lower incidence of some pathologies such as cardiovascular disease, obesity, type 2 diabetes, cancer, rheumatoid arthritis (Aparicio-Soto et al. 2016). It is clear that virgin olive oil is an essential ingredient and the main source of dietary lipids in the Mediterranean Diet. Olive oil should have a high nutritional quality and a unique composition with a high content of monounsaturated fatty acids and antioxidants (e.g., phenolic compounds, carotenoids or vitamin E), especially in the case of Extra Virgin Olive Oil (EVOO). In fact, a diet high in monounsaturated fatty acids provides sufficient fluidity for the biological membranes, which reduces lipid peroxidation exposure. In addition, the antioxidant components of olive oil are able to trap free radicals and protect lipid peroxidation and thereby reduce the development of atherosclerosis. In terms of the heart, olive oil lowers plasma LDL cholesterol levels and increases HDL cholesterol levels, reducing the risk of heart disease (Alarcón-de-la-Lastra et al. 2001). The olive oil's quality will be related to the region in which it is produced, to the variety of the olive tree, the year of production, the maturation of the fruits, the extraction processing and the preservation. Indeed, the quality of an olive oil is determined taking into account several physical-chemical parameters (e.g., the determination of free fatty acids (acidity expressed as a percentage of oleic acid) and the organoleptic properties evaluated during the sensory analysis. The acceptable levels of these parameters in olive oil are regulated by the European Union (EU Regulation 1348/2013), i.e., when forecasting the quality of virgin olive oil, account should be taken of the geographical area in which the olive tree is planted, agronomic parameters, weather, physicochemical parameters, organoleptic parameters and sensory analysis (Dias et al. 2016). The agronomic factors should also be taken into account, such as soil type, pH, olive variety, olive color, fruit ripening, fruit weight and presence (or absence) of infestation, such as olive fly, which increases the levels of free fatty acids and peroxides, contributing to a reduction in olive oil quality (Pereira et al. 2004, Piscopo et al. 2016).
Among physical-chemical parameters the percentage of free fatty acids and the rancidity or detention be must considered

(Reboredo-Rodriguez et al. 2016). European Union provides that free acid content of olive oil must be less than or equal to 0.8% for extra virgin olive oils, less than 2% for virgin olive oils and less than 3% for ordinary olive oils (EU Regulation 1348/2013). Regarding to the peroxide content, it should be less than 20 $mEqO_2/kg$ for extra virgin and virgin oils (Reboredo-Rodriguez et al. 2016). The content of esterified unsaturated fatty acids is a very important contribution to the quality of olive oil and responsible for the beneficial properties of it. The concentration depends on the production region, the climate, the variety and the degree of ripeness of the olives (Jiménez et al. 2013, Borges et al. 2017). Another parameter to consider in the quality and beneficial effects of olive oil is the amount of tocopherols (vitamin E) present, which are the most important fat-soluble antioxidants in olive oil and increase oxidative stability during storage (Reboredo-Rodriguez et al. 2016). The degradation products of chlorophyll can be used as markers for the quality and traceability of virgin olive oils. The obtained data can be used to determine whether a virgin olive oil has been properly stored or whether the virgin olive oil has been cut with refined oils (Borges et al., 2017).

The sensory analysis included tasting events performed by a trained panel involving sensory difference tests and descriptive analysis attributes. The measurement of minor flavor variations is often separated into two groups, i. e. positive and negative attributes. The first contains attributes such as fruity, bitter and spicy, while the other contains sediment/mold or acid rancidity (Jiménez et al. 2013). According to the European Union (EU Regulation 1348/2013) olive oil is classified taking into account to the median of negative attributes and the median of the fruity and others positive attributes (e.g., the extra virgin olive oil, has a median negative attribute equal to 0 and a median of positive attributes greater than 0). To overcome these problems, this work presents a computerized tool that uses a *Logic Programming* based approach to knowledge representation and reasoning (Neves 1984, Neves et al. 2007), complemented by a computerized framework based on *Artificial Neural Networks* (*ANNs*). *ANNs* were inspired by the architecture and internal features of the human brain and nervous system. *ANNs* can be described as a connected structure of basic computing units (artificial neurons or nodes) with learning capabilities. The neural multi-layer feed-forward network architecture is one of the most popular *ANN* structures. It consists of two or more layers of artificial neurons, including an input layer, an output layer and a number of hidden layers with a certain number of active neurons connected by modifiable weights. The number of nodes in the input level determines the number of independent variables, and the number of nodes in the output level indicates the number of dependent variables (Haykin 2009). Several studies have shown how *ANNs* can successfully be used to model data and capture complex relationships between inputs and outputs (Haykin 2009, Vicente et al. 2012).

This paper is divided into four sections, the first of which is an opening to publicize the problem to be solved, followed by a background in which related issues are the subject of attention. The third section introduces the case study and the solution to the problem using *ANNs*. Finally, a conclusion is presented and guidelines for future work are outlined.

KNOWLEDGE REPRESENTATION AND REASONING

Many approaches to *Knowledge Representation and Reasoning* (*KRR*) have been proposed using the *Logic Programming* (*LP*) epitome, namely in the area of *Model Theory* (Kakas et al. 1998; Pereira and Han 2009), and *Proof Theory* (Neves 1984, Neves et al. 2007). In the present work the *Proof Theoretical Approach* in terms of an extension to the *LP* language is followed. A *LP* under this setting is a finite set of clauses in the form, viz.

$$
\begin{aligned}
&\{\\
&\quad p \leftarrow p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m\\
&\quad ?\ (p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m)\ \ (n, m \geq 0)\\
&\quad exception_{p_1}\\
&\quad \cdots\\
&\quad exception_{p_j}\ \ (0 \leq j \leq k),\ being\ k\ an\ integer\ number\\
&\} :: scoring_{value}
\end{aligned}
$$

where "?" is a domain atom denoting falsity, the p_i, q_j, and p are classical ground literals, i.e., either positive atoms or atoms preceded by the classical negation sign ¬ (Neves 1984). Indeed, ¬ stands for a strong declaration that speaks for itself, and not denotes negation-by-failure, or in other words, a flop in proving a given statement, once it was not declared explicitly. Under this formalism, every program is associated with a set of abducibles (Kakas et al. 1998; Pereira and Han 2009), given here in the form of exceptions to the extensions of the predicates that make the program, viz.

$$
\begin{aligned}
&exception_{p_1}\\
&\cdots\\
&exception_{p_j}\ \ (0 \leq j \leq k),\ being\ k\ an\ integer\ number
\end{aligned}
$$

that stand for information or knowledge that cannot be ruled out. On the other hand, clauses of the type, viz.

$$?\ (p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m)\ \ (n, m \geq 0)$$

also named invariants or restrictions to complain with the universe of discourse, set the context under which it may be understood. The term $scoring_{value}$ stands for the relative weight of the extension of a specific predicate with respect to the extensions of its peers ones that make the inclusive or global program. Indeed, *KRR* aims at the understanding of the information's complexity and the associated inference mechanisms. Automated reasoning capabilities enables a system to fill in the blanks when one is dealing with incomplete information, where data gaps are common, i.e., the fundamentals and the attributes of the logical functions go from discrete to continuous, allowing for the representation or handling of unknown, vague, or even self-contradictory information/knowledge. In this work a data item is to be understood as find something smaller inside when taking anything apart, i.e., it is mostly formed from different elements, namely the Interval Ends where their values may be situated, the *Quality-of-Information* (*QoI*) they carry, and the *Degree-of-Confidence* (*DoC*) put on the fact that their values are inside the intervals just referred to above. These are just three of over an endless element's number. Undeniably,

one can make virtually anything one may think of by joining different elements together or, in other words, viz.

- What happens when one splits a data item? The broken pieces become data item for another element, a process that may be endless; and
- Can a data item be broken down? Basically, it is the smallest possible part of an element that still remains the element.

Therefore, the proposed approach to this issue, put in terms of the logical programs that elicit the universe of discourse, will be set as productions of the type, viz.

$$predicate_{1\le i\le n} - \bigcap_{1\le j\le m} clause_j\Big(\Big([A_{x_1}, B_{x_1}](QoI_{x_1}, DoC_{x_1})\Big), \cdots \cdots, \Big([A_{x_m}, B_{x_m}](QoI_{x_m}, DoC_{x_m})\Big)\Big) :: QoI_j :: DoC_j$$

n, $\cap$, m and A_{x_m}, B_{x_m} stand for the cardinality of the predicates' set, conjunction, predicate's extension, and the Interval Ends where the predicates attributes values may be situated, respectively. The metrics $[A_{x_m}, B_{x_m}]$, QoI and DoC show the way to data item dissection (Fernandes et al. 2015; Silva et al. 2016), i.e., a data item is to be understood as the data's atomic structure. It consists of identifying not only all the sub items that are thought to make up an data item, but also to investigate the rules that oversee them, i.e., how $[A_{x_m}, B_{x_m}]$, QoI_{x_m}, and DoC_{x_m} are kept together and how much added value is created.

CASE STUDY

This study focused on 33 olive oil samples from several regions from five Mediterranean countries, namely Portugal (Beira Baixa and Alentejo), Spain (Andalusia), Italy (Sicily), Greece and Turkey produced with autochthonous varieties of these countries, such as Arbequina, Cobrançosa, Galega Vulgar, Madural, Picual and Verdeal cultivars (Pereira et al. 2004; Jiménez et al. 2013; Peres et al. 2013; Valli et al. 2014; Reboredo-Rodríguez et al. 2015;Köseoglu et al. 2016; Piscopo et al. 2016; Borges et al. 2017). Thus, in order to develop an intelligent system for olive oil quality screening a knowledge database was build up. Such database is provided in the form of extensions of the relations (or tables) illustrated in Figure 1 and merge information about analytical parameters and sensorial data. In Figure 1, the *Fruit Ripening* and *Biological Attacks* columns from *Olive Oil Quality* table are filled with zero and one denoting *Early Maturing*/*No* and *Full Maturation*/*Yes*, respectively. In *Production Process* table the value zero is assigned to *Hot Extraction* and *Refining Existence*, while the value one stands for *Cold Extraction* and *Refining Absence*.

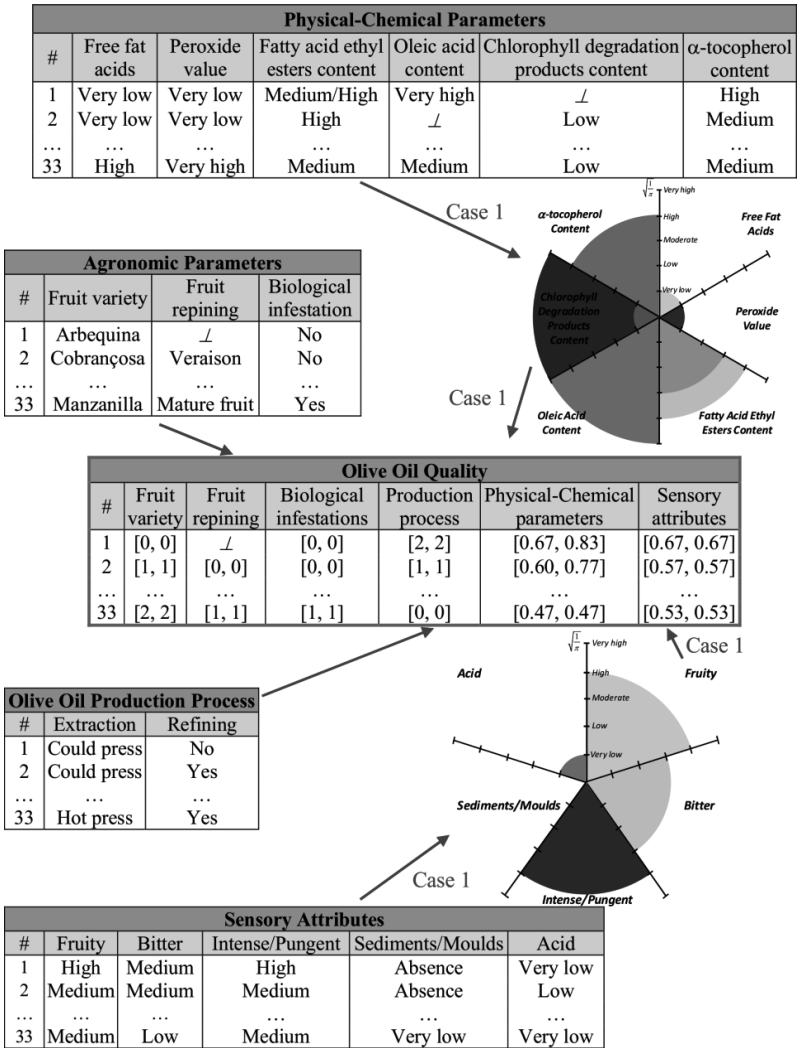


Figure 1: A Fragment of the Olive Oil's Quality Knowledge Base

It should be noted that high values of the fatty acid ethyl esters, oleic acid and the α-tocopherol contents contribute to increase the quality of olive oil. Conversely, high levels of free fatty acids, peroxides and chlorophyll degradation products lead to a decrease in quality. Regarding sensory analysis, the parameters fruity, bitter and intense/ /pungent have a positive contribution to olive oil quality while the contribute of the sediments/moulds and acid is negative (Reboredo-Rodriguez et al. 2016). In order to consider these opposing tendencies, the procedure suggested by Vicente et al. (2017) was adopted. Thus, the contribution of the items with negative impact on the subject in analysis are set as $1/n$ minus the correspondent area, where n denotes the number of issues regarding the subject in study (i.e., 6 and 5 for physical-chemical parameters and sensory attributes, respectively).

Although the bulk of the data is known, there are some items that are unknown, be in a set or belong to an interval (e.g., in case 1 the *Fruit Repining* is unknown, represented by the symbol $\perp$, while the *Physical-Chemical Parameters* ranges in the interval [0.67, 0.83]). The extensions of the relations just referred to above make the knowledge base of one's universe of discourse (Figure 1). Thus, the objective function mapping the system behavior can now be set in terms of the extension of the predicate olive oil quality (oo_{qual}), viz.

$$oo_{qual}: F_{ruit}V_{ariety}, F_{ruit}R_{epining}, B_{iological}I_{nfestations}, P_{roduction}P_{rocess}, P_{hysical}C_{hemical}P_{arameters}, S_{ensory}A_{ttributes} \rightarrow \{0, 1\}$$

where 0 and 1 denote *true* and *false*, respectively. Its extension, with respect to an olive oil sample characterized by feature vector $F_{ruit}V_{ariety} = 1$, $F_{ruit}R_{epining} = \perp$, $B_{iological}I_{nfestations} = 1$, $P_{roduction}\ P_{rocess} = 1$, $P_{hysical}\ C_{hemical}\ P_{arameters} = [0.47, 0.67]$, $S_{ensory}\ A_{ttributes} = [0.53, 0.67]$ is now given in the form, viz (Fernandes et al. 2015).

$$\begin{aligned}
&\{\\
&\neg\, oo_{qual}\Big(\big([A_{FV}, B_{FV}](QoI_{FV}, DoC_{FV})\big),\ \big([A_{FR}, B_{FR}](QoI_{FR}, DoC_{FR})\big),\ \cdots,\ \big([A_{SA}, B_{SA}](QoI_{SA}, DoC_{SA})\big)\Big)\\
&\qquad\qquad \leftarrow not\ oo_{qual}\Big(\big([A_{FV}, B_{FV}](QoI_{FV}, DoC_{FV})\big),\ \big([A_{FR}, B_{FR}](QoI_{FR}, DoC_{FR})\big),\ \cdots,\ \big([A_{SA}, B_{SA}](QoI_{SA}, DoC_{SA})\big)\Big)\\
&oo_{qual}\underbrace{\Big(\big([1,1](1_{[1,1]},\ DoC_{[1,1]})\big), \big([0,\ 1](1_{[0,1]},\ DoC_{[0,1]})\big), \cdots, \big([0.53,\ 0.67](1_{[0.53,\ 0.67]},\ DoC_{[0.53,\ 0.67]})\big)\Big)}_{attribute\grave{}s\ values\ ranges} :: 1 :: DoC\\
&\qquad\underbrace{[0,24] \qquad\qquad [0,1] \qquad\qquad \cdots \qquad [0,1]}_{attribute\grave{}s\ domains}\\
&\} :: 1
\end{aligned}$$

Program 1: Feature Vector's Decoding

whose attribute's values may now be set to the interval 0…1 (Program 2), viz.

$$\begin{aligned}
&\{\\
&\neg\, oo_{qual}\Big(\big([A_{FV}, B_{FV}](QoI_{FV}, DoC_{FV})\big),\ \big([A_{FR}, B_{FR}](QoI_{FR}, DoC_{FR})\big),\ \cdots,\ \big([A_{SA}, B_{SA}](QoI_{SA}, DoC_{SA})\big)\Big)\\
&\qquad\qquad \leftarrow not\ oo_{qual}\Big(\big([A_{FV}, B_{FV}](QoI_{FV}, DoC_{FV})\big),\ \big([A_{FR}, B_{FR}](QoI_{FR}, DoC_{FR})\big),\ \cdots,\ \big([A_{SA}, B_{SA}](QoI_{SA}, DoC_{SA})\big)\Big)\\
&oo_{qual}\underbrace{\Big(\big([0.04,\ 0.04](1_{[0.04,0.04]}, 1_{[0.04,0.04]})\big), \big([0,\ 1](1_{[0,1]}, 0_{[0,1]})\big), \cdots, \big([0.53,\ 0.67](1_{[0.53,0.67]}, 0.99_{[0.53,0.67]})\big)\Big)}_{attribute\grave{}s\ values\ ranges\ once\ normalized\ and\ respective\ QoI\ and\ DoC\ values} :: 1 :: 0.83\\
&\qquad\underbrace{[0,1] \qquad\qquad [0,1] \qquad\qquad \cdots \qquad [0,1]}_{attribute\grave{}s\ domains\ once\ normalized}\\
&\} :: 1
\end{aligned}$$

Program 2: A Feature Vector's Logical Form

With regard to the computational model, ANNs were selected for their dynamic properties such as adaptability, robustness and flexibility. Figure 2 shows how the normalized values of the *Interval Ends* and the respective *DoC* and *QoI* values operate as inputs to the *ANN*. The result is the evaluation of the quality of the olive oil as well as a measure of the trust that can be expected from such a prediction. The *ANN* topology consists of an input layer with 6 nodes (corresponding to 24 nodes in the preprocessing), a hidden layer with 5 nodes and a 2-node output layer.

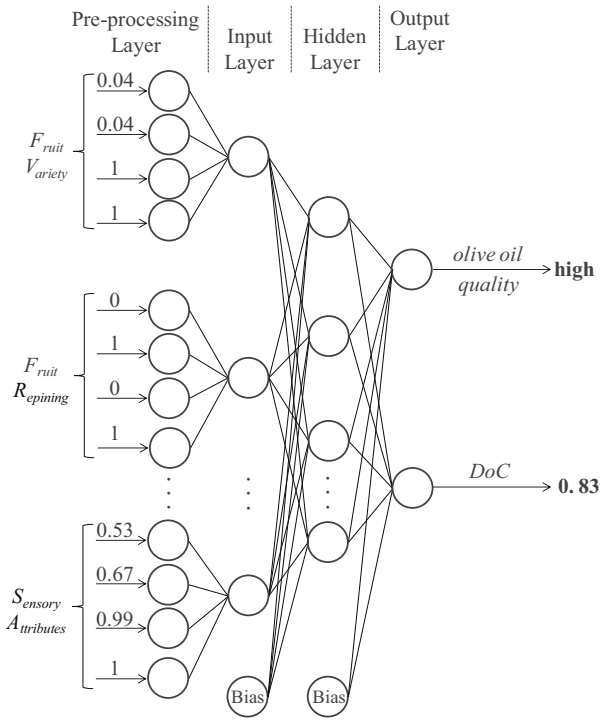


Figure 2: ANN's Topology

A set with 33 records were used to test the model. The dataset was divided in exclusive subsets through the six-folds cross validation (Haykin 2009). In order to guarantee the statistical significance of the attained results 20 (twenty) experiments were applied in all tests. The back-propagation algorithm was used in the learning process of the ANN. As the activation function in the pre-processing layer it was used the identity one, while in the other layers was passed down the sigmoid. The model accuracy was 81.8% (i.e., 27 instances correctly classified in 33). Such results seem to suggest that the ANN model exhibits a good performance in assessing olive oil's quality.

CONCLUSIONS

This work presents an Artificial Intelligence based system for assessing olive oil quality, being the information acquired from physical-chemical parameters, sensory analysis and agronomic factors. It is grounded on a formal framework based on *Logic Programming* for *Knowledge Representation and Reasoning*, complemented by an *Artificial Neural Network* stance to computing. In fact, this is one of the added values resulting from the complementary between *Logic Programming* and *ANNs* remarkable information processing capabilities, such as nonlinearity, high parallelism, robustness, error and fault tolerance, learning ability, generalization, and the possibility of handling incomplete information or knowledge.

ACKNOWLEDGEMENTS

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – *Fundação para a Ciência e Tecnologia* within the Project Scope: UID/CEC/00319/2013.

REFERENCES

Alarcón-de-la-Lastra, C.; M.D. Barranco; V. Motilva; and J.M. Herrerías. 2001. "Mediterranean Diet and Health: Biological Importance of Olive Oil." *Current Pharmaceutical Design* 7, No.10, 933-950.

Aparicio-Soto, M.; M. Sánchez-Hidalgo; M.Á. Rosillo; M.L. Castejón; and C. Alarcón-de-la-Lastra. 2016. "Extra virgin olive oil: a key functional food for prevention of immune-inflammatory diseases." *Food & Function* 7, No.11, 4492-4505.

Borges, T.H.; J.A. Pereira; C. Cabrera-Vique; L. Lara; A.F. Oliveira; and S. Seiquer. 2017. "Characterization of Arbequina virgin olive oils produced in different regions of Brazil and Spain: Physicochemical properties, oxidative stability and fatty acid profile." *Food Chemistry* 215, No.11, 454–462.

Dias, L.G.; N. Rodrigues; A.C.A. Veloso; J.A. Pereira; and A.M. Peres. 2016. "Monovarietal extra‐virgin olive oil classification: a fusion of human sensory attributes and an electronic tongue." *European Food Research and Technology* 242, No.2, 259-270.

EU Regulation No.1348/2013. Commission Implementing Regulation of 16 December 2013. Amending Commission Regulation (EEC) No.2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. *Official Journal of the European Union*, L338/31-L338/67.

Fernandes, F.; H. Vicente; A. Abelha; J. Machado; P. Novais; and J. Neves. 2015. "Artificial Neural Networks in Diabetes Control". In *Proceedings of the 2015 Science and Information Conference* (London, UK, Jul. 28-30). IEEE, Picataway, N.J., 362-370.

Haykin, S. 2009. *Neural Networks and Learning Machines*. Pearson Education, Upper Saddle River, N.J.

Jiménez, B.; A. Sánchez-Ortiz; M. L. Lorenzo; and A. Rivas. 2013. "Influence of fruit ripening on agronomic parameters, quality indices, sensory attributes and phenolic compounds of Picudo olive oils." *Food Research International* 54, No.2, 1860-1867.

Kakas, A.; R. Kowalski; F. Toni. 1998. "The role of abduction in logic programming". In *Handbook of Logic in Artificial Intelligence and Logic Programming*. Vol. 5, D. Gabbay; C. Hogger and I. Robinson (Eds.). Oxford University Press, Oxford, 235-324.

Köseoglu, O.; D. Sevim; and P. Kadiroglu. 2016. "Quality characteristics and antioxidant properties of Turkish monovarietal olive oils regarding stages of olive ripening." *Food Chemistry* 212, 628-634.

Neves, J. 1984. "A logic interpreter to handle time and negation in logic databases". In Proceedings of the 1984 Annual Conference of the ACM on the 5th Generation Challenge (San Francisco, CA, Oct. 8-10). ACM, New York. N.Y., 50-54.

Neves, J.; J. Machado; C. Analide; A. Abelha; and L. Brito. 2007. "The halt condition in genetic programming." *Lecture Notes in Artificial Intelligence* 4874, 160-169.

Pereira, J.A.; M.R. Alves; S. Casal; and M.B.P.P. Oliveira. 2004. "Effect of olive fruit fly infestation on the quality of olive oil from cultivars Cobrançosa, Madural and Verdeal Transmontana." *Italian Journal of Food Science* 16, No.3, 355-365.

Pereira, L. and H. Anh. 2009. "*Evolution prospection*." *Studies in Computational Intelligence* 199, 51-64.

Peres, F.; H.H. Jeleń; M.M. Majcher; M. Arraias; L.L. Martins; and S. Ferreira-Dias. 2013. "Characterization of aroma compounds in Portuguese extra virgin olive oils from Galega Vulgar and Cobrançosa cultivars using GC-O and GC×GC-ToFMS." *Food Research International* 54, No.2, 1979-1986.

Piscopo, A.; A.D. Bruno; A. Zappia; C. Ventre; and M. Poiana. 2016. "Characterization of monovarietal olive oils obtained from mills of Calabria region (Southern Italy)." *Food Chemistry* 213, 313-318.

Reboredo-Rodríguez, P.; C. González-Barreiro; B. Cancho-Grande; E. Valli; A. Bendini; T.G. Toschi; and J. Simal-Gandara. 2016. "Characterization of virgin olive oils produced with autochthonous Galician varieties." *Food Chemistry* 212, 162-171.

Reboredo-Rodríguez, P.; C. González-Barreiro; B. Cancho-Grande; G. Fregapane; M.D. Salvador; and J. Simal-Gandara. 2015. "Characterisation of extra virgin olive oils from Galician autochthonous varieties and their co-crushings with Arbequina and Picual cv." *Food Chemistry*, 176, 493-503.

Silva, A.; H. Vicente; A. Abelha; M.F. Santos; J. Machado; J. Neves; and J. Neves. 2016. "Length of Stay in Intensive Care Units – A Case Base Evaluation." *Frontiers in Artificial Intelligence and Applications* 286, 191-202.

Valli, E.; A. Bendini; M. Popp; and A. Bongartz. 2014. "Sensory analysis and consumer acceptance of 140 high-quality extra virgin olive oils." *Journal of the Science of Food and Agriculture* 94, No.10, 2124-2132.

Vicente, H.; M.R. Martins; M. Duarte; P. Miguel; J.L.Grañeda; F. Caldeira; J. Vilhena; J. Neves; and J. Neves. 2017. "A Case Base View of Heart Failure Predisposition Risk." *Advances in Intelligent Systems and Computing* 571, 312-323.

Vicente, H.; S. Dias; A. Fernandes; A. Abelha; J. Machado; and J. Neves. 2012. "Prediction of the Quality of Public Water Supply using Artificial Neural Networks." *Journal of Water Supply: Research and Technology – AQUA* 61, 446-459.

AUTHOR BIOGRAPHIES

HENRIQUE VICENTE was born in S. Martinho do Porto, Portugal and went to the University of Lisbon, where he studied Chemistry and obtained his degrees in 1988. He joined the University of Évora in 1989 and received his Ph.D. in Chemistry in 2005. He is now Auxiliary Professor at the Department of Chemistry at the University of Évora. He is a researcher at the Évora Chemistry Center and his current interests include Water Quality Control, Lakes and Reservoirs Management, Data Mining, Knowledge Discovery from Databases, Knowledge Representation and Reasoning Systems, Evolutionary Intelligence and Intelligent Information Systems.

M. ROSÁRIO MARTINS was born in Évora, Portugal. She was obtained her degree in Pharmaceutical Sciences at the Faculty of Pharmacy of the University of Lisbon, in 1990. She works at University of Évora since 1991 and received her Ph.D. in Chemistry in 2004. She is Assistant Professor at the Department of Chemistry at the University of Évora, since 2004. she is a researcher at Laboratory HERCULES (University of Évora) and her main research scientific areas include Clinical Biochemistry, Toxicology and Biotechnology.

MIGUEL SANTOS was born in Lausanne, Switzerland. He was obtained her degree in Biotechnology at University of Évora, in 2015. He participated in some fungus studies at the Laboratory HERCULES – University of Évora. His current interests are in the area of agriculture, olive growing and oenology.

RAFAELA FERNANDES was born in Santo Isidoro, Mafra. She obtained her degree in Biotechnology at the School of Sciences and Technology of the University of Évora in 2015. At the same year she did a curricular internship entitled "New Approaches for the study of microbial communities present in Cultural Heritage" at Laboratory HERCULES (University of Évora). She is currently the production manager in the field of organic cosmetics at Bio4Natural.

JOÃO COSTA was born in Cumeira, Caldas da Rainha, in 1993. He took a bachelor degree on Biotechnology in the University of Évora. During those time he worked with Mycorriza Fungi – Genetic Analysis.

JORGE RIBEIRO is Adjunct Professor in the School of Technology and Management Polytechnic of the Institute of Viana do Castelo (IPVC) – Portugal, since 2006. He holds a PhD (Informatics Engineering) in 2011 at the Department of Electronic and Computation of the University of Santiago de Compostela, Spain. He is affiliated with the Algoritmi – Research and Development Unit of the School of Engineering – University of Minho. His main research interests are Artificial Intelligence, Software Engineering and Knowledge Discovery in Databases and he has been author and co-author of some papers in the field of Data Mining, Software Engineering, Knowledge Representation, Evolutionary Systems and DataMining using Geographic Information Systems.

JOSÉ NEVES is Full Professor of Computer Science at Minho University, Portugal, since 1983. José Neves is the Deputy Director of the Division for Artificial Intelligence. He received his Ph.D. in Computer Science from Heriot Watt University, Scotland, in 1983. His current research interests relate to the areas of Knowledge Representation and Reasoning, Evolutionary Intelligence, Machine Learning, Soft Computing, aiming to construct dynamic virtual worlds of complex symbolic entities that compete against one another in which fitness is judged by one criterion alone, intelligence, here measured in terms of a process of quantification of the quality of their knowledge, leading to the development of new procedures and other metaheuristics and their application in complex tasks of optimization and model inference in distinct areas, namely in the Healthcare Arena and The Law.

COLLABORATIVE SYSTEMS MANAGEMENT

CLOUD COMPUTING ENVIRONMENTS FOR SIMULATION OF ADAPTABLE STANDARDIZED WORK AND ELECTRONIC WORK INSTRUCTIONS IN INDUSTRY 4.0

Fábio Silva, Ricardo Martins, Marco Gomes, António Silva, José Machado, Paulo Novais and Cesar Analide
Algoritmi Centre/Department of Informatics
University of Minho
Braga
Portugal
E-mail: {fabiosilva, marcogomes, jmac, pjon, analide}@di.uminho.pt
{ricardo.martins, asilva}@algoritmi.uminho.pt

KEYWORDS

Industry 4.0, Smart Factory, Electronic work instructions, Standard Work.

ABSTRACT

Industry 4.0 is a term connotated with the development of software services able to percept and act in industrial environments. A common characteristic is the availability of data and information in the format of streams which are available for building new industrial processes. In this paper, an architecture of an application able to process different inputs in the creation of electronic work instruction and the optimisation of standard work is presented. Taking into consideration the principles of industry 4.0, the information is stored in a web application which manages user access, content creation and update, review process and optimisation procedures.

INTRODUCTION

The Industry 4.0 represents the upcoming industrial revolution that will set a new context for future factories environments. It will fuel the right conditions, in which, human abilities can be amplified as machines help to process, analyse, and evaluate the abundance of data that will shape the modern industries (Brettel, Friederichsen, Keller, & Rosenberg, 2014). Its consequences have been widely reported and will be focused on transforming supply chains, business models, and business processes significantly. Meanwhile, many current manufacturing systems are not yet, ready to manage or overcome technical issues and challenges that are posed by this radical shift towards a new industrial paradigm. Indeed, many companies and organisations are exposed to a dilemma: to wait for a more mature Industry 4.0 implementation or to start an early adaptation but risk to commit fatal errors (Schmidt et al., 2015). One way to help solve this dilemma is to invest in scientific research of potential use in Industry 4.0 and implicit technologies and their effects in traditional models and processes of manufacturing.

The work presented here tries to answer these issues by exploring how using cutting-edge technologies can optimise manufacturing processes, allowing the reallocation of more time of human resources in high-level thinking, creativity, and decision-making.

It is implicit in this work that updating production processes by combining disruptive concepts such as Cyber-Physical Systems, Internet of Things, Big Data and Cloud Computing and other cutting-edge technologies will meet the challenge of improving existing methodologies and procedures, to increase competitiveness and productivity through systematic innovation.

The object of study of this work is focused on two critical aspects of the planning process in a factory: Work Instructions (WI), and Standard Work (SW). These are essential tools that support production and play pivotal roles in process engineering. On the one hand, WI provides information about the sequence of operations needed to perform a task correctly. On the other hand, SW allows an active line balancing, allocating tasks to operators and workstations and yield relevant information for decision-making in the manufacturing process. It is important to point out that in many factories, these tools are manually performed, and the knowledge regarding WI, production quantities and resources available is not articulated in real-time during the implementation of SW. Therefore, the objective is clear: designing a technological strategy to improve the processes of elaboration of Work Instructions and Standard Work (SW), according to the principles of manufacturing methodologies and the context of the Industry 4.0 initiative.

Thus, this paper aims to provide the description and analysis of a system designed and developed under Industry 4.0 umbrella to support and simulate the automatic WI creation with real-time updates, as well as, the integration of models and algorithms to assign tasks to workstations and SW planning according to production needs.

STATE OF THE ART

Industry 4.0

According to Kolberg and Zühlke (Kolberg & Zühlke, 2015), the term industry 4.0 has been created in Germany in 2014 and refers to an industry network-based approach, where components and machines are getting intelligent and intra-connected by communication infrastructure and services.

The concept is related to the idea of the smart factory, which uses resources from information technology. Carrió (Carrió, 2015) claims that this is a new phase in the industry,

beginning with the advancement of technology and the innovations brought about by it, namely:

- Increasing the volume of data and information within organisations;
- Increased computer processing and storage capacity;
- Cloud information storage capacity;
- Improvement in the Man/Machine relationship that began with robots and can now be exemplified with 3D printers.

Industry 4.0 has revolutionised how companies operate. However, there are some adverse effects, such as job creation since this context requires a greater demand for skilled employees, according to Carrió (Schmidt et al., 2015).

In fact, the changes brought about by the advancement of technology are irreversible, and it is up to companies to adapt to this new phase, making their processes increasingly efficient.

Industry 4.0

The term cloud manufacturing, as defined for Li et al (Li, Zhang, & Chai, 2010), can be expressed as a service-oriented, knowledge-based smart manufacturing system with high efficiency and low energy consumption.

Thus, according to Li (Li, Zhang, Wang, et al., 2010) and emphasized for Xu (Xu, 2012), cloud manufacturing is related to an evolution of the manufacturing model based on usage of cloud computing, Internet of Things (IoT), virtualization, and service-oriented technologies to transform the manufacturing resources into services that can be consumed across different production levels. It covers the entire product life cycle supply chain and is widely considered as a parallel, networked, and intelligent manufacturing system (the manufacturing cloud). An example of cloud manufacturing proposition is presented in Figure 1

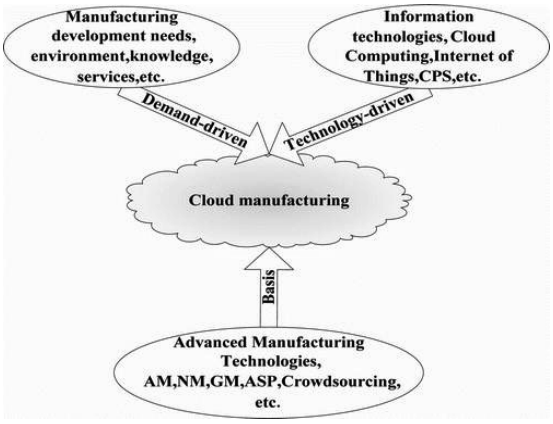


Figure 1: Cloud Manufacturing (Zhang et al., 2014).

Thus, according to Zhang (Zhang et al., 2014), on-demand manufacturing services are available to all types of end-users through the manufacturing cloud.

In cloud manufacturing, different technologies can coexist intelligently and connected to the cloud, providing services that can be categorised and aggregated.

For example, for Atzori (Atzori, Iera, & Morabito, 2010), IoT provides solutions such as RFID tags, sensor and actuator networks, embedded systems and intelligence in smart objects to collect information, supported by high-performance computing solutions and existing enterprise systems, virtual servers and service-oriented architecture (SOA) that can allocate resources on-demand, as mentioned by Vecchiola (Vecchiola, Pandey, & Buyya, 2009).

Decision Optimization

According to definition available in ESourcingWiki.com (Lamoureux, 2008), decision optimisation is the application of rigorous analytical techniques to a well-defined scenario to arrive at the absolute best decision out of a multitude of possible alternatives in a rigorous, repeatable, and provable fashion.

Decision Optimization has been used to find the best solution by exploring all the possibilities for various usage scenarios. According to (Shao, Brodsky, & Miller, 2015), a feasible solution is an instantiation of values from corresponding domains of decision variables that satisfy all the constraints. Among all feasible solutions, an optimal solution is one that makes the objective minimal or maximal, as required. Using decision optimisation to analyse the performance of a manufacturing process is challenging because the model abstractions only have indirect connections to the elements of a manufacturing process.

There are many technologies and tools for optimisation available in the literature. In general, identifying and using them properly is a difficult task because users need have to some previous knowledge and skills or be trained for using these tools. Also, end users do not have an intuitive tool which enables them to use problem-solving techniques while modelling problems and designing operations at the same time.

Some examples of optimisation tools are OptQuest for ARENA (Inc, 2018), AnyLogic (OptTek Systems, 2018), SIMUL8 (Corporation, 2018) and CPLEX optimizer (IBM, 2018). These optimisation tools use search strategies such as genetic algorithms, neural networks, scatter search, tabu search, and simulated annealing. However, for (Klemmt, Horn, Weigert, & Wolter, 2009), optimisation problems that can be expressed in the closed analytical form, simulation-based optimisation is not as efficient as MP/CP solver-based optimisation concerning optimality of results and time complexity.

ARCHITECTURE FOR PRODUCTION AND CONTROL PLANNING

The system described herein follows a client-server architecture that aims to improve the processes of elaboration of Work Instructions (WI) and Standard Work (SW), according to the Industry 4.0 initiative. The objective is to provide decision support for the managers, using new tools and techniques that aim to increase the company efficiency.

The systems developed, include a web application with the MVC software pattern, security, authorisation, relational entities, web services and optimisation models based on linear programming and custom heuristics. These components are required to connect to each other as the operation of the application is dependent on this. The web

server acts as a data repository and controls the process of information collection and distribution.

As shown in

Figure 2, the system consists of a *Core Server* that is the central component of the architecture, which is responsible for handling requests from each client store and retrieve information. The application core server is responsible for the work related to information management and delivery to and from databases, and its visual interfaces representation. The server task list provides encompass, security layer, the transformation of data and information, intelligent services, autocompleting, search and retrieve, role-based authorisation and management of standard work and work instructions. As depicted it consists of four components: the *Authentication Agent*, the *Standard Work Handler*, the *Database Handler* and the *SW and WI Engine*.

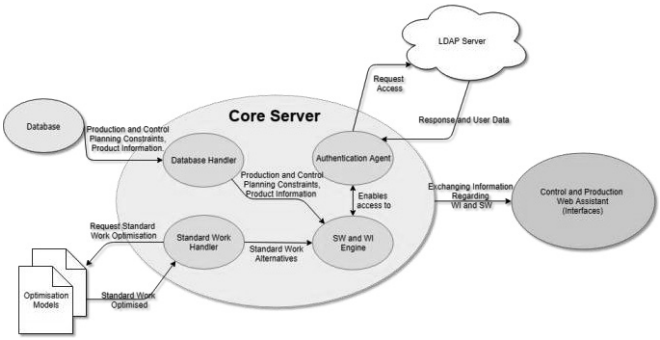


Figure 2: Architectute for Production and Control Planning.

Authentication Process

The *Authentication Agent* is the component responsible for requesting authentication and authorisation of the different types of users of the system, namely administrators and operations manager for the plant. This component is responsible for obtaining the user's credentials and verifying user authentication and authorisations from an external authentication server. Moreover, within the authentication process, user data such as name, department, phone number, the thumbnail is extracted and updated in the application.

After the authentication process is completed, it is possible to manage the information about assembly lines, workstations, families of products, products, work instructions and standard works. As such, one can create, edit and delete all this information, according to the type of authorisation in the system. So, authorisation is used to control the creation and access to information of each authenticated user. Different levels of authorisation are defined by various roles to which authenticated users are compared. Thus, the information provided by the server is tailored to the needs of each specific user, so the Control and Production Web Assistant provide different views, restricting access to information for each user group.

Work instructions and Standard Work Optimisation

The system provides decision support to create WI and SWs. The solution developed ensures real-time access to updated information, easy access to accurate information and optimisation methods, integration with other company systems and decision-making support.

To support the process of creating work instructions, the Database Handler provides information, such as product details and list of assembly lines and workstations, to fill in the data produce work instructions.

After the process of creating work instructions, the *Standard Work Handler* is responsible for standard work optimisation using the optimisation models stored on the server. Furthermore, it is possible to keep different standard work optimisations based on different optimisation models according to the user's needs. This module uses information about production and control planning constraints, product details and work instructions, provided by the *Database Handler*. The work instructions provide information such as operations descriptions, operations sequence, assembly line and products assigned. With this information, the optimisation models, developed using mathematical, heuristics and constraint programming, generate a solution, the optimisation engine automatically generates an optimal standard work, considering the given model, in mathematical logic.

Control and Production Web Assistant

The *Control and Production Web Assistant* provides the views to access the information such as assembly lines, products and families, work instructions and standard works. It is also responsible for delivering access to electronic work instructions approved for each assembly line as well as the access to the team leader to view the approved standard work. The general public only has access to public and approved work instructions and standard work for informative purposes.

Additional components are also developed such as a communication framework to allow communication and report or error and suggestions to process owners across all users. As for the notification system, it is used to alert the user about new messages from other users periodically. The communication system also provides steps to fill data entry points for creating work instructions and standard works.

This component is optimised for different devices such as smartphones, tablets and personal computers. It uses responsive web design to resize, hide, shrink, enlarge, or move the content to make it fit properly on any screen.

RESULTS

The system described in this paper has been tested using data from real production environments at the testing company. For this effect, a transcript of assembly lines, workstation and fabric instructions was introduced in the platform according to the company guidelines. The tests conducted were designed in accordance with the managers from actual assembly lines and their actual work regarding the production of work instructions. For the purposes of the results presented in this paper, the electronic work instructions produced were done from scratch, simulating a new product being prepared for assembly line production.

Electronic Work Instructions

After initial learning and adjusting to the system, assembly managers were able to replicate electronic work instructions in the platform. Functionalities tested included the addition fabric instructions, quality information and multimedia such images and videos inside work instructions. The visualisation of these instructions was integrated

successfully into the assembly monitors after being approved and published o the developed platform.
The review process was able to handle new versions of the electronic work instructions and display newly published version of the work instructions in assembly lines in real time electronically. Moreover, the modification of electronic work instructions triggers visual warnings to viewers (e.g.: operators, assembly line team leaders and managers).
Multimedia was also tested, and satisfactory results were achieved when automatic multimedia display presented in assembly lines. It was possible to enlarge images and visualise short animations to clear any manufacturing instruction doubts..

Standard Work Optimisation

From the electronic work instructions created, standard work was created specifying a number of operators in the assembly line, production quantities and assigning tasks and workstations to each operator.
To do it, the system must collect data generated and managed by itself but also data retrieved from the user and relevant to the process. From the data produced and maintained by the system, some constraints and assumptions on data must be applied, namely:

- One task must be assigned to only one worker;
- The working time of each worker must be less than equal to the target cycle time;
- Workstations have a limited number of workers assigned;
- The compatibility between each workstation;
- Dependent tasks (set of tasks that must be assigned to the same worker);
- One worker can have working time greater than target cycle time;
- Distances between the workstations;
- Worker routes, collisions and precedences (Product Tree).

This data is related to production and control planning constraints, product details and work instructions and also operations descriptions, operations sequence, assembly line and products assigned. However, this information by itself is not enough to run the models and generate a solution. Some specific information is also required. The missing data is, therefore, obtained through a form displayed to the user in the optimisation interface. The input data retrieved, such the quantity to produce, is subsequently processed to provide some relevant information, such as the minimum number of workers per workstation and the maximum quantity to produce that satisfies some production conditions.
After validating optimisation data, the system is able to pass this information to optimisation models stored in the system. Both linear programming and heuristic models were tested successfully. Nevertheless, the success of the optimisation is mostly dependent on the algorithm as the system has the responsibility for invoking the models and processing results.
From the compiled results, assembly line managers demonstrated the preference for heuristic optimisation as it portrayed more feasible work distributions between operators and fewer collisions. This can also be attributed to the difficulty in expressing optimisation constraints in linear models that satisfy all requirements enumerated.

CONCLUSIONS

Industry 4.0 requires the availability of information in real time and updated. The system presented here describes an of these principles to the creation of electronic work instructions and standard work. Automatic optimisation models were also integrated to produce automatic work distributions according to the company guidelines.
The system was successfully tested in assembly lines integrating with existing infrastructure and maintaining information updated. Initial tests with the assembly line managers demonstrated the usefulness of the system and ease of the electronic process created. Notification systems, multimedia display and work distribution dashboards were also successfully demonstrated.
For future work, the platform architecture is planned to be extended with new modules which will connect to external services existing in the company. These services complement the current system while having different objectives. Examples are assembly line simulation demonstration and production control.

Acknowledgements

This work is supported by: European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project nº 002797; Funding Reference: POCI-01-0247-FEDER-002797].

REFERENCES

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, *54*(15), 2787–2805.

Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2014). How virtualization, decentralization and network building change the manufacturing landscape: An Industry 4.0 Perspective. *International Journal of Mechanical, Industrial Science and Engineering*, *8*(1), 37–44.

Carrió, R. T. (2015). Industria 4.0:¿ qué impacto tiene en la producción y el empleo? *La Revista Del Foment*, (2146), 38–41.

Corporation, S. (2018). SIMUL8 Simulation Software - for visual process simulation modeling.

IBM. (2018). CPLEX Optimizer | IBM Analytics.

Inc, O. S. (2018). Opt quest Simulation optimization.

Klemmt, A., Horn, S., Weigert, G., & Wolter, K.-J. (2009). Simulation-based optimization vs. mathematical programming: A hybrid approach for optimizing scheduling problems. *Robotics and Computer-Integrated Manufacturing*, *25*(6), 917–925.

Kolberg, D., & Zühlke, D. (2015). Lean automation enabled by industry 4.0 technologies. *IFAC-PapersOnLine*, *48*(3), 1870–1875.

Lamoureux, M. G. (2008). Sourcing Decision Optimization.

Li, B.-H., Zhang, L., & Chai, X.-D. (2010). Introduction to cloud manufacturing. *ZTE Communications*, *16*(4), 5–8.

Li, B.-H., Zhang, L., Wang, S.-L., Tao, F., Cao, J. W., Jiang, X. D., Chai, X. D. (2010). Cloud manufacturing: a new service-oriented networked manufacturing model. *Computer Integrated Manufacturing Systems*, *16*(1), 1–7.

OptTek Systems, I. (2018). Arena Simulation.

Schmidt, R., Möhring, M., Härting, R.-C., Reichstein, C., Neumaier, P., & Jozinović, P. (2015). Industry 4.0-potentials for creating smart products: empirical research results. In

International Conference on Business Information Systems (pp. 16–27). Springer.

Shao, G., Brodsky, A., & Miller, R. (2015). Modeling and optimization of manufacturing process performance using Modelica graphical representation and process analytics formalism. *Journal of Intelligent Manufacturing*, 1–15.

Vecchiola, C., Pandey, S., & Buyya, R. (2009). High-performance cloud computing: A view of scientific applications. In *Pervasive Systems, Algorithms, and Networks (ISPAN), 2009 10th International Symposium on* (pp. 4–16).

Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing*, *28*(1), 75–86.

Zhang, L., Luo, Y., Tao, F., Li, B. H., Ren, L., Zhang, X., Liu, Y. (2014). Cloud manufacturing: a new manufacturing paradigm. *Enterprise Information Systems*, *8*(2), 167–187. https://doi.org/10.1080/17517575.2012.683812

Process-oriented Inconsistency Management in Collaborative Systems Modeling

István Dávid[1,2] Joachim Denil[1,2] Hans Vangheluwe[1,2,3]

[1]University of Antwerp, Belgium
[2]Flanders Make, Belgium
[3]McGill University, Montréal, Canada
{istvan.david, joachim.denil, hans.vangheluwe}@uantwerpen.be

KEYWORDS

Concurrent engineering, model-based systems engineering, heterogeneous systems, model consistency

ABSTRACT

Engineered products have reached a complexity which requires explicit modeling and analysis of various aspects such as performance, safety, and energy-efficiency, before system realization. This allows stakeholders in various domains to collaborate on a "virtual product", using their domain-specific modeling languages and tools. This variety of modeling languages and tools, if not managed properly, can give rise to inconsistencies between stakeholder models, resulting in an incorrect product. Managing inconsistencies, therefore, is a key enabler to efficient collaborative engineering. Explicitly modeling the engineering process in which the virtual product models are manipulated allows for a detailed analysis of the root causes of inconsistencies and of the impact of applying various (in-)consistency management techniques on the process.

INTRODUCTION

The complexity of currently engineered systems has increased drastically over the last decades. To tackle this complexity, a virtual product can be designed before realizing the real product. The design of the virtual product is achieved by model-based techniques, and typically in heterogeneous collaborative settings. This means stakeholders of different domains interact with the virtual product through the models in their own views and viewpoints (Corley et al., 2016). Pertinent examples include the engineering of mechatronic and cyber-physical systems (CPS), and Internet-of-Things (IoT) systems, in which the hardware and software elements have to be modeled, simulated and verified together.

Such collaborative endeavors are, however, severely hindered by inconsistencies between the engineering artifacts such as domain-specific models, requirement specifications and other documents.

According to Herzig et al. (2014), *an inconsistency is present if two or more statements are made that are not jointly satisfiable.* Note, that this definition is not limited to models in a pure software settings, but is valid in multi-disciplinary engineering as well. Spanoudakis and Zisman (2001) also characterize the notion of an inconsistency using the joint non-satisfiability criterion, but in addition, they shed more light on the origin of inconsistencies by relating inconsistencies to *overlapping elements of different software models.*

In our view as well, inconsistencies stem from the joint non-satisfiability of statements. In our view, however, it is the ontological and linguistic *properties* (Vanherpen et al., 2016) of the system that cannot be the jointly satisfied. The link between the aforementioned overlaps and system properties has been first identified in previous work by Persson et.al. (Persson et al., 2013).

The problem of inconsistencies is exacerbated by the disparity of the domain-specific modeling languages and modeling tools involved in the flow of the engineering work. Stakeholders, quite often, do not use the same vocabulary when describing the various aspects of the system, resulting in overlaps between terminologies, and consequently, in their models. Because these inter-domain overlaps are not easy to identify, inconsistencies spanning the various involved domains are harder to detect.

Inconsistencies, if not managed properly, may lead to an incorrect product, which does not satisfy the properties required by the specifications. In extreme cases this may for example lead to safety breaches in mission critical systems.

Managing inconsistencies, therefore, is a must in every engineering process, especially in the collaborative ones. Wrongly executed inconsistency management, however, may have severe repercussions on the time it takes to produce a correct product. For example, dealing with incompatible sub-system interfaces during the integration phase may require additional iterations over costly engineering activities. Consequently, analyzing the impact of various inconsistency management techniques is desirable.

Explicit modeling of the engineering process enables combining the two facets of proper and efficient inconsis-

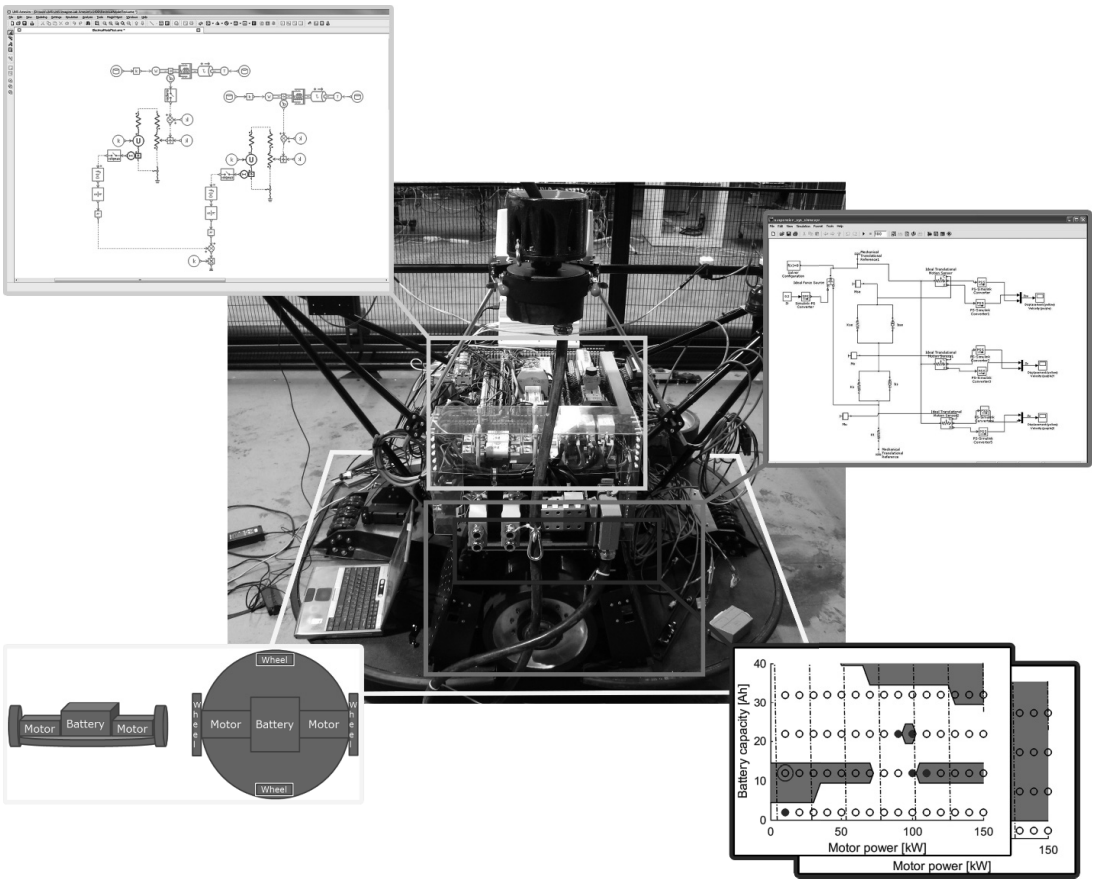


Figure 1: Autonomous Guided Vehicle (AGV): an example heterogeneous system which requires stakeholder collaboration across various domains.

tency management. On the one hand, the process model enables thorough understanding of where and when specific inconsistencies arise. On the other hand, it also enables a quantitative assessment of the impact of introducing a specific inconsistency management strategy to the process.

Here, we focus on how inconsistencies should be managed in collaborative, heterogeneous settings in order to guarantee the correctness of the end product, yet enabling an efficient flow of engineering work. We present PROxIMA, a process-oriented framework for managing inconsistencies. The main features of the framework are (i) its support for the optimal selection of inconsistency management techniques across the process, and (ii) enforcing the overall consistency of the virtual product by process enactment and tool interoperability. The feasibility of our approach has been demonstrated by means of the engineering of an Autonomous Guided Vehicle (AGV), shown in Figure 1.

MANAGING INCONSISTENCIES

Inconsistencies may have different causes. Human error is the top cause, with examples such as making design mistakes and communication shortcomings. In more elaborate cases, inconsistencies may arise from imprecise semantics of modeling languages or from a mismatch between the semantics of different modeling languages (Huzar et al., 2005).

As proposed by Finkelstein (2000), instead of simply just removing inconsistencies from the design, we need to think about "managing consistency". One of the core activities of (in-)consistency management is the detection of inconsistencies. It is preferred that detection is achieved as early as possible, because the earlier the inconsistencies are detected, the lower the costs of resolving them. Early detection of inconsistencies also provides more freedom in choosing the appropriate resolution and tolerance techniques Van Der Straeten (2005); Dávid et al. (2016b).

A significant amount of research has been dedicated to solving semantic inconsistencies on the syntactic level (for example (Adourian and Vangheluwe, 2007; Bhave et al., 2010)). These approaches, however, are prone to lose vital information during the approximation and abstraction steps taken while translating semantic properties to linguistic structures and parameters. We argue that reasoning over explicitly modeled semantic properties suits the problem of tackling inconsistencies better, as demonstrated by, e.g., Herzig and Paredis (2014).

Prevention

Preventive approaches advocate avoiding inconsistencies before they occur. A pertinent example is using locking mechanisms in collaborative modeling settings, for example, by storing the models into a version control system. Such a technique, however, does not allow true parallelism in the engineering process, as entire models/files are locked for one stakeholder, putting others' work on hold, thus rendering the process less performant. Property-based locking (Chechik et al., 2015) alleviates this problem by reducing the scope of the lock to selected parts of the model, by the use of graph patterns. Additionally, such a fine-grained locking mechanism solves the problem of overlocking, where multiple stakeholders would lock too many models/files and a deadlock or livelock could occur. To support true parallelism in the engineering flow, contract-based techniques have been recently proposed (Vanherpen et al., 2016). In this approach, the accepted value ranges of the various system parameters are being negotiated by the stakeholders concerned with the respective parameters, and the negotiated values are persisted in a contract. The contract asserts what different stakeholders guarantee towards the others with respect to a specific parameter. By establishing a contract prior to the actual engineering work, the consistency of the virtual product, with respect to the properties detailed in the contract, can be guaranteed.

Allow-and-resolve

Instead of spending effort on the a priori negotiation of a contract, other approaches deal with inconsistencies in situ, i.e., inconsistencies are allowed to occur and subsequently are detected and later resolved, with the possibility of tolerating them (Balzer, 1991) for a certain period. After that period, the inconsistencies are either naturally resolved or need to be resolved explicitly. Dependency analysis between models has been widely applied in collaborative settings, where inconsistencies typically occur due to the usage of different overlapping stakeholder models (Qamar et al., 2012). In such techniques, a pivot model (correspondence model) relates the elements of the models and by this, enables propagating change information from one model to the others. In its most basic form, such a technique can provide a stakeholder with the information that a change in another stakeholder's model may have resulted in an inconsistency, and consequently, investigation/resolution of the problem is required. If more information about the relationship between the models (such as algebraic relationships between their parameters) is known, propagation of the change information may also trigger automated resolution of the inconsistency. SysML is a typical choice for establishing pivot models at the architectural level. Recently, the combination of SysML and ontological reasoning has been proposed in order to support the detection of semantic inconsistencies (Feldmann et al., 2014).

To manage an inconsistency, one has to choose an appropriate management technique, and to decide when to apply it. This decision can be viewed as an optimization problem and as such, requires an objective function. Typical examples include minimizing the time to market, minimizing the overall cost per work item, maximizing the utilization of resources, etc. Whichever metric we choose as an objective function for the optimization, the only way to derive them is by modeling and simulating the underlying engineering process. To satisfy the objective function, the process is transformed to an improved one, where every inconsistency is managed.

The challenge: appropriate selection of inconsistency management techniques

The management of inconsistencies is achieved by selecting an appropriate management pattern for each inconsistency instance. The selection is made from a catalogue of management patterns, such as reordering activities, introducing automated or manual consistency inspections after a critical region of the process, applying an engineering contract (Sangiovanni-Vincentelli et al., 2012) on a specific part of the process, etc.

Since the same type of inconsistency may be managed by applying different different management patterns, the selection of the most appropriate one should happen through quantified cost measures. We use the expected transit time of the process (to approximate the time-to-market) as a cost measure and prefer processes with lower transit time, as they result in accelerated product development. This problem is translated to a constraint solving and optimization problem which finds the best process alternative while managing every potential source of inconsistencies, discussed in the following section.

THE PROxIMA FRAMEWORK

The PROxIMA (PRocess Optimization + Inconsistency MAnagement) (Dávid et al., 2016a, 2017) framework provides tools for modeling numerous aspects of the engineering process (Figure 2); transforming the process into an optimized form; and enacting the optimized process while interfacing with engineering tools. The most important capabilities of the framework are

- modeling the engineering process with both (control) flow of engineering activities, and (data) flow of engineering artifacts in it;
- modeling the attributes and properties of the engineered system, and relating them to engineering activities and artifacts;

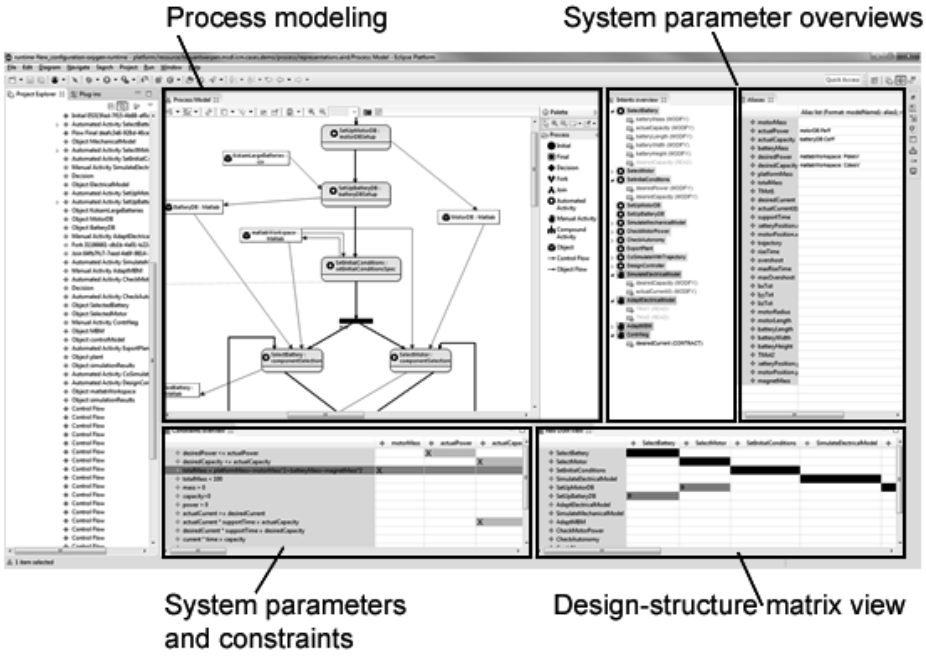


Figure 2: The modeling tool allows specifying various aspects of the underlying engineering process.

- resource and cost modeling;
- process optimization – finding the most cost-efficient process which still guarantees the correctness of the product by satisfying every system property;
- process enactment – exection and monitoring of the optimized process.

The tool was built on the Eclipse platform, using Python libraries for symbolic mathematical evaluations and tool interoperability.[1]
In the following, we discuss the typical steps required for inconsistency management in our framework.

Modeling the process

The modeling of the process is carried out using a visual modeling tool, built with the Sirius framework. Activities, their precedence, the input and output models are captured using the FTG+PM formalism (Lúcio et al., 2013). The formalism enables the usage of process models ("PM") in conjunction with the model of languages and transformations (the formalism-transformation graph - "FTG") used throughout the process. Formalisms and transformations serve as types to the processes: objects of the process are instances of languages of the FTG; and activities of the process realize transformations. In addition, system parameters which are subjects to potential inconsistencies, are charted as well and linked to the engineering activities which interact with them (e.g., by reading the value of the parameter or modifying it). This information is vital in identifying the root causes and scope of inconsistencies. In addition, resources (human and automated) can be modeled as well, such as stakeholders who enact the different activities, machines, tools, licenses, etc.
The optimality of the process is expressed by cost models. Multiple types of costs can be associated with activities, models, and tools. These costs are treated as different objectives to the optimization, resulting in a Pareto-front.

Off-line inconsistency management: process optimization by design-space exploration

After the process is modeled, the first phase of the inconsistency management takes place. The process is transformed so that as much inconsistencies are managed as possible, but also resulting in the best performing process, and respecting the resource constraints. The search for the resulting process is carried out using rule-based multi-objective design space exploration (DSE), in which the space of different process alternatives is

[1]The tool is available as open-source software (EPL) from `http://istvandavid.com/icm`.

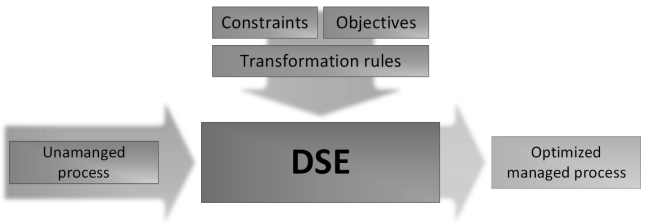


Figure 3: Design-Space Exploration.

searched to find the most appropriate one (Figure 3). The DSE engine takes the original, *unmanaged* and produces an *optimized managed* process as an output. This is achieved by applying pre-defined *model transformations rules* on the original process model in multiple iterations, thus constituting a search plan. Instead of an exhaustive approach, this search is governed by a hill climbing algorithm. There are two types of transformations:

- management transformations transform the process by applying management patterns onto it (e.g., introducing a contract negotiation activity in the process);
- optimization transformations, which try to parallelize the process as much as possible, also considering resource and scheduling constraints.

In every iteration, the consistency and performance of the new process is assessed so it can be decided whether the new process is an appropriate one. Managing inconsistencies is a hard *constraint* of the optimization, while maximizing the overall efficiency is a soft *objective*. This means, every potential inconsistency will be managed in the process, but this may sacrifice efficiency. Due to the stochastic nature of the process, assessing the performance of a candidate process is not straightforward. The process model is mapped onto the queueing network formalism of MathWorks' SimEvents and simulated to obtain a quantitative performance metric (Figure 4).
The following rules are used to map the engineering process onto the SimEvents formalism.

- **Activities** are translated to a Server processing a single token in the SimEvents formalism. The service time of the token in the server is based on the provided cost. The service time is either a constant value or a value sampled from a distribution.
- **Fork** nodes are translated to replicate nodes that process an incoming token (and all of it properties, like the total service time) to each of the outgoing branches.
- The **Join** node uses a combination of queues to let the tokens wait for the other branches. An entity combiner combines all tokens when they are available.
- **Decision** nodes are translated to an output switch element that routes to the available paths. The chosen path is sampled from the information provided in the process model. The merge node uses a path combiner to combine the incoming paths.

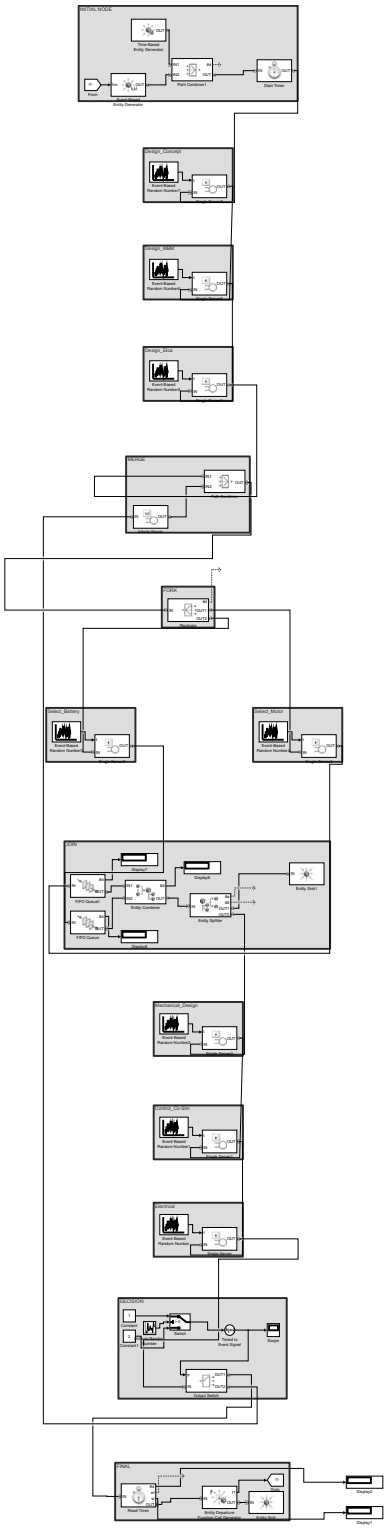


Figure 4: A quantitative SimEvents simulation model, generated from the original FTG+PM process model.

- Because we only allow for a **single process** to be executed at the same time, the logic in the final node allows for the creation of a new token in the initial node.
- The **control flow** between these newly created entities is equal to the control flow edges in the process model.

For each of the tokens, the total service time is recorded. As SimEvents is a stochastic formalism, multiple tokens are used to calculate the total cost of the process. We use the average service time of all tokens as the cost of the process.

On-line inconsistency management: process enactment and monitoring

Process enactment is commonly defined as the use of software to support the execution of operational processes, which enables mixing automated and manual activities. (CMMI Product Team, 2010) The PROxIMA framework provides an engine for enacting previously defined (and optimized) processes with the additional support for interoperability with a selection of engineering tools, such as Matlab/Simulink or Papyrus. During the enactment, the constraints of the system's parameters are continuously monitored. Whenever a violation of a constraint occurs, the stakeholders are notified about the occurrence of an inconsistency.
By employing symbolic mathematics, constraints can additionally be maintained in an incremental fashion and used for guiding the engineering work. Whenever the value of a system attribute can be derived from a combination of constraints and previously assigned attributes, the engine will provide this information to the stakeholder, thus aiding engineering decisions.

RELATED WORK

Inconsistency management has been a topic of interest for a long time in the domains of software engineering, mechatronic design and cyber-physical systems, due to their typically multi-view approach to system design. Di Ruscio et al. (2017) identify the research directions, challenges, and opportunities of collaborative MDSE and conclude that inconsistency management is one of the main enablers of efficient collaboration.
Multiple authors point out that managing inconsistencies should be carried out with processes in mind as well. Persson et al. (2013) identify consistency between the various views of cyber-physical system design as one of the main challenges in design of such complex systems. This is due to relations between views, with respect to their semantic relations, process and operations which often overlap. Multi-paradigm modeling Vangheluwe et al. (2002) advocates using the most appropriate formalism(s), at the most appropriate level(s) of abstraction, while also explicitly modelling the processes manipulating the models. The framework presented in this paper aims at the problem of semantics inconsistencies with a focus on processes.
Other approaches also acknowledge the role of semantic techniques in inconsistency management, and try to relate semantic concepts to the linguistic concepts of modeling. Hehenberger et al. (2010) organize structural design elements and their relations into a domain ontology to identify inconsistencies. A limited set of semantic properties are expressed with linguistic concepts which enables reasoning over semantic overlaps to a sufficient extent. Similarly, Chechik et al. (2015) introduce the notion of approximate properties: linguistic properties expressed as graph patterns which are accurate enough to appropriately approximate a semantic property. Approximate properties suitable to implement smart locking mechanisms in collaborative model-based design as they introduce a trade-off between the computational resources to obtain or check a property, and the accuracy of the results. As opposed to these, our approach makes semantic properties first-class artifacts and relates them to processes, instead of linguistic model elements, which enables management of a richer class of inconsistencies.
Ontologies have been used for inconsistency management by Kovalenko et al. (2014) to support automated detection of defects between domain-specific models. Similarly, Feldmann et al. (2014) use the OWL language in conjunction with a SysML-based approach to formally represent the design of a production system and evaluate the compatibility of domain-specific models in a collaborative setting. These approaches are complementary to ours: incorporating relationships between ontological properties for reasoning over inconsistencies is a planned extension to our work.
As opposed to the above techniques, inconsistency management in collaborative modeling is more frequently addressed on the linguistic level. Qamar et al. (2012) approach inconsistency management by making inter- and intra-model dependencies explicit. Dependencies are direct results of semantic overlaps and are used to notify stakeholders about possible inconsistencies when dependent properties change. Our approach introduces an indirection between models and properties by relating them to specific activities that during working over models also access properties with specific intents. Blanc et al. (2008) approach the detection of inconsistencies from a model operation based point of view, where models are stored as sequences of change events and inconsistencies are expressed in terms of Create Read Update Delete (CRUD) operations. Our approach generalizes this approach by introducing intents that are analogous with model operations, but they express change operations in terms of activities and properties.
In our work, we opted for the FTG+PM formalism for modeling processes. As compared to the widely used BPMN2.0 Object Management Group (OMG) (2011) or

BPEL-based process modeling frameworks (e.g., jBPM), our formalism allows modeling details more relevant to engineering scenarios in MDE settings. Models and transformations are first-class citizens in the FTG+PM, which enables deeper understanding of inconsistencies and more control over the enacted process.

Our framework provides simple undo/redo actions to revert to the latest consistent state of the models, but there have been other approaches to inconsistency resolution published. Mens et al. (2006) propose expressing the steps of inconsistency detection and resolution as graph transformation rules. Critical pair analysis is used to analyse potential dependencies between the detection and resolution of inconsistencies. It is, however, unclear whether critical pair analysis scales to industrial-size problems. Eramo et al. (2016) present an approach where each of the consistent alternatives are maintained throughout the process and removed when a decision is made and an alternative becomes infeasible. Almeida da Silva et al. (2010) investigate the possibilities of managing deviations of enacted processes from their respective specifications. It is not within the scope of our work, but indeed, deviations from the specified process are big threat to the validity of any process-oriented engineering approach. The efforts put into analyzing and optimizing a process model can be easily undone by deviating from (and sometimes even completely abandoning) the specification of the process.

CONCLUSIONS AND FUTURE WORK

In this paper, we presented an approach and framework for managing inconsistencies in collaborative and potentially highly parallelized, concurrent engineering settings. Our approach leverages the underlying engineering process and the various related information which enables reasoning over inconsistencies, their origins, impact and severity in a novel way.

We support the automated process of inconsistency management by a prototype tool. The approach has been evaluated over a case study of a mechatronic system, and Autonomous Guided Vehicle (AGV).

The approach discussed here can be an efficient enabler for collaborative engineering. At the same time, however, making the various facets of the engineering process explicit is a labor intensive and tedious task, which heavily frontloads the project. To alleviate the costs, the task of process modeling can be automated to a relatively high extent. The fundamental structure of the engineering process can be derived from business level processes.

In the future, we plan to augment our framework with performance simulation techniques for the process models. Our ongoing work focuses on metrics such as the transit time of tasks. The impact of inconsistencies on the process is another performance metric we plan to support through simulation. We also plan to enable explicit reasoning about the trade-off between different inconsistency management strategies during the process optimization phase (Dávid et al., 2016a) and during process enactment (Dávid et al., 2017). Another direction in our research is to support our approach with a library of inconsistency resolution techniques. We aim to develop a semi-automated selection of resolution methods, which will require detailed cost models.

ACKNOWLEDGEMENTS

This research has been partially funded by a BOF DOCPRO-1 grant of the University of Antwerp, Belgium.

REFERENCES

Adourian C. and Vangheluwe H., 2007. *Consistency Between Geometric and Dynamic Views of a Mechanical System.* In *Proc. of the 2007 Summer Computer Simulation Conf.* Society for Computer Simulation Int., San Diego, SCSC '07. ISBN 1-56555-316-0, 31:1–31:6.

Almeida da Silva M.A.; Bendraou R.; Blanc X.; and Gervais M.P., 2010. *Early Deviation Detection in Modeling Activities of MDE Processes*, Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-16129-2, 303–317. doi:10.1007/978-3-642-16129-2_22.

Balzer R., 1991. *Tolerating Inconsistency.* In *Proceedings of the 13th International Conference on Software Engineering.* IEEE Computer Society Press, ICSE '91. ISBN 0-89791-391-4, 158–165.

Bhave A.; Krogh B.; Garlan D.; and Schmerl B., 2010. *Multi-domain modeling of cyber-physical systems using architectural views. AVICPS 2010*, 43.

Blanc X.; Mounier I.; Mougenot A.; and Mens T., 2008. *Detecting model inconsistency through operation-based model construction.* In *Software Engineering, 2008. ICSE '08.* ISSN 0270-5257, 511–520. doi:10.1145/1368088.1368158.

Chechik M.; Dalpiaz F.; Debreceni C.; Horkoff J.; Ráth I.Z.; and Varró D., 2015. *Property-based methods for collaborative model development.*

CMMI Product Team, 2010. *CMMI for Development, Version 1.3, Tech. Rep. CMU/SEI-2010-TR-033.*

Corley J.; Syriani E.; Ergin H.; and Van Mierlo S., 2016. *Cloud-based multi-view modeling environments.* In *Modern Software Engineering Methodologies for Mobile and Cloud Environments*, IGI Global. 120–139.

Dávid I.; Denil J.; Gadeyne K.; and Vangheluwe H., 2016a. *Engineering Process Transformation to*

Manage (In)consistency. In *COMMitMDE 2016*. http://ceur-ws.org/Vol-1717/, 7–16.

Dávid I.; Meyers B.; Vanherpen K.; Van Tendeloo Y.; Berx K.; and Vangheluwe H., 2017. *Modeling and Enactment Support for Early Detection of Inconsistencies in Engineering Processes*. In *COMMitMDE 2017*.

Dávid I.; Syriani E.; Verbrugge C.; Buchs D.; Blouin D.; Cicchetti A.; and Vanherpen K., 2016b. *Towards Inconsistency Tolerance by Quantification of Semantic Inconsistencies*. In *COMMitMDE 2016*. 35–44.

Di Ruscio D.; Franzago M.; Muccini H.; and Malavolta I., 2017. *Envisioning the future of collaborative model-driven software engineering*. In *Proceedings of the 39th International Conference on Software Engineering Companion*. IEEE Press, 219–221.

Eramo R.; Pierantonio A.; and Rosa G., 2016. *Approaching Collaborative Modeling as an Uncertainty Reduction Process*. In *COMMitMDE 2016*. 27–34.

Feldmann S.; Kernschmidt K.; and Vogel-Heuser B., 2014. *Combining a SysML-based modeling approach and semantic technologies for analyzing change influences in manufacturing plant models. Procedia CIRP*, 17, 451–456.

Finkelstein A., 2000. *A Foolish Consistency: Technical Challenges in Consistency Management*. In *Database and Expert Systems Applications*, Springer, *LNCS*, vol. 1873. ISBN 978-3-540-67978-3, 1–5. doi:10.1007/3-540-44469-6_1.

Hehenberger P.; Egyed A.; and Zeman K., 2010. *Consistency Checking of Mechatronic Design Models*. In *30th Computers and Information in Engineering Conf.* ASME, vol. 3. ISBN 978-0-7918-4411-3, 1141–1148. doi:10.1115/DETC2010-28615.

Herzig S.J. and Paredis C.J., 2014. *Bayesian Reasoning Over Models*. In *MoDeVVa 2014*. 69–78.

Herzig S.J.; Qamar A.; and Paredis C.J., 2014. *An approach to identifying inconsistencies in model-based systems engineering. Procedia Comp Sci*, 28, 354–362.

Huzar Z.; Kuzniarz L.; Reggio G.; and Sourrouille J.L., 2005. *Consistency Problems in UML-based Software Development*. In *Proceedings of the 2004 International Conference on UML Modeling Languages and Applications*. Springer-Verlag, Berlin, Heidelberg, UML'04. ISBN 3-540-25081-6, 1–12. doi:10.1007/978-3-540-31797-5_1.

Kovalenko O.; Serral E.; Sabou M.; Ekaputra F.J.; Winkler D.; and Biffl S., 2014. *Automating Cross- Disciplinary Defect Detection in Multi-Disciplinary Engineering Environments*. In *Knowledge Engineering and Knowledge Management*, Springer. 238–249.

Lúcio L.; Mustafiz S.; Denil J.; Vangheluwe H.; and Jukss M., 2013. *FTG+PM: An Integrated Framework for Investigating Model Transformation Chains*. In *SDL 2013: Model-Driven Dependability Engineering*, Springer, *LNCS*, vol. 7916. ISBN 978-3-642-38910-8, 182–202. doi:10.1007/978-3-642-38911-5_11.

Mens T.; Van Der Straeten R.; and D'Hondt M., 2006. *Detecting and resolving model inconsistencies using transformation dependency analysis*. In *International Conference on Model Driven Engineering Languages and Systems*. Springer Berlin Heidelberg, 200–214.

Object Management Group (OMG), 2011. *BPMN 2.0 Specification*. `http://www.bpmn.org/`. Accessed: 2018-05-06.

Persson M.; Törngren M.; Qamar A.; Westman J.; Biehl M.; Tripakis S.; Vangheluwe H.; and Denil J., 2013. *A Characterization of Integrated Multi-View Modeling in the Context of Embedded and Cyber-Physical Systems*. In *EMSOFT*. IEEE. ISBN 9781479914432, 1–10. doi:10.1109/EMSOFT.2013.6658588.

Qamar A.; Paredis C.J.; Wikander J.; and During C., 2012. *Dependency modeling and model management in mechatronic design. Journal of Computing and Information Science in Engineering*, 12, no. 4, 041009.

Sangiovanni-Vincentelli A.; Damm W.; and Passerone R., 2012. *Taming Dr. Frankenstein: Contract-Based Design for Cyber-Physical Systems. European Journal of Control*, 18, no. 3, 217 – 238. ISSN 0947-3580. doi:http://dx.doi.org/10.3166/ejc.18.217-238.

Spanoudakis G. and Zisman A., 2001. *Inconsistency management in software engineering: Survey and open research issues*. In *in Handbook of Software Engineering and Knowledge Engineering*. World Scientific, 329–380.

Van Der Straeten R., 2005. *Inconsistency management in model-driven engineering. An Approach Using Description Logics (PhD thesis), Vrije Universiteit Brussel, Brussels, Belgium*.

Vangheluwe H.; De Lara J.; and Mosterman P.J., 2002. *An introduction to multi-paradigm modelling and simulation*. In *Proc. of the AIS'2002 conf.* 9–20.

Vanherpen K.; Denil J.; Dávid I.; Meulenaere P.D.; Mosterman P.J.; Torngren M.; Qamar A.; and Vangheluwe H., 2016. *Ontological reasoning for consistency in the design of cyber-physical systems*. In *CPPS 2016*. 1–8. doi:10.1109/CPPS.2016.7483922.

SIMULATION ANALYSIS IMPROVES OPERATIONS AT EMERGENCY DEPARTMENT AND SURGICAL SUITE

Gabriel Carreño, Srinivas Rajanna, and Edward J. Williams

Simulation Analytics
PMC
15726 Michigan Avenue
Dearborn MI 48126
United States of America
E-mail: {gcarreno|srajanna|ewilliams}@pmcorp.com

KEYWORDS

Emergency department, surgical suite, discrete-event process simulation, health care.

ABSTRACT

Discrete-event simulation now has a long and distinguished track record of guiding improvements to queueing systems subject to severe operational and budget constraints and also held to lofty expectations of service speed and quality. Historically, this track record began with simulation applications to manufacturing operations. These applications have now expanded far beyond manufacturing plants, to include supply-chain distribution systems, transport terminals, service industries (e.g., hotels and restaurants), and health care delivery in hospitals, clinics, and doctors' and dentists' offices.

Relative to the application of discrete-event simulation discussed in this paper, a large medical center in the Midwest region of the United States availed itself of the ability of simulation to guide improvements to and expansion of the emergency department and the surgical suite over a planning horizon of nearly a decade. This medical center serves a large urban area and the surrounding rural areas, and anticipates considerable pressure of increasing demand for its services. This increasing demand is attributable to significant demographic trends such as increasing density of population and gradually increasing average age of this population.

INTRODUCTION

During the last half-century, discrete-event process simulation has made enormous contributions to productivity, efficiency, and economic viability of a wide variety of activities in many economic sectors. Historically, the first, and still many of the most conspicuous, such contributions pertained to manufacturing, whether in large assembly plants or "custom-order" job shops; example applications abound in the literature. More recently, simulation has been used extensively in service industries (e.g. restaurants), government operations (e.g., courthouses), supply chain operations (e.g., configuration and operation of warehouses), and the health care industry. In the health care industry, simulation is making high-value contributions to the configuration, staffing, and operation of hospitals, clinics, long-term care centers, and private medical practices. For example, (Baskaran, Bargiela, and Qu 2013) and (Yankovic and Green 2011) applied simulation to the cost-effective scheduling of nurses, thereby addressing the chronic shortages and high costs of skilled nurses. Discrete-event simulation was used to analyze and improve the performance of a hospital blood laboratory, as documented in (Kadı, Kuvvetli, and Çolak 2016). Recently, (Zhang, Hanchi, and Meijer 2017) used simulation to identify the factors most heavily influencing the length of patients' stays in a surgery center, thereby improving the reliability of scheduling. As an example of planning for most unwelcome contingencies, (Zehrouni et al. 2017) describes the use of simulation to propose a viable emergency plan in case of a major flood in the region a hospital serves – disasters simultaneously increase demand for emergency services and impede provision of them.

In the present study, which covered a *much* longer time horizon than most simulation studies in either health care or manufacturing, managers and analysts at a large medical center in the Midwest region of the United States sought simulation guidance and predictive analytics for their emergency department and surgical suite – and the interrelationships between them. Patients initially entering the emergency department often proceed to surgery after doctors stabilize their medical condition. Therefore, *two* simulation models were built and analyzed – one for the emergency department and one for the surgical suite – and the outputs of these models were analyzed concurrently. Due to demographic trends, management of the medical center is confident that demand for services will increase steadily and significantly through a planning horizon ending in 2024. These managers looked to simulation to guide decisions on expansion plans and choices of operational procedures.

The remainder of this paper is organized as follows: The next section describes, at a high level, the operations at the medical center, particularly the emergency department and the surgical suite. The following two sections explain the steps undertaken to analyze the input data and to build,

verify, and validate the model. Then we present highlights of the results and indicate likely directions of future work.

OVERVIEW OF OPERATIONS

The regional medical center serves residents of parts of three midwestern states. Further, by providing an air-ambulance service, it can and does provide lifesaving emergency service to rural and farming regions nearby. The simulation study described in this paper, at the client's request, concentrated on two of the largest and most vital components of the medical center: The emergency department and the surgical suite. A key objective of the study was to provide advice to upper management on how to best ensure support of projected increased patient volumes through 2024. The emergency department, as is typical, must address urgent issues of trauma (e.g., automobile accidents), cardiac emergencies (e.g., heart attack), chest pain (e.g., pneumonia), and many others, as described in (Rossetti, Trzcinski, and Syverud 1999). The surgical suite, comprising multiple operating rooms, has similar complex operational and staffing concerns, similar to those described in (Lowery and Davis 1999). One of the operating rooms is designated as cardiovascular – and reserved for that type of surgery only. Furthermore, at least one operating room within the surgical suite must always be available (if not indeed in actual use!) for a suddenly arising emergency surgery. The surgical suite also includes the post-anesthesia care unit [PACU].

INPUT DATA AND ITS ANALYSIS

Extensive input data from medical center records, encompassing the calendar year 2016, was made available to the simulation analysts. For the emergency department, these data included arrival time, medical problem presented, the extent of services required (time devoted to the patient by nurse(s) and doctor(s), and any other services such as X-rays), and the disposition of the patient (i.e., to home, to a private doctor, or into a hospital, or into the medical center surgery center. For the surgical suite, these data included date and time, operation performed, operating room used, time spent by surgeons, anesthesiologists, and nurses, and time the patient subsequently spent in the PACU. These data were entered into Microsoft Excel® workbooks and thence read into the respective Simio® models. Examples of these data tables, as provided by the clients, are shown in the Appendix (Figure 1 and Figure 2). Therefore, in this project, there was no need to fit closed-form probability distributions to input data sets (Law 2016).

Additionally, the medical center documented standard operational procedures, such as the typical routes traveled by and services received by various types of patients. For example, the emergency department of course operated continuously, 24 hours a day, 7 days a week. Except for emergency surgeries, the operating rooms within the surgical suite operated 7AM to 5PM Mondays, Wednesdays, and Fridays, with slightly reduced hours (7AM-3PM Tuesdays and Thursdays). Particularly for the surgical suite model, preliminary analysis with Microsoft Excel® (including macros) derived key information such as the proper proportional allocation of patients to type of surgery and length of stay for the various types of surgery.

MODEL CONSTRUCTION, VERIFICATION, AND VALIDATION

After discussion of more than half a dozen alternatives (there are many as documented in (Abu-Taieh, Evon M. O. and Asim Abdel Rahman El Sheikh 2007)) between the clients and the simulation analysts, the well-known simulation software Simio [SIMulation using Intelligent Objects], known to several client engineers via their university studies, was selected for construction of the two simulation models. This software provides ease of use, high modeling power, easy specification of extensive experimentation to compare scenarios, and high-quality animation (Joines and Roberts 2015).

For the emergency department model, typical hourly demand curves for each day of the week were constructed, based upon the historical 2016 data. Lengths of stay were calculated as a range for each patient type and modeled as a triangular distribution. The clients specifically requested that the triangular distribution be used, and this request was deemed reasonable based on histograms of the pertinent data set. Likewise, for the surgical suite model, the year-2016 case volumes and turnaround times, subclassified by patient type and surgical specialty, were scheduled across each day of the week within the model. Post-surgery dwell times in the PACU were incorporated into this model similarly.

Various well-known methods of verification and validation were used (Sargent 2013). These included directional analyses (if arrival rate increases, do queue lengths increase?), examination of the animation, and allowing only one entity (patient) to enter the model and examining the path it takes. Very importantly, in view of the unusually extensive historical data available, both models were run "with yesterday's data." Simio® provides a valuable feature called an "arrival table," allowing convenient specification of arrivals according to a historical logbook (Smith, Sturrock, and Kelton 2017).

Since validation included having the models precisely (within 3%) duplicate "yesterday's observed historical results," both models very quickly achieved high credibility with the clients. This credibility was also enhanced by realistic, yet not overelaborate, animations. Next, each model was run for one year (indeed, as in the (Kadı, Kuvvetli, and Çolak 2016) study) to obtain key performance metrics, including queue lengths, queue residence times, resource utilizations, and total patients served. In lieu of specifying a warm-up period (the usual procedure when modeling a steady-state system such as this medical center, in contrast to a terminating system such as a bank or retail store) (Robinson 2014), the analyst, using the extensive historical data available, captured the times of entry and exit

into each service center, and then calculated the service center's utilization.

RESULTS OF THE SIMULATION MODEL

Results reported by the emergency department model included average and maximum times in system and in queue, average and maximum queue lengths, and utilization of resources (both personnel such as doctors, nurses, and aides; and resources of equipment (e.g., X-ray machines). These results were presented both overall and subdivided by time of day and day of week. They were compared for various possible arrival rates, reflecting uncertainty of the increased demand for services during the planning horizon specified in the study.

Results (including the same performance metrics as for the emergency department, plus the percentages of operations starting at the originally scheduled time) obtained from the surgical suite model included comparative simulations with six, seven, and eight operating rooms; each of these scenarios was run with both average (to include both arithmetic mean and median) and third-quartile (75^{th} percentile) patient volume demand levels. For each of these scenarios, utilizations of the operating rooms were reported both in aggregate and by day of the week, and further relative to type of operation (e.g., cardiac) performed. Analogous results were reported for the preparatory rooms and the PACU.

CONCLUSIONS AND FURTHER WORK

Conclusions from this simulation study provided significant reassurances to the client relative to the need for very large capital expenditures, which – if needed – must be undertaken early in the planning horizon. Specifically, relative to the emergency department, the simulation model and analysis of its output confirmed that:

1. The medical center will be able to handle projected volumes of patients with the 24 treatment rooms already available and currently underutilized;
2. The low utilization of these rooms will assure significant capacity for volume growth without investing in more physical space (the eight hallway rooms were barely used during the simulation runs);
3. The current number of employees working in registration and triage will continue to be sufficient throughout the planning horizon.

Relative to the surgical suite, analysis of the model similarly confirmed that:

1. The medical center will be able to handle the 2024 projected patient volume with eight operating rooms at 65%-70% utilization;
2. Overall capacity can be increased by allowing non-cardiovascular cases to be processed in the operating room currently dedicated "cardiovascular" whenever doing so is medically feasible;
3. The projected number of pre-operation rooms and PACUs will be sufficient to handle projected volume increases through 2024.

Toward the end of this simulation study, on client request, a cursory analysis of the medical center's parking lot capacity was undertaken. Inasmuch as public transport is sorely lacking in the region the hospital serves, almost all patients not arriving by ambulance arrive by private vehicle (very likely driven by a relative or friend). Likewise, these friends and relatives often come to the medical center to visit patients. Very plausibly, future work will include a much more detailed study of this parking lot, its capacity, and the number of parking places needed as "handicapped parking."

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the collaboration and cooperation of client managers and industrial engineers, as well as the support and guidance of PMC supervisors and project managers. Additionally, constructive criticisms from several anonymous referees helped the authors significantly improve this paper.

REFERENCES

Abu-Taieh, Evon M. O. and Asim Abdel Rahman El Sheikh. 2007. Commercial Simulation Packages: A Comparative Study. In *International Journal of Simulation: Systems, Science & Technology* 8(2): 66-76.

Baskaran, Beetha, Andrzej Bargiela, and Ron Qu. 2013. A Study of Cost Effective Scheduling of Nurses Based on the Domain Transformation Method. In *Proceedings of the 27th European Conference on Modelling and Simulation*, eds. Webjørn Rekdalsbakken, Robin T. Bye, and Houxiang Zhang, 309-314.

Joines, Jeffrey A. and Stephen D. Roberts. 2015. *Simulation Modeling with Simio: A Workbook*, 4th edition. Simio® LLC.

Kadı, Deniz, Yusuf Kuvvetli, and Selçuk Çolak. 2016. Performance Analysis of a University Hospital Blood Laboratory via Discrete Event Simulation. In *Simulation: Transactions of the Society for Modeling and Simulation International* 92(5):473-484.

Law, Averill M. 2016. A Tutorial on How to Select Simulation Input Probability Distributions. In *Proceedings of the 2016 Winter Simulation Conference*, eds. Theresa M. Roeder, Peter I. Frazier, Robert Szechtman, and Enlu Zhou, 103-117.

Lowery, Julie C. and Jennifer A. Davis. 1999. Determination of Operating Room Requirements Using Simulation. In *Proceedings of the 1999 Winter Simulation Conference*, Volume 2, eds. Phillip A. Farrington, Harriet Black Nembhard, David T. Sturrock, and Gerald W. Evans, 1568-1572.

Robinson, Stewart. 2014. *Simulation: The Practice of Model Development and Use*, 2nd edition. Basingstoke, United Kingdom: Palgrave Macmillan.

Rossetti, Manuel D., Gregory F. Trzcinski, and Scott A. Syverud. 1999. Emergency Department Simulation and Determination of Optimal Attending Physician Staffing Schedules. In *Proceedings of the 1999 Winter Simulation Conference*, Volume 2, eds. Phillip A. Farrington, Harriet Black

Nembhard, David T. Sturrock, and Gerald W. Evans, 1532-1540.
Sargent, Robert G. 2013. An Introduction to Verification and Validation of Simulation Models. In *Proceedings of the 2013 Winter Simulation Conference*, eds. Raghu Pasupathy, Seong-He Kim, and Andreas Tolk, 321-327.
Smith, Jeffrey S., David T. Sturrock, and W. David Kelton. 2017. *Simio and Simulation: Modeling, Analysis, and Applications*, 4th ed. CreateSpace Independent Publishing.
Yankovic, Natalia and Linda V. Green. 2011. Identifying Good Nursing Levels: A Queuing Approach. In *Operations Research* 59(4): 942-955.
Zehrouni, Afafe, Vincent Augusto, Thierry Garaix, Raksmey Phan, and Xiaolan Xie. 2017. Health Care Emergency Plan Modeling and Simulation in Case of Major Flood. In *Proceedings of the 2017 Winter Simulation Conference*, eds. Victor Chen, Andrea D'Ambrogio, Gregory Zacharewicz, and Navonil Mustafee, 2994-3005.
Zhang, Chen, Hamza Hanchi, and Sebiastiaan Meijer. 2017. Identifying Influential Factors of Patient Length of Stay in a Surgery Center: A Simulation Modelling Approach. In *Proceedings of the Second International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing*, eds. Doris Lindörfer, Hassan Khachfe, and Birgit Gersbeck-Schierholz, 28-34.

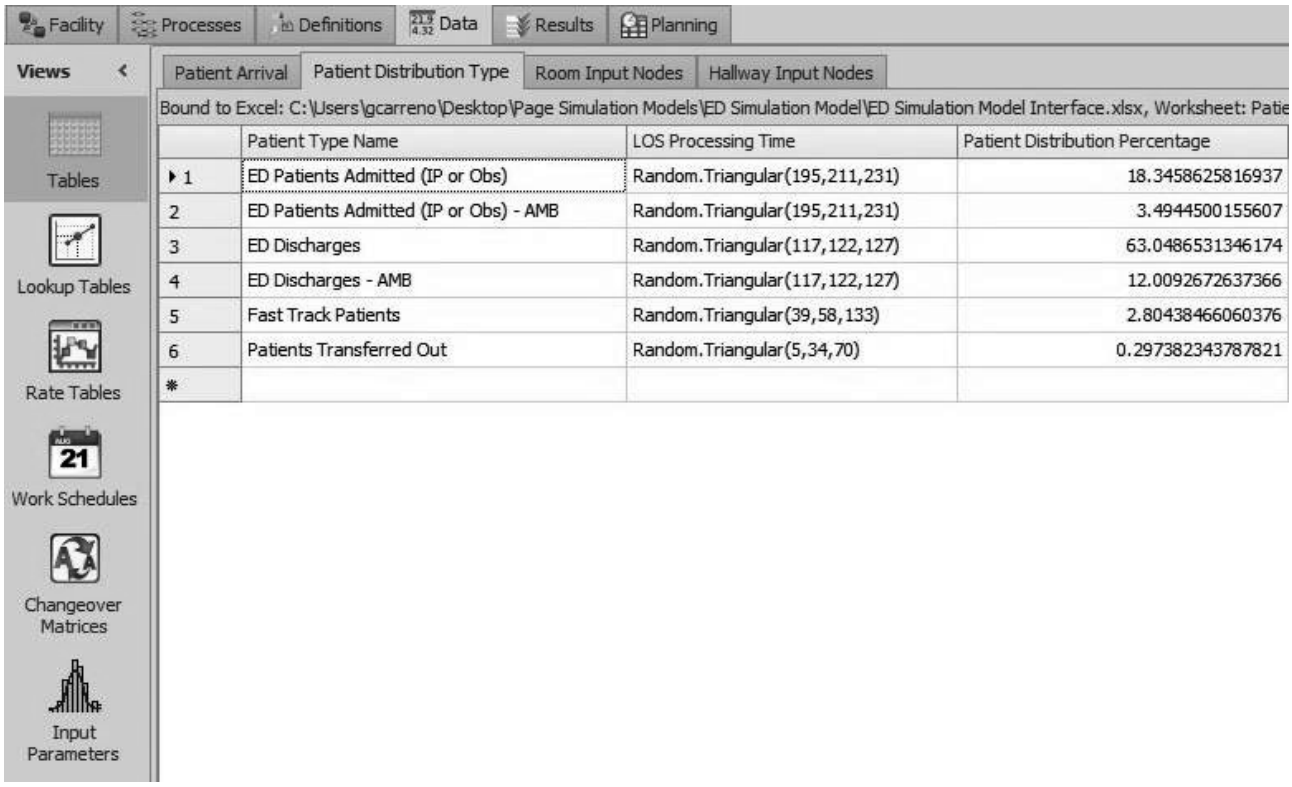
AUTHOR BIOGRAPHIES

GABRIEL CARREÑO holds a bachelor's degree in Industrial Engineering (Texas State University, 2015) with a minor in applied mathematics. From January 2015 to May 2015 he worked for his Capstone Design Project building a material handling simulation model to optimize the warehouse of a major North American supermarket chain. In August 2015 he joined Production Modeling Corporation (PMC) in Dearborn, Michigan, as an Industrial Engineer. He has been working with various companies through PMC on material handling, scheduling, and simulation projects using a variety of analytical software tools such as ProPlanner®, Preactor®, AutoMod®, SIMUL8®, Simio®, and Witness®. His email address is gcarreno@pmcorp.com.

SRINIVAS RAJANNA is the Director of Program Management at PMC. He has over fifteen years of experience in simulation, lean, production, process improvement, six-sigma, theory of constraints, supply chain, and managing projects. He graduated from Bangalore University with a bachelor of engineering in mechanical engineering. He holds a master's degree in industrial engineering from West Virginia University and an MBA from the Eli Broad Graduate School of Management, Michigan State University. He is certified in production and inventory management (CPIM) from APICS. Srinivas has broad industry experience that includes healthcare, automotive, aerospace, semiconductor, consumer, nuclear, manufacturing, and pharmaceutical. He has experience providing solutions that include developing throughput improvement roadmap to meet the target production, assessing the operation strategies, conducting material flow studies to reduce traffic congestion, optimizing the utilization of staff and equipment, applying lean strategies in manufacturing, healthcare, and service industries, using analytical techniques such as flow charts, value stream mapping, and process mapping etc. His email address is srajanna@pmcorp.com.

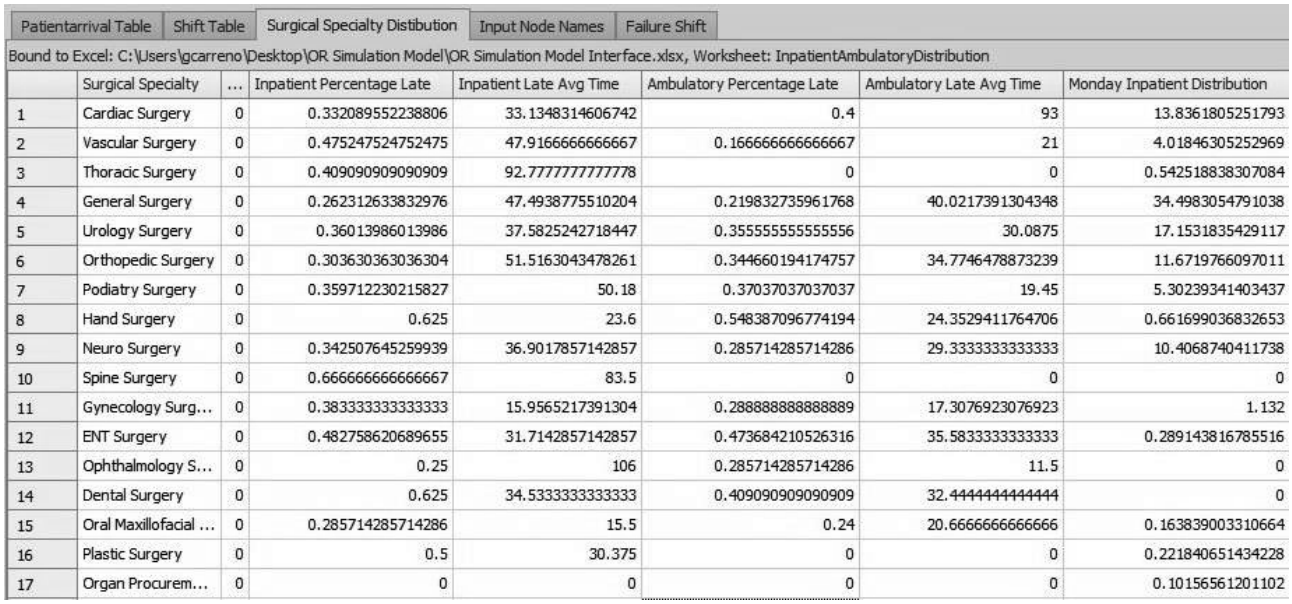
EDWARD WILLIAMS holds bachelor's and master's degrees in mathematics (Michigan State University, 1967; University of Wisconsin, 1968). From 1969 to 1971, he did statistical programming and analysis of biomedical data at Walter Reed Army Hospital, Washington, D.C. He joined Ford Motor Company in 1972, where he worked until retirement in December 2001 as a computer software analyst supporting statistical and simulation software. After retirement from Ford, he joined PMC, Dearborn, Michigan, as a senior simulation analyst. Also, since 1980, he has taught classes at the University of Michigan, including both undergraduate and graduate simulation classes using GPSS/H™, SLAM II™, SIMAN™, ProModel®, SIMUL8®, or Arena®. He is a member of the Institute of Industrial Engineers [IIE], the Society for Computer Simulation International [SCS], and the Michigan Simulation Users Group [MSUG]. He serves on the editorial board of the International Journal of Industrial Engineering – Applications and Practice. During the last several years, he has given invited plenary addresses on simulation and statistics at conferences in Monterrey, México; İstanbul, Turkey; Genova, Italy; Rīga, Latvia; and Jyväskylä, Finland. He served as a co-editor of Proceedings of the International Workshop on Harbour, Maritime and Multimodal Logistics Modelling & Simulation 2003, a conference held in Rīga, Latvia. Likewise, he served the Summer Computer Simulation Conferences of 2004, 2005, and 2006 as Proceedings co-editor. He was the Simulation Applications track coordinator for the 2011 Winter Simulation Conference. A paper he co-authored with three of his simulation students won "best paper in track" award at the Fifth International Conference on Industrial Engineering and Operations Management, held in Dubai, United Arab Emirates, in March 2015. His email addresses are ewilliams@pmcorp.com and williame@umich.edu.

APPENDIX



	Patient Type Name	LOS Processing Time	Patient Distribution Percentage
1	ED Patients Admitted (IP or Obs)	Random.Triangular(195,211,231)	18.3458625816937
2	ED Patients Admitted (IP or Obs) - AMB	Random.Triangular(195,211,231)	3.4944500155607
3	ED Discharges	Random.Triangular(117,122,127)	63.0486531346174
4	ED Discharges - AMB	Random.Triangular(117,122,127)	12.0092672637366
5	Fast Track Patients	Random.Triangular(39,58,133)	2.80438466060376
6	Patients Transferred Out	Random.Triangular(5,34,70)	0.297382343787821
*			

Figure 1. One of several Data Tables for the Emergency Department Model



	Surgical Specialty	...	Inpatient Percentage Late	Inpatient Late Avg Time	Ambulatory Percentage Late	Ambulatory Late Avg Time	Monday Inpatient Distribution
1	Cardiac Surgery	0	0.332089552238806	33.1348314606742	0.4	93	13.8361805251793
2	Vascular Surgery	0	0.475247524752475	47.9166666666667	0.166666666666667	21	4.01846305252969
3	Thoracic Surgery	0	0.409090909090909	92.7777777777778	0	0	0.542518838307084
4	General Surgery	0	0.262312633832976	47.4938775510204	0.219832735961768	40.0217391304348	34.4983054791038
5	Urology Surgery	0	0.36013986013986	37.5825242718447	0.355555555555556	30.0875	17.1531835429117
6	Orthopedic Surgery	0	0.303630363036304	51.5163043478261	0.344660194174757	34.7746478873239	11.6719766097011
7	Podiatry Surgery	0	0.359712230215827	50.18	0.37037037037037	19.45	5.30239341403437
8	Hand Surgery	0	0.625	23.6	0.548387096774194	24.3529411764706	0.661699036832653
9	Neuro Surgery	0	0.342507645259939	36.9017857142857	0.285714285714286	29.3333333333333	10.4068740411738
10	Spine Surgery	0	0.666666666666667	83.5	0	0	0
11	Gynecology Surg...	0	0.383333333333333	15.9565217391304	0.288888888888889	17.3076923076923	1.132
12	ENT Surgery	0	0.482758620689655	31.7142857142857	0.473684210526316	35.5833333333333	0.289143816785516
13	Ophthalmology S...	0	0.25	106	0.285714285714286	11.5	0
14	Dental Surgery	0	0.625	34.5333333333333	0.409090909090909	32.4444444444444	0
15	Oral Maxillofacial ...	0	0.285714285714286	15.5	0.24	20.6666666666666	0.163839003310664
16	Plastic Surgery	0	0.5	30.375	0	0	0.221840651434228
17	Organ Procurem...	0	0	0	0	0	0.10156561201102

Figure 2. One of several Data Tables for the Surgical Suite Model

Mobile Computing in Patient Relationship Management – A Case Study

Pedro Moreira, Daniela Oliveira, Filipe Miranda, António Abelha and José Machado
Department of Informatics
University of Minho, Portugal
E-mail: antonio.pedro.moreir@gmail.com,
daniela777@live.com,pt,
filipemanuelmotamiranda@gmail.com,
abelha@di.uminho.pt,
jmac@di.uminho.pt

KEYWORDS
Information Technologies, PRM, push notifications, mobile application, e-Health.

ABSTRACT

Around fourteen years ago, the first hospital in Portugal started to use a patient notification system, to alert patients of their medical events at the hospital, such as appointments, surgeries, exams, treatments through text messages. This notification system is used nowadays, but it faces a big problem: a huge amount of money spent for the telecommunication companies involved. Although each message cost a fraction of cents, it can easily reproduce its value in more than 50,000 euros per year per institution. Since technology and the use of smartphones has been evolving in such a quick way, it is estimated that in no more than 10 years, almost all the Portuguese population will use smartphones or have access to them. For those reasons, the main purpose of the present work, is to design and develop a mobile application in order to substitute the previous notification systems, through push up notifications on the app and by email, that can be saved on the smartphone calendar, translating in no costs associated with the notifications sent by the hospitals. The main motivation is, therefore, suppressing these costs for the hospital, bring the patients closer integrating other systems on the app and make the notification alert more efficient. Thereby, the mobile app will be able, not only to manage each notification and notify the patients, but also to check its medical event history and to schedule medical appointments.

INTRODUCTION

As the Portugal healthcare system moves to a more technological care system, the importance of engaging patients in that journey must equally intensify. These days, engaging patients to improve their subjective satisfaction will not be enough for providers who want to maximize their value. The optimization relies on developing strong and long-term relationships with patients, ensuring a daily follow-up to the patient. An example of that can be the implementation of a good Patient Relationship Management (PRM) system, just as other companies in non-healthcare industries, with the concept of customer relationship management (CRM). We can think of the patients as real customers, and actually focus on providing the most oriented and individual service as possible. As the result of suppressing the costs of the notification system implemented in most hospitals of the country, the idea of the new patient notification system, evolved to a PRM mobile application that can not only notify the patients and help them manage their medical agenda, but also bring the hospital closer to its reach by providing services on the app that require a phone call or physical presence. Therefore, it is highly predictable that this mobile app can bring future work in terms of integrating more and more services in order to innovate the public healthcare system and turn it more efficient and up-to-date.

In "*How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX*", a concise overview of health and healthcare smartphone applications is described. This review considers the mobile phone applications a significant and potential advance in different points in healthcare. One of relevant benefits is at the economic level, as well as its portability, continuous data stream, and the capability through sufficient computing power to support multimedia software applications (Boulos et al. 2011).

This paper presents an insight into a prototype development and impact analysis of this mobile application in two hospitals in Portugal in Porto and in Vila Verde, Braga. The main focus is to highlight the potential of this more up-to-date method and the advantages, when comparing to the old one.

In the State of the Art section, similar works are briefly described. Moreover, in section 3, it will be presented the research methodologies. Next section, the project designed and in Section 5, the conclusion and future work.

STATE OF THE ART

This section pretends to focus on the main topics addressed throughout the work as well as related technology from SNS – *"Serviço Nacional de Saúde" (*National Health Service).

Information Technologies (IT)

First, IT have been pointed out as a tool capable of revolutionize the provision of healthcare treatments by improving significantly its efficiency (Lee, McCullough, and Town 2013). In a general way, IT represents all the

activities and solutions available by computer resources with the focus to storage, transmit, access, produce and use the information available. In the last two decades, e-Health concept has emerged and has gradually obtained a lot of importance. Eysenbach claims that e-Health is the "intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology" (Eysenbach 2001). Considering that information, IT in healthcare can promote data sharing in a more efficient way, turning medical decision making more accurate. It can also create a new relationship between healthcare professionals, organizations and patients, taking down barriers and improving communication. This can lead to the decrease the risk of medical errors and improve the efficiency of its services, from healthcare service, to administration or economic management.

AIDA ("*Agência para Integração, Difusão e Arquivo de informações médicas*")

AIDA stands for agency to integrate, share and store medical information. AIDA is a multi-agent system, devoloped by Minho University investigators, representing the central unit of interoperability in many major hospitals and medical institutions in Portugal. This platform is highly efficient and modular, able to adapt according to the specific needs of each institution it operates in (Peixoto et al. 2012). Even though it's a complex system, its composed by simple subsystems, known as intelligent agents responsible for communication tasks among different systems (Cardoso et al. 2014). This platform is able to integrate other information systems (IS) such as electronic health record, administration IS, doctor support IS, among others. In order to enhance the communication and consequently its interoperability, this platform uses Health Level 7 (HL7), a normalized communication protocol, which is a set of international standards for clinical and administrative transfer of data between software applications. AIDA, ensures the interoperability between different intelligent systems such as SAM, SONHO, PCE, SAPE, RIS, LIS (Cardoso et al. 2014; Machado et al. 2006). It also includes a module called AIDA-PRM, which is responsible for the communication with patients, via text message to alert them of their medical events (University of Minho n.d.). This project wants to be an alternative and more up-to-date notification system.

MySNS – SPMS (Shared Services of Health Ministry)

In a similar way to increase the proximity of the patient to the health institution and its transparency of its services, SPMS, which are the shared services of the health ministry, developed this mobile application. My SNS has become an official link from the National Health Service (SNS) to the citizens and it works as a mobile connection to the digital health services provided by SNS. The mobile app, allows the user to see the SNS news, a map and a list of the public Hospitals, public health Centres and pharmacies in the country. It also gives direct access to services provided by SNS, such as Health Contact Centre (SNS 24). Besides that, the app, notifies the users of meteorological variations associated to its location but has no working functionalities in services like the one this project is aiming to implement (SPMS n.d.). The next section presents the work being developed.

RESEARCH METHODOLOGIES

In the field of Information Technologies (ITs), a research project and development of a product should be based in rigorous research methodologies and technologies up to the final product, as well as its final analysis in terms of the solution presented. The most suited methods and tools have been chosen based on the advantages, limitations that will be shown in the next sections.

This project is being developed based on the research methodology Design Science Research (DSR), commonly used for constructing and evaluating useful IT solutions in a more rigorous way (Hevner and Chatterjee 2010). The appropriate methodologies, technologies and tools chosen to elaborate and define the solution are included in each design phase presented in this study.

In first subsection, DSR research method will be presented, then in second subsection, a Proof of Concept in the form of a SWOT analysis in order to predict future problems or disadvantages as well as enhancing the solution by using the potential of itself.

Design Science Research (DSR)

DSR is a scientific research methodology that aims to obtain the best product possible. This product, has to be able to match each goal proposed initially and solve the problem. The resulting artefact will be the technologically feasible solution, to relevant problems and its technical features have to be demonstrated in favor of its accreditation when being evaluated. The research has to be clarifying and efficient as possible in order to make the process flow as it is supposed to (Hevner and Chatterjee 2010). The next figure illustrates the ideal DSR investigation methodology model in the context of this case study, where the motivation to find the

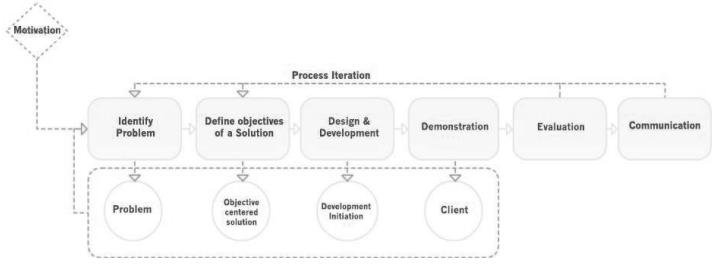


Figure 1: DSR Methodology Process Model (Hevner and Chatterjee 2010).

best solution to the problem (case study) is to eradicate the notification costs for the hospitals in the years ahead as well as enhancing the Patient/Hospital relationship. Although it is represented the complete methodology in figure1, part of the phase of demonstration and the evaluation and communication are future work, since it is only a prototype yet.

The choice of this methodology was based on facts well supported and justified by the case study created in order develop a suitable solution for the problem. After the first version of the final artefact is created to the client, it will be evaluated and criticized by the users, to verify the pre-established requirements and possible enhancements for future work (Gregor and Hevner 2013). DSR is considered to be a cyclic methodology starting by identifying the problem and creating the case study, predicting the solution based on the research and investigation made. The solution is then presented to the user or client and gives its feedback. In case the feedback does not match the pre-requirements, the process starts again (Vaishnavi, Kuechler, and Petter 2017). The starting point is problem and motivation identification, that translated to be the amount of money spent in the notification process and the goal to find a solution that can deliver the notifications to the patients without cost in the process in a more efficient and valuable way (hospital and patient). Therefore, the prototype created is a mobile application capable of notificate each patient for its medical events, as well as providing hospital services in order to enhance this relationship. The next phase consists in testing the app in "Vila Verde" Hospital. After testing, it will be evaluated its effect in order to find gaps or enhancements opportunities. Finally, the artefact will be defunded in a dissertation project. The final products in healthcare obtained through DSR methodology guarantee that each solution found for a specific problem in IT, cover the needs of the institution working with. This case study can therefore provide the health institution with an appropriate and well-founded solution, based on methods and technologies that have already been explored and adapted to solve the problem in question, and also stimulate new knowledge for the organization and the scientific community.

Proof of Concept – SWOT Analysis

The Proof of Concept (PoC) research methodology consists of a practical model that can prove or validate the concept established through analysis or even technical articles. Therefore, in this specific case of a project under development it goes on to verify whether this concept or theory might be successful and feasible and, on the other hand, being susceptible of exploitation in a useful way (Sergey, Alexandr, and Sergey 2015). A PoC allows to demonstrate in practice the concepts, methodologies and technologies involved in the elaboration of a given project and thus validate the proposed solution by proving its feasibility and usefulness for the purpose for which it is intended by defending its potential. In order to verify the viability, utility, quality and the efficiency of the mobile app, given the fact that not everyone has a smartphone and the functionalities addressed can be hard to implement, it was necessary to make a SWOT analysis, previous to the conclusion of the work, considering to be an advantage to improve the usability and the ergonomics (Pereira et al. 2013).

Strengths:

- No costs in the process of sending each notification;
- High usability, intuitive and easy to learn;
- Easy access to history medical data of the patients, as well as the hospital services in which the patient was;
- High scalability;
- Possibility to save notification on a digital agenda;
- Easy adaptability to different health institutions.

Weaknesses:

- Requires internet connection;
- Requires a smartphone or a computer.

Opportunities:

- Modernization and organizational development undergoing in hospitals;
- Potential to integrate other hospital services;

Threats:

- Lack of acceptance to resort to new technologies by generation who does not have access to a smartphone.

MOBILE APPLICATION - PRM

This project intends to develop a mobile application for the patients, in order to substitute the current notification system and give them access to useful information, as well as services from the hospital. It will be possible to be notified for appointments, surgeries, exams and tests. Besides, scheduling appointments will be a functionality as well. To complement that, a web application will be created for administration purposes, in order to manage the users, and monitor services utilization with a purpose to identify certain patterns to use them as decision making material to improve hospital services. Due to its versatility, the mobile application can be improved over time, with new functionalities, that can be predicted with a great potential tool to a win-win situation.

User Monitoring System – Web Application

The purpose of this application is to be able to manage the users, edit its information, creating users, deleting accounts, consult appointments of a certain user and include a Business Intelligence (BI) tab. This single page application will be developed in AngularJS, that simplifies the connection between the interface and the data storage place, using the ng-model not making necessary to be explicit about properties or dependencies. Since the application will

be running in a Linux machine, it will be used an Apache server called MAMP, with a MySQL Database (DB) management system. After that it will be developed a restful server, in PHP to support the access to the DB. In order to develop a fluid and dynamic application, the AngularJS framework will contribute to that (Darwin, Peter Bacon and Kozlowski 2013). The modularity and extensibility of this JavaScript framework allows the development of several and other new applications.

PRM – Mobile Application

The notification system of the hospital is made through PRM-AIDA, which is a module of AIDA working on CHP and Vila Verde hospitals. The goal is to substitute, in long term, this notification system for this mobile app. The users will have to register with their national health number and their own email, in order to have access to the app. Once the user has verified access, it will go to menu with four tabs:

1) User Details: page containing personal information, such as ID number, name, address, email, mobile number. In this page, the user can verify its info and update it whenever needed to. Although sections as name, ID number can only be edited by administration.

2) Notifications: page containing every notification received, grouped by type of medical event such as exams, appointments, surgeries, both organized by date received. There will be the options to mark notifications as "checked", or not, as well as save them to google/smartphone calendar.

3) History of events: page containing the history of medical events, grouped by category.

4) Agenda: page containing field to request medical appointments.

The application will be developed in Xamarin, which was considered to be the best framework to use, since it is multi-platform oriented, allowing to build native iOS and Android apps, making it possible to use in every smartphone available. It uses Ruby and C# as codebase. The application will use Apache as the Web server, PHP as the object-oriented language. MySQL will be used as the relational database management system (RDBMS) with the same data warehouse of the hospitals that is expected to implement this app. For the push notifications, it will be used the technology Firebase Cloud Messaging Android, as it is described by a model of implementation of Google. For that reason, it will be necessary to add API FirebaseMessaging using Gradel (Google 2016).

CONCLUSION

The number of patients missing appointments or any other medical events, can represent a huge loss of money for the health institution, so improving the notification system whilst providing other services to the patients will certainly have a positive impact on lowering the missing rate. Although this technology might seem to futuristic for now, only by starting to introduce this into the national healthcare system, will we get people to adapt and be able to embrace this technology in their everyday life. Given the studies found, it is expected that the rate on non-smartphone users will be significantly low in the next decade. Nevertheless, it came to the conclusion that this time of adaptation, will require that some users receive their notification through SMS or by email.

The implementation of this case study, will allow the development of a management tool, namely a user-friendly mobile app in order to improve patient relationship with the hospital. The mobile application itself can represent a difficult system to implement since not every patient has a smartphone, knows how to work with them or has access to mobile data. Nevertheless, it is believed that in no much more than a decade those problems will be residual. The Web application will make possible to gather information of the mobile app usage as well as behaviour patterns of the services requested, in order to better understand customer needs, in this case, patient's needs. In the coming years, the expansion of the mobile and web application is expected in terms of new functionalities such as a financial functionality in order to use medical insurances number or different paying methods to healthcare services.

ACKNOWLEDGEMENTS

This work has been supported by Compete: POCI-01-0145-FEDER-007043 and FCT within the Project Scope UID/CEC/00319/2013.

REFERENCES

Boulos, Maged, Steve Wheeler, Carlos Tavares, and Ray Jones. 2011. "How Smartphones Are Changing the Face of Mobile and Participatory Healthcare: An Overview, with Example from ECAALYX." *BioMedical Engineering OnLine* 10(1): 24. http://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-10-24.

Cardoso, Luciana et al. 2014. "The Next Generation of Interoperability Agents in Healthcare." *International Journal of Environmental Research and Public Health* 11(5): 5349–71. http://www.mdpi.com/1660-4601/11/5/5349.

Darwin, Peter Bacon and Kozlowski, Pawel. 2013. *AngularJS Web Application Development*. Packt Publishing.

Eysenbach, G. 2001. "What Is E-Health?" *Journal of Medical Internet Research* 3(2): e20. http://www.jmir.org/2001/2/e20/.

Google. 2016. "Set Up a Firebase Cloud Messaging Client App on Android." https://firebase.google.com/docs/cloud-messaging/android/client (November 20, 2017).

Gregor, Shirley, and Alan R Hevner. 2013. "Positioning and Presenting Design Science Research for Maximum

Impact." *MIS Quarterly* 37(2): 337–55.

Hevner, Alan, and Samir Chatterjee. 2010. 22 *Design Research in Information Systems*. Boston, MA: Springer US. http://link.springer.com/10.1007/978-1-4419-5653-8.

Lee, Jinhyung, Jeffrey S. McCullough, and Robert J. Town. 2013. "The Impact of Health Information Technology on Hospital Productivity." *The RAND Journal of Economics* 44(3): 545–68. http://doi.wiley.com/10.1111/1756-2171.12030.

Machado, Jose, Antonio Abelha, Jose Neves, and Manuel Santos. 2006. "Ambient Intelligence in Medicine." In *2006 IEEE Biomedical Circuits and Systems Conference*, IEEE, 94–97. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4600316.

Peixoto, Hugo, Manuel Santos, António Abelha, and José Machado. 2012. "Intelligence in Interoperability with AIDA." In , 264–73. http://link.springer.com/10.1007/978-3-642-34624-8_31.

Pereira, Rui, Maria Salazar, António Abelha, and José Machado. 2013. "SWOT Analysis of a Portuguese Electronic Health Record." In , 169–77. http://link.springer.com/10.1007/978-3-642-37437-1_14.

Sergey, A. Baydali, D. Budnickiy Alexandr, and A. Terehov Sergey. 2015. "Proof of Concept Center — A Promising Tool for Innovative Development at Entrepreneurial Universities." *Procedia - Social and Behavioral Sciences* 166(7): 240–45. http://www.science-gate.com/IJAAS/V4I7/Gustomo.html.

SPMS. "SPMS - Serviços Partilhados Do Ministério Da Saúde, EPE, MySNS." https://www.sns.gov.pt/apps/mysns/ (November 20, 2017).

University of Minho. "AIDA-Intelligence in Healthcare." https://sites.google.com/site/aidasuite/ (November 20, 2017).

Vaishnavi, Vijay, Bill Kuechler, and Stacie Petter. 2017. "Design Science Research in Information Systems." *Design Science Research in Information Systems:*: 1–66. http://www.desrist.org/design-research-in-information-systems/.

TRAJECTORY OPTIMIZATION

Real time trajectory matching and outlier detection for assembly operator trajectories

Bauters Karel [1,2]
Johannes Cottyn[1,2]
Hendrik Van Landeghem[1]
Department of Industrial Systems Engineering and Products Design, Ghent University [1]
Flanders Make [2]
B-9052, Zwijnaarde (Gent),
Belgium
E-mail: karel.bauters@ugent.be

KEYWORDS
Trajectory matching, dynamic time warping, anomaly detection, manual assembly

ABSTRACT

Flexible, reactive and adaptive manufacturing systems are a prerequisite to cope with the demand for low volumes of highly customized products of today's market. For years, manufacturing companies have been using real-time data capturing systems, such as RFID, to gather the necessary data to obtain insights in their production processes, mainly in the domain of quality control and inventory management. However, very few work has been done on monitoring an assembly operator during his work cycle in real-time. This paper presents a method to match operator trajectories, obtained through a multi-camera vision system, in real-time to predefined models. This way, the performance of the operator can be assessed online and problematic or anomalous work cycles can be detected. This information can then be used to support the operator in his pursuit for continuous improvement by pointing out improvement potential.

INTRODUCTION

Over the recent years, the consumer market has made a shift towards more customized and highly variant products. To answer these demands, production systems need to be very reactive and flexible. Contemporary flexible production systems are able to react and adapt their behavior to the current circumstances, based on real-time data and information obtained through a variety of sensor systems. Industrial applications of data gathering using sensor systems can be found in inventory control (Visich, et al. 2009) and job floor control (Arkan and Van Landeghem 2013). RFID technology for example, is used to provide accurate real-time process data which can be used to keep the manufacturing execution system (MES) up-to-date. These examples are all part of the current paradigm shift in manufacturing companies towards Industry 4.0. This fourth industrial revolution is based on digital transformation and cyber-physical systems to overcome the challenges posed by the changing market demands (Sogeti Labs, 2016).
Also within assembly line work stations, sensor systems are being used to monitor the progress of the production process and use this information to update the central production database (Wang 2012). This information can also be used to provide the operator with contextualized work instructions and information about the required inspections and test procedures. These systems are mainly focused on the product rather than on the operator performing the assembly tasks. However, gaining insights in the performance of the operator could provide the assembly line worker with critical information to support him in the continuous improvement of his work methods. Up until today, this kind of information can only be obtained through manual analysis of video-images or at the work station. These analysis methods are prohibitively time-consuming and therefore not tailored to the flexibility and reactivity requirements of contemporary work stations.
Recently researchers presented an analysis system for manual assembly work stations in which multiple cameras are used to track the operators' position in the work station throughout the complete work cycle (Bauters, et al. 2018). These trajectories are then classified into clusters based on their similarity in order to detect outliers or anomalous work cycles. These outliers are work cycles in which irregularities or problems took place and are therefore interesting subjects for further investigation. By pointing towards these anomalous work cycles, this system can significantly decrease the time needed to perform the manual analysis. Furthermore the system calculates a number of performance indicators and visualizes the operators' performance indicators in an operational dashboard to unveil improvement potential.
One of the disadvantages of this system is the fact that the analysis of the video-images is still done offline. This is because the classification method relies on the dynamic time warping algorithm (DTW) to calculate the similarity between different trajectories. This method yields better classification results for this application than other existing similarity measures (Bauters, et al. 2018). However, calculation time for DTW is exponential, making it impossible to calculate the warping between a large number of rather long trajectories in real-time.
In this paper we present a method to match an operators' trajectory to a number of pre-recorded model trajectories in real time. This method makes it possible to detect anomalies in real-time and immediately ask the operator for feedback on what exactly happened during that particular work cycle. Furthermore it enables us to assess to operators' performance in real-time and suggest improvements to his work procedure.
The remainder of this paper is structured as follows. In the next section, a description of the data sets used in this research is given. Afterwards the real-time trajectory matching method is explained before presenting some results. Finally, the

conclusions of this research are presented and some ideas for further research are proposed.

DATA SETS

Two different data sets have been used to validate the trajectory matching methodology. In this section, these data sets are briefly discussed.

Experimental data set

The first data set is created by recording a human operator performing simulated assembly tasks in a laboratory setting. The parts produced in this experiment consist of a Duplo® base block on which different patterns of Lego® blocks are placed. The Duplo® base blocks are delivered to the work station using a conveyer belt that mimicks a moving assembly line. The Lego® blocks are stored in different locations in a picking rack equipped with a pick-to-light system. This way, each different product produced yields a different pattern or trajectory followed by the operator (Bauters, et al. 2018). An overview of the laboratory setting is given in Figure 1.

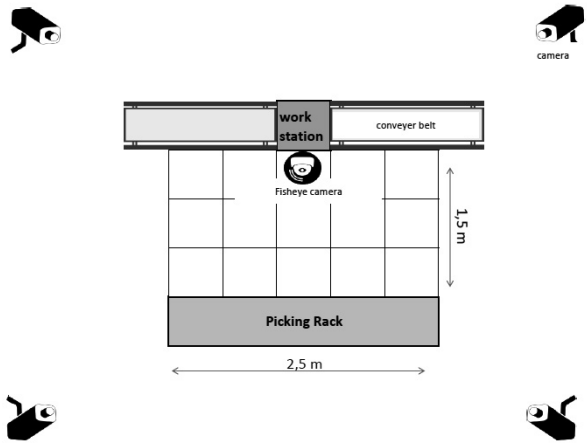


Figure 1: overview of the laboratory setting for data set 1

In this case, a system of multiple cameras (5 in total) was used to determine the operators' position throughout his work cycle. To do this, the principle of voxel carving is used to create a visual hull of the operators' body in every frame of the video sequence as described by several researchers (Laurentini 1994) (Slembrouck, et al. 2015). The position is then determined by projecting the center of mass of this visual hull on the ground plane (x, y). To filter the noise in the resulting trajectories, a Gauss kernel smoothing approach was implemented. Figure 2 shows the resulting visual hull, based on the images obtained through the five cameras.

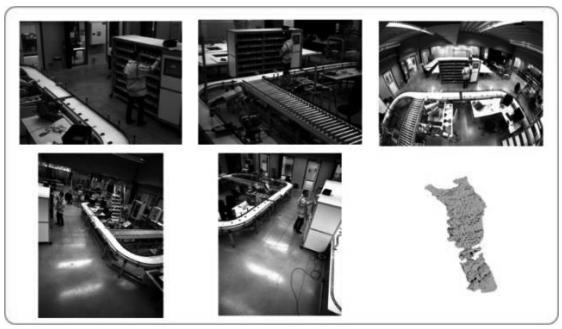


Figure 2: Output of the multi-camera system

This dataset contains 22 different trajectories. Two different patterns can be observed as well as 2 anomalous work cycles. In this paper, this dataset is mainly used to show the ability of the developed methodology to accurately distinguish normal trajectories from anomalous work cycles in real-time.

Omni1

The second data set used in this research is a data set containing over 200 trajectories of people walking through a lab as described in (Morris and Trivedi 2011). All trajectories were recorded within a 24 hour period without the knowledge of the people entering and leaving the laboratory. This dataset was constructed with one single omni-directional camera. Figure 3 provides an overview of the laboratory setting.

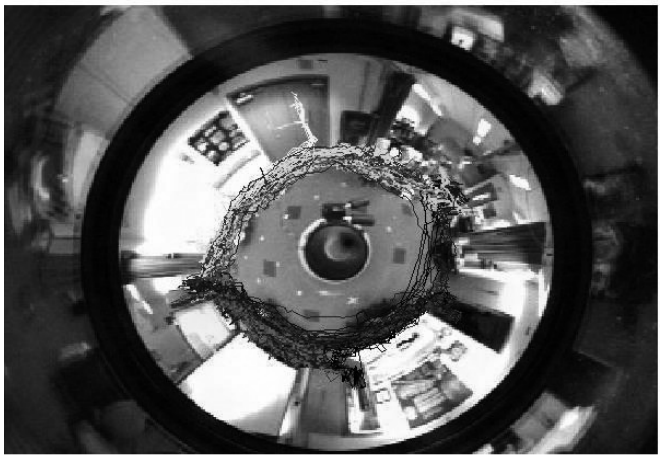


Figure 3: overview of the laboratory setting for data set 2

This dataset contains 7 different activities or patterns in total. All 206 trajectories in this data set are labeled, meaning that for each trajectory the class of the activity performed by the subject is known. In this research, this data set was mainly used to validate whether the trajectory matching methodology is capable of handling a higher number of trajectory models.

Real-time trajectory

For both data sets discussed above, determining the location of the operator/human subject is done off-line. Indeed, obtaining robust and accurate location data from video-images in real-time at high sample rates with existing image processing algorithms remains a challenging task. Allthough a lot of research in the field of image processing is being performed and real-time localization algorithms based on video images are expected to become available in the near

future. Also, there exist a number of different sensors which could deliver exactly the same information.
To overcome this problem, in this research we choose to use trajectory data which is calculated off-line and feed a new location to the system at fixed time intervals, which simulate the frame rate of the cameras. This way, we are able to prove that the developed methodology is capable of handling real-time data once it becomes available.

METHODOLOGY

In this section, the existing method for off-line classification of work cycle trajectories is briefly discussed. Afterwards the challenges encountered when using this method in real-time are clarified and the newly developed method for real-time trajectory matching is presented.

Off-line classification

The inherent variation in the assembly process leads to trajectories that vary in speed and length, even if the tasks performed by the operator are the same. Therefore, one can not simply use the Euclidean distance between concurrent points in two trajectories as a distance measure. To overcome this challenge, trajectories are compared using dynamic time warping (DTW). DTW is a technique that originally was implemented in speech recognition applications, but by now it has successfully been used to cope with deformations in all kinds of (multi-dimensional) time-dependent data (Müller 2007). The idea behind DTW is to find an optimal warping path that minimizes the distance between two trajectories, taking into account a number of warping constraints. The DTW distance can recursively be calculated: given two time series $X := (x_1, x_2, \dots, x_N)$ and $Y := (y_1, y_2, \dots, y_M)$ with respective lengths $N, M \in \mathbb{N}$, the cost of the alignment between these two time series can be calculated as follows:

$$C(X_i, Y_j) = \delta(x_i, y_j) + \min\{C(X_{i-1}, Y_{j-1}), C(X_{i-1}, Y_j), C(X_i, Y_{j-1})\}$$

where X_i an Y_j are the respective subsequences $(x_1, x_2, \dots, x_i)$ and $(y_1, y_2, \dots, y_j)$ and $\delta(x_i, y_j)$ is the Euclidean distance between two two-dimensional points x_i and y_j. $C(X_n, Y_m)$ determines the DTW distance between the two trajectories. Figure 4 visualizes how this alignment works for two one-dimensional time series. More detailed information on the implementation can be found in (Bauters, et al. 2018).

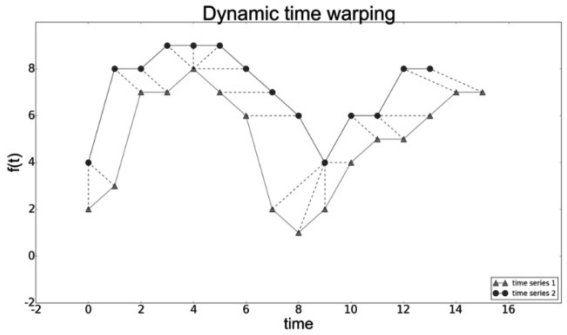


Figure 4: example of DTW alignment between two time-series

The distinction between regular or normal work cycles and anomalies for a set of trajectory sequences, is made based on a hierarchical clustering procedure. Hierarchical clustering methods are used to find a similarity structure in a dataset by initially dividing the data set in n clusters, with n being the number of objects in the data set. The most similar clusters are then merged into a new cluster and this is repeated until all objects are grouped into the same cluster. The similarity structure of the data set is typically visualized in a dendrogram Figure 5, showing the sequence in which clusters are merged. The distance between two merged clusters is indicated by the height of the links in the dendrogram.

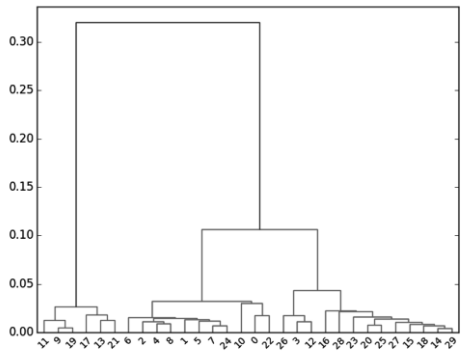


Figure 5: example of a dendrogram

To determine where to cut the dendrogram and thus decide whether the objects in two clusters actually represent the same or a different activity, an adapted version of the permutation testing method proposed by Bruzesse (Bruzesse and Vistocco 2015) is applied. This procedure is based on the assumption that, if two clusters contain similar objects, the distance between two-randomly sampled sets of objects from these clusters will not be significantly different from the distance between the original clusters.
Applying this method on a set of work cycle trajectories, this set is divided into groups or clusters of similar trajectories and single-item clusters which we call outliers or anomalies. The similar trajectories all represent the same assembly process under normal circumstances and can be used to build a model that serves as a template for real-time trajectory matching later on. To build this model, an average trajectory of all sequences in the cluster is calculated. This average trajectory is iteratively calculated using the DBA algorithm as proposed by Petitjean (Petitjean, et al. 2014). The resulting model for one of the 7 patterns in the Omni1 data set is shown in Figure 6.

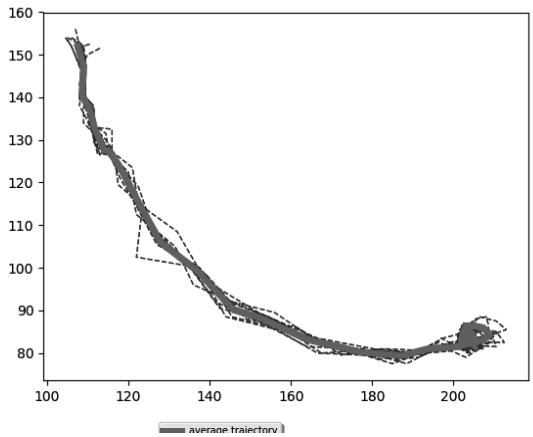


Figure 6: average trajectory model

Real-time trajectory matching

A naïve approach to try to match sequences to the previously calculated models, would be to calculate the DTW distance between the incoming sequence and all of the models, each time a new point is added to the new sequence. The sequence can then be matched to the model resulting in the minimal DTW distance. However, there are a number of downsides to this approach: (1) The calculation of the DTW has a O(n*m) computation time complexity, where n is the length of the incoming sequence and m represents the length of the model. Performing the DTW calculation for every new datapoint, leads to computation times far exceeding the framerate of the camera system, especially if the sequences are becoming longer and the number of trajectory models is high. (2) The incoming trajectory sequence only represents a fraction of the full work cycle. One can rightfully question wether matching such a partial sequence to the model of a complete work cycle actually provides meaningful results.

To overcome these challenges another approach was taken, based on Keogh's lower bound calculation for DTW (Keogh and Ratanamahatana 2005). The idea behind the approach is to calculate a lower bound for the DTW distance between the incoming sequence and a subsequence of the model that has the same length as the incoming trajectory. Based on this low-complexity lower bound calculation, it is possible to eliminate trajectory models from the set of possible candidate matches.

To calculate the Keogh lower bound, a bounding envelope is constructed for each of the trajectory models (Capitani and Ciaccia 2006). Let $M(a_1, \ldots, a_m)$ be a trajectory model of length m and $Env(M)$ is the envelope around M defined by two time series $Up(M)$ and $Low(M)$. Then $Up(M)$ and $Low(M)$ as follows:

$$Up(M) = \max\left(M_j \middle| j \in [\max(1, i-b), \min(m, i+b)]\right)$$
$$Low(M) = \min\left(M_j \middle| j \in [\max(1, i-b), \min(m, i+b)]\right)$$

In other words, $Up_i(M)$ and $Low_i(M)$ are respectively the maximum and minimum values of M in the interval [i-b, i+b], where b is a user-defined parameter and taking into account the border effects. The squared Keogh LB distance between a subsequence M_n of the model M and an incoming sequence S_n of length n, is defined as follows:

$$LB_{Keogh}(Env(M), S)^2 = \sum_{i=1}^{n} \begin{cases} \left(S_i - Up_i(M)\right)^2 & if\ S_i > Up_i(M) \\ 0 & if\ Low_i(M) \le S_i \le Up_i(M) \\ \left(S_i - Low_i(M)\right)^2 & if\ Low_i(M) > S_i \end{cases}$$

It can be proven that the Keogh LB distance is a lower bound for the DTW distance for 1-dimensional time series. However, the trajectory sequences under investigation in this case, are 2-dimensional. This issue can be overcome by constructing separate envelopes for the x and y component of the model and performing the Keogh LB distance calculation on both the x and y component of the incoming sequence. Rath and Manmatha (Rath and Manmatha 2002) proved that in this case:

$$LB_{Keogh}(Env(M_x), S_x)^2 + LB_{Keogh}\left(Env\left(M_y\right), S_y\right)^2 \le DTW(M_x, S_x)^2 + DTW\left(M_y, S_y\right)^2 = DTW(M, S)^2$$

The Keogh LB calculations for the respective x- and y-component of a sequence and model in the Omni1 data set are visualized in Figure 7 and Figure 8.

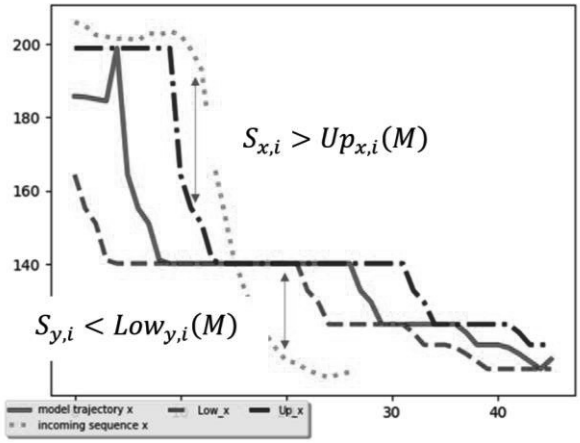


Figure 7: Keogh LB calculation x component

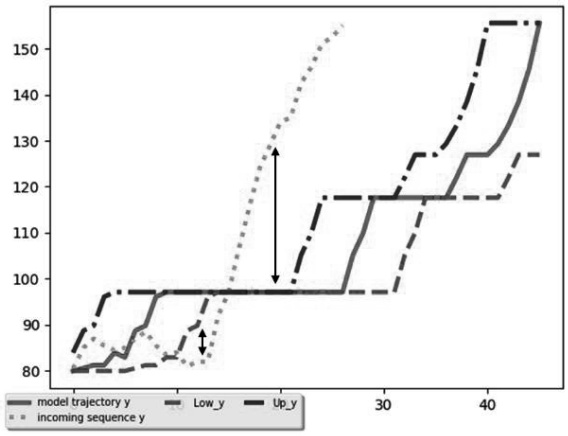


Figure 8: Keogh LB calculation y component

This lower bound calculation requires less computation time than the full DTW calculation. Based on this knowledge, the real-time trajectory matching methodology was developed. In this methodology, the lower bound distance between the incoming sequence and an even long subsequence of all the candidate models is calculated. For the model yielding the best lower bound distance, the DTW distance is calculated and saved as the *best_so_far* distance. Subsequently the LB distances of all candidate models are compared to this *best_do_far* DTW distance and candidate models for which the LB distance is higher than the *best_so_far* are eliminated from the set of candidate models, under the assumption that those models are unlikely to provide a good match for the incoming sequence. The outline of the method is provided in Figure 9.

Real_Time_Trajectory_Matching(M: [(x_1, y_1), …, (x_m, y_m)], S: [(x_1, y_1), …, (x_n, y_n), …])

1. **Initialization**
2. Best_so_far ← inf.
3. Incoming_traj ← []
4. **Start**
5. **For** new data point:
6. LBs := [Keogh_LB for model in traj_models]
7. **Best_model** ← traj_models[min(LBs)]
8. Best_so_far = DTW(Best_model)
9. **For** model in traj_models:
10. **If** LB>Best_so_far:
11. **Remove** from traj_models
12. **End if**
13. **End for**
14. **Return** best_model
15. **End for**

Figure 9: outline of the real time trajectory matching algorithm

As shown in Figure 9, the algorithm uses the Keogh LB to estimate what the best matching model is. This way, only one DTW calculation needs to be performed per new incoming data point. The algorithm was then further sped up by implementing a warping window for the DTW calculation.
To detect outliers, the average distance between the average trajectory of the best matching model and all the trajectories used to build up that model (*avg_dist*), is calculated together with the standard deviation σ on those distances. Once the incoming trajectory is fully completed, the DTW distance between the new trajectory and the average trajectory of the best matching model is compared to *avg_dist.* If DTW(incoming, avg_traj) > avg_dist + z.σ , the incoming trajectory is considered to be an outlier.

RESULTS

Experimental data set

The first data set contains two regular trajectories, one in which parts are only picked on the left side of the rack and one for which the necessary parts are stored on the left and right side of the picking rack. In the anomaluous work cycles, the operator travels back-and-forth alongside the rack to set right a picking mistake. The models and outliers are visualized in following Figure 10.
For this data set, the proposed method was able to classify all incoming segments correctly. For every segment, the average calculation time per frame was logged. The average calculation time is 0.07710 seconds, with a maximum time of 0.089 seconds. Twenty frames per second are obtained through the camera system. However, trajectories can safely be downsampled up to a factor 10, without compromising the classification results (Bauters, et al. 2018). Therefore, it can be concluded that the proposed method is capable of performing real-time trajectory matching on this particular data set.

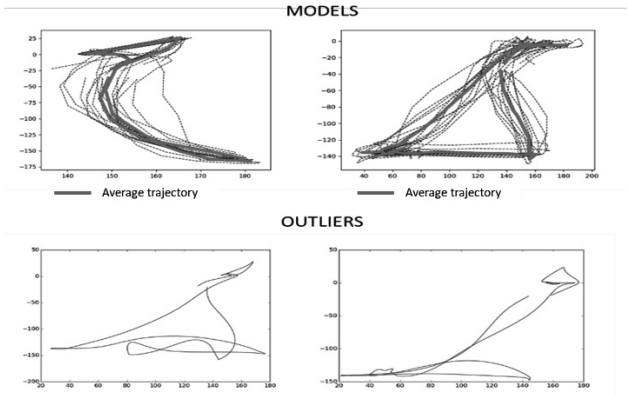


Figure 10: models and outliers experimental data set

Omni1 data set

This data set contains seven models. Therefore one would expect the average calculation time per frame to be higher. The opposite however is true. Due to the fact that the trajectories in this data set are generally shorter than the ones in the first data set, the average calculation time per frame only amounts up to 0.0678 seconds, with a maximum of 0.077 seconds.
On the downside, applying the proposed method on this data set only yields an accuracy 94.3% percent. In the experiments described in this paper, no false negatives (no matching pattern was found when it does exist) were detected. The 5.7% mistakes detected are trajectories that are matched to the wrong model (false positive). This can be explained by the fact that some of the models in this data set share common subtrajectories. Sometimes this results in a slightly higher similarity of the incoming trajectory to a subtrajectory of the wrong model. This occasionally leads to the preliminary elimination of the actual best matching model.

CONCLUSIONS AND FUTURE RESEARCH

In this paper, a method for real-time trajectory matching and outlier detection was presented. The aim of the method is to develop a system that is able to monitor assembly line work station operators and detect problems and mistakes in real-time. The monitoring of the operator is done using a multi-camera video analysis system. By detecting difficulties and problems on-line and linking this to real-time operator feedback and video images, a vast amount of valuable information for improving the process and/or redesigning the work station is created. Until today, this kind of information can only be obtained through manual analyses of video recordings and interviews with operators, which are heavily time-consuming.
The developed outlier detection method is based on dynamic time warping. The Keogh lower bound concept was used to speed up the similarity measurement to enable real-time outlier detection. The method was validated on two different experimental data sets. Results show that the proposed system is capable to accurately detect outliers in real-time.
Further research will focus on accelerating the video analysis in order to evolve to a (near) real time analysis tool. The vision technology and the 3D-model of the operator created

by the visual hull method can also be used to perform an ergonomics analysis of the work cycle. This would be a valuable extra to the system.

REFERENCES

Arkan, Ihsan, and Hendrik Van Landeghem. 2013. "Evaluating the performance of a discrete manufacturing process useing RFID: A case study." *International journal of Robotics and Computer-integrated manufacturing* 29 (6): 502-512.

Bauters, Karel, Johannes Cottyn, Dieter Claeys, Maarten Slembrouck, Peter Veelaert, and Hendrik Van Landeghem. 2018. "Automated work cycle classification and performance measurement for manual work stations." *International journal of robotics and compuetr-integrated manufacturing.*

Bruzesse, D., and D. Vistocco. 2015. "DESPOTA: Dendrogram slicing through a permutation test approach." *Journal of classification* 32 (2): 285-304.

Capitani, Paolo, and Paolo Ciaccia. 2006. "Warping the time on data streams." *Data & knowledge engineering* 62: 438-458.

Keogh, Eamonn, and Chotirat Ann Ratanamahatana. 2005. "Exact indexing of dynamic time warping." *Knowledge and information systems* 7 (3): 358-386.

Laurentini, A. 1994. "The visual hull concept for silhouette-based image understanding." *IEEE transactions on Pattern Analysis and Machine Intelligence* 16 (2): 150-162.

Morris, Brendan Tran, and Mohan Manubhai Trivedi. 2011. "Trajectory learning for activity understanding: Unsupervised, Multilevel, and Long-Term adaptive approach." *IEEE Transactions on pattern analysis and machine intelligence* 33 (11): 2287-2302.

Müller, M. 2007. "Dynamic Time Warping." *Information Retrieval for music and motion* (Springer) 69-84.

Petitjean, F., G. Forestier, G. Webb, A. Nicholson, Y. Chen, and E. Keogh. 2014. "Dynamic time warping averaging of time series for faster and more accurate classification." *IEE international conference on data mining.* Shenzen: IEEE.

Rath, T., and R. Manmatha. 2002. "Lower-bounding of dynamic time warping distances for multivariate time series." Tech Report MM40, University of Massachusetts.

Slembrouck, M., J. N. Castaneda, G. Allebosch, D. Van Cauwelaert, P. Veelaert, and W. Philips. 2015. "Gigh performance multi-camera tracking using shapes-from-silhouettes and occlusion removal." *Proceedings of the 9th international conference on distributed smart cameras.* New York.

Visich, J. K., B. M. Khumawala, P. M. Reyes, and S. Li. 2009. "Empirical evidence of RFID impacts on supply chain performance." *International journal of operations & production management* 29 (12): 1290-1315.

Wang, L. 2012. "RFID-based information technology abd labagement in agriculture and food supply chains." *International journal of information technology and management* 11 (3): 225-239.

WEB REFERENCES

Industry 4.0: Factory to smactory, disruption meets the manufacturing industry.
http://labs.sogeti.com/industry-4-0-factory-to-smactory-disruption-meets-the-manufacturing-industry
Accessed on: 2018-05-03

A CAPACITY STUDY FOR VESSEL TRAFFIC USING AUTOMATIC IDENTIFICATION SYSTEM DATA

Matthias Deceuninck, Kurt De Cock, Stijn De Vuyst
Department of Industrial Systems Engineering and Product Design
Ghent University
Technologiepark 903, 9052 Zwijnaarde,
Belgium
email: matthias.deceuninck@ugent.be

Mark Vantorre, Katrien Eloot
Maritime Technology Division
Ghent University
Technologiepark 904, 9052 Zwijnaarde
Belgium

KEYWORDS
Simulation, Maritime traffic, Capacity study

ABSTRACT

In this study, we created a simulation model to assess the overall impact of implementing a one-way traffic policy due to construction works. The inputs of the simulation model are found by performing statistical analysis on data from the Automatic Identification System (AIS). The aim of this study is twofold: (a) map the vessel traffic during the reference period and (b) analyse the congestion for the new traffic conditions. We use a non-homogeneous Poisson process with piecewise linear intensity to model the arrival process. For scenarios with varying arrival intensities, we compare the vessels' waiting times as well as the maximum queue lengths. The latter is important for upstream traffic since there are space constraints.

INTRODUCTION

In an effort to improve the accessibility of city and port, the Flemish Government launched the Master Plan 2020 to unscramble the traffic knot in the Antwerp region (The Oosterweel Link, 2018). One of the projects involves the construction of canals tunnels passing under the Albert Canal, one of Belgium's most important and busy waterways. Due to these construction works, two-way traffic will no longer be possible in a section of the canal. In this respect, the motivation behind this study is to investigate the impact of implementing a one-way traffic policy. To forecast the vessel traffic in the canal, AIS data was collected. Since 2002 most vessels are required to carry an AIS transceiver on board which broadcasts information such as position, speed and direction through dedicated VHF frequencies. This information formed the inputs of our simulation model.
Simulation methods have been widely applied for the modelling of vessel traffic on waterways because they enable studies of more complex systems. In the literature, extensive simulation models have been developed to investigate the effects of numerous factors on performances measures such as capacity and waiting times. Golkar et al. (1998) used simulation to evaluate the capacity of the Panama Canal under different operating conditions. In Thiers and Janssens (1998), a detailed maritime traffic simulation model was developed for the port of Antwerp including navigation rules, tides and lock operations. Merrick et al. (2003) used simulation to perform a traffic density analysis in the San Francisco Bay area. The model tried to assess the overall impact of an expansion in ferry services which was a proposal of the California legislature. The Istanbul Channel has also received a lot of attention (e.g. Köse et al. 2003, Almaz et al. 2006, Özbaş and Or 2007). For example, Köse et al. (2003) developed a simulation model to test the effect of arrival intensity on the waiting times.
This paper is divided into five sections. In the next section, we discuss the data analysis. The third section describes the simulation model used in this paper. Various scenarios are investigated in the fourth section. Finally, conclusions are drawn in the last section.

METHODOLOGY

In this study, we employed the AIS data collected in the Albert Canal during the month of August 2016. The dataset contains the AIS data of all vessels passing one of the six intersections depicted in Figure 1. For each passage, the following statistics were registered:name, width in meter (Ship Beam), length in meter (Ship Length), speed, position and UTC (Coordinated Universal Time). After data cleaning, 17312 entries from 897 unique vessels were kept for analysis. The dimensions and traffic types of the vessels are shown in Figure 2. The average length of a vessel is equal to 82.08m. During the reference period, 61% of the vessels were cargo ships (AIS ship type numbers 70-79), 18% tankers (80-89) and only 1% passenger ships (60-69).

Figure 1: View of intersections where AIS data is collected.

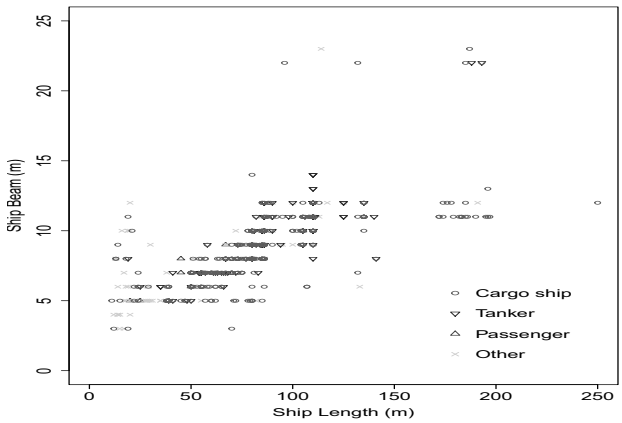


Figure 2: (Corrected) Dimensions and ship type of the vessels observed during the reference period.

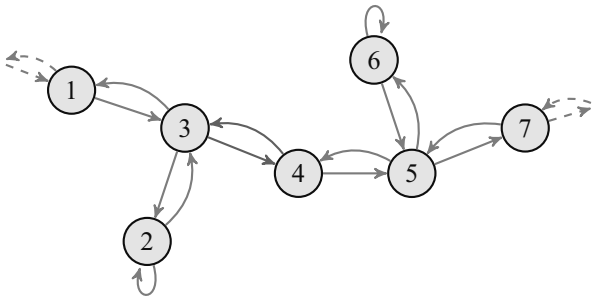


Figure 3: Graphic representation of possible movements.

A directed graph can be used to represent the traffic in the waterway. This is shown in Figure 3 where the arcs represent all the possible movements of the vessels and the nodes subdivide the waterway into sections of homogeneous capacity. Vessels enter or leave the system at a boundary node (nodes 1 and 7) or at one of the docks (nodes 2 and 6). We want to stress that the nodes do not fully correspond with the intersections in Figure 1. The arc between nodes 3 and 4 corresponds with the narrowed waterway where one-way traffic will be implemented. In the remainder of this paper, we will denote this section of the channel as the construction zone. The length of the construction zone is approximately equal to 880m. Furthermore, we will refer to vessels moving in the direction 1→7 (7→1) as upstream (downstream) traffic.

For the analysis, we are mainly interested in the traffic through the construction zone as this will be the section with congested traffic. Figure 4 depicts the vessel traffic for the first week of august. The black lines denote the length of the vessel that is passing the section at that moment in time, with the positive and negative axis respectively corresponding to upstream and downstream traffic. To get an idea of the traffic intensity over the course of a week, we also plot the KDE (kernel density estimation) for the upstream/downstream (red, solid) and total (blue, dashed) traffic. For n ships with arrival times t_i (in hour) and length l_i (meter), $i = 1, \ldots, n$, with mean $\bar{l}$, the instantaneous arrival rate, expressed in ship meters per hour (shm/h), at time t is estimated as

$$\hat{\gamma}_h(t) = \frac{1}{nh} \sum_{i=1}^{n} \frac{l_i}{\bar{l}} K(\frac{t - t_i}{h}), \tag{1}$$

with the so-called kernel $K(t)$ being a 'cosine' window. A crucial parameter is the bandwidth h (in hours) since this parameter determines the smoothness of the resulting estimate. Intuitively one wants to choose h as small as the data allows. A small h results in low bias but increases the variance of the estimates. In Figure 4, we set h equal to 1 hour and find for the first week a mean of 311 shm/h with a maximum of 716shm/h. We can clearly observe some daily seasonality with multiple peaks. Furthermore, it can be seen that there is significantly less traffic on Sundays.

Finally, Table 1 gives an overview of the vessel traffic through the construction zone coming from all possible directions. In the last row, we can see that the total traffic through the construction zone has a mean of 339.37 shm/h with the highest arrival intensity being 533.83 ship meters in 30 minutes ($h = 0.5$h).

Path	mean ships/h	peak(.5h)	peak(1h)	peak(2h)
$1 \rightarrow 4$	1.95	7.92	5.95	4.63
$2 \rightarrow 4$	0.02	0.72	0.61	0.35
$6 \rightarrow 3$	0.08	1.40	0.82	0.50
$7 \rightarrow 3$	1.86	5.78	4.97	3.93
$3 \leftrightarrow 4$	3.91	14.04	11.01	9.25

Path	mean shm/h	peak(.5h)	peak(1h)	peak(2h)
$1 \rightarrow 4$	176.88	717.11	538.16	418.86
$2 \rightarrow 4$	0.94	42.11	35.79	20.23
$6 \rightarrow 3$	5.78	101.66	59.53	36.77
$7 \rightarrow 3$	155.77	595.22	440.36	355.43
$3 \leftrightarrow 4$	339.37	1069.66	903.06	758.57

Table 1: Vessel traffic through the construction zone during the reference period.

SIMULATION MODEL

The simulation software package FlexSim 2016 is used for the implementation of the simulation model of the maritime traffic. Simulation allows us to analyse and compare the results of different scenarios. In this section, we describe the arrival process, traffic control measures and other features of the model.

Arrival process

The following input data are generated for each vessel entering the system from a boundary node: arrival time, dimensions, speed and path. Instead of using the real data directly, we generate artificial scenarios where all input factors are randomly generated based on the probability distributions obtained from the data. This allows us to investigate scenarios

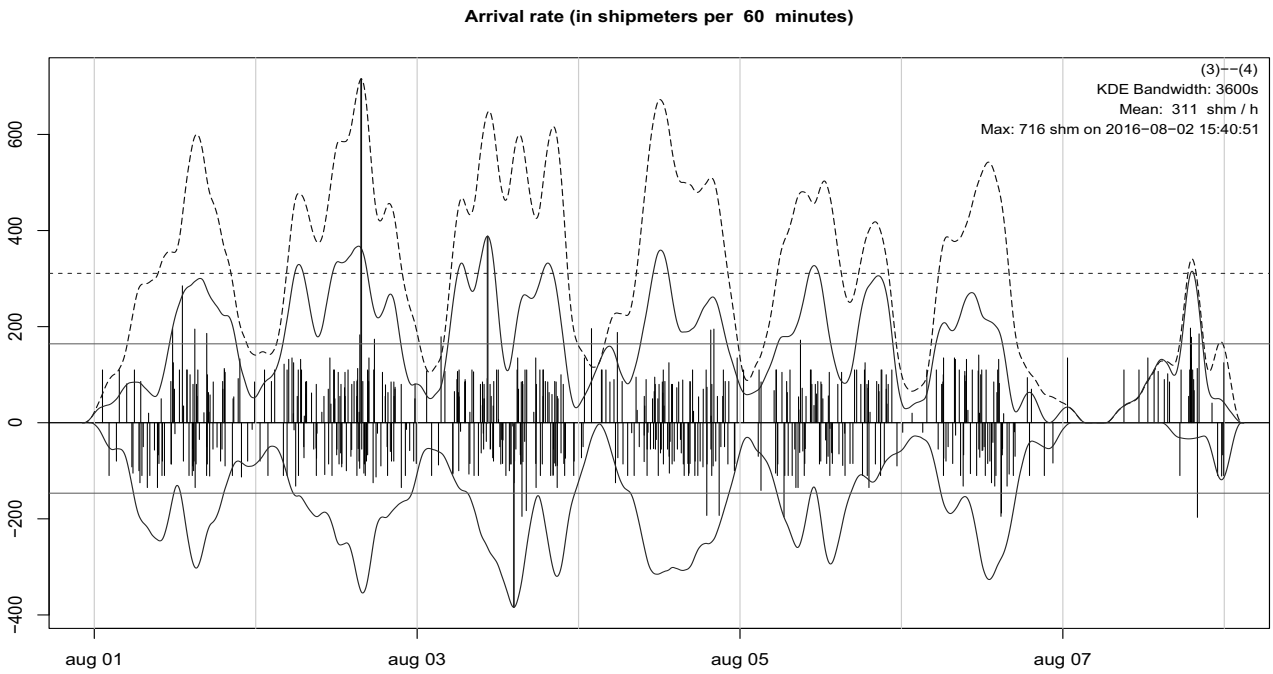


Figure 4: Upstream (positive vertical axis) and downstream (negative vertical axis) vessel traffic through the construction zone combined with their kernel density estimates ($h = 1$ hour). The red (solid) lines correspond with the upstream and downstream traffic while the blue (dashed) lines give the mean and KDE of the total traffic.

in which the traffic has similar characteristics as in Figure 4 but with a different intensity: $\lambda = \alpha\lambda^{\text{ref}}$ with the multiplier α varying from 1 to 1.5. This may be necessary since follow-up studies found that the traffic was considerably higher during the subsequent months (+11%).

An important question that arises is the modelling of the non-stationary arrivals. As discussed earlier, a time-dependent arrival process is observed from the data with both daily and weekly seasonalities. Let $A_k(t)$ denote the arrival process at node k. We assume that $A_k(t)$ follows a non-homogeneous Poisson process with a piecewise-linear intensity function: $A_k(t) \sim \mathcal{P}(\lambda_k(t))$. That is, the interarrival times are independent and exponentially distributed with intensity $\lambda_k(t)$. A piece-wise linear function is chosen to simplify the model as such complex time series are prone to over-fitting for a small dataset.

Our approach consists of partitioning each weekday into 1-hour intervals, calculate for each hour the average intensity and then interpolate between the obtained values. Let $\lambda^{\text{ref}}_{k,j}$ denote the average traffic intensity during the jth interval at node k, then for $j = 1, 2, \ldots 24$, we have

$$\lambda^{\text{ref}}_{k,j} \frac{1}{D_{\text{ref}}} \sum_{i=1}^{n} \sum_{d=1}^{D_{\text{ref}}} \mathbb{1}_{\{j-1 \leqslant t_i \,(\text{mod } 24) < j\}}\,, \tag{2}$$

with D_{ref} the number of days in the reference period (excluding weekends) and $\mathbb{1}_{\{.\}}$ the indicator function which evaluates to 1 if its argument is true and to 0 if this is not the case. We exclude weekends from the dataset because there is generally less traffic and we are interested in the performance

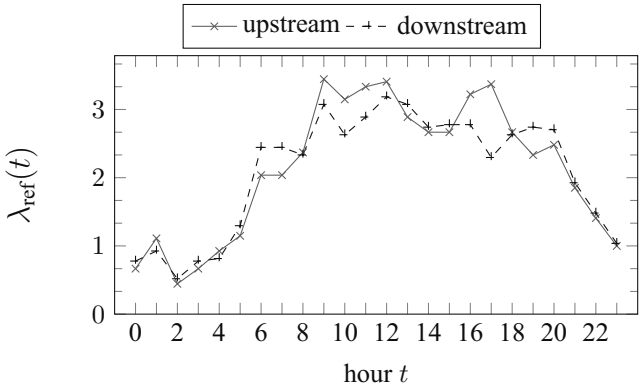


Figure 5: Intensity function $\lambda_{\text{ref}}(t)$ estimate for upstream and downstream vessel traffic (excluding weekends).

measures during congestion. Using linear interpolation, the instantaneous intensity function $\lambda^{\text{ref}}_k(t)$ is then equal to

$$\lambda^{\text{ref}}_k(t) = \lambda_{\text{ref},j^*} + (t - j^*)(\lambda^{\text{ref}}_{k,j^*+1} - \lambda^{\text{ref}}_{k,j^*}) \quad t \geq 0\,, \tag{3}$$

with j^* rounded down to the nearest hour: $j^* = \lfloor t \,(\text{mod } 24) \rfloor$. The resulting intensity function for upstream and downstream traffic are given in Figure 5. Most traffic is between 6am and 8 pm.

Traffic control measures

Traffic through the construction zone is reduced to one lane. Temporary traffic lights are installed at nodes 3 and 4 which

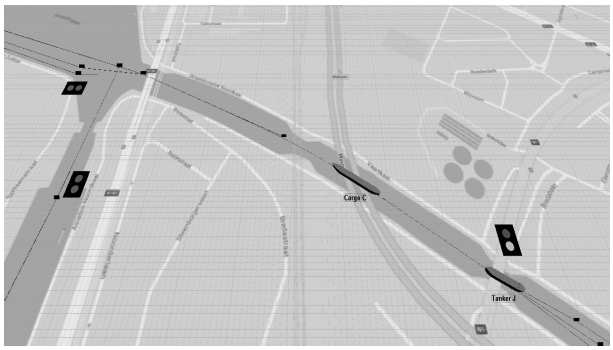


Figure 6: A snapshot of the simulation model.

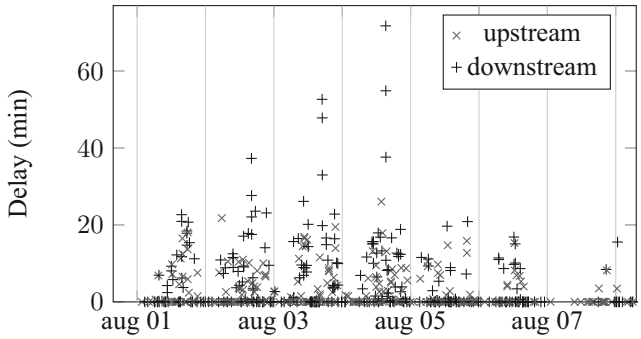


Figure 7: Delay times for base case scenario.

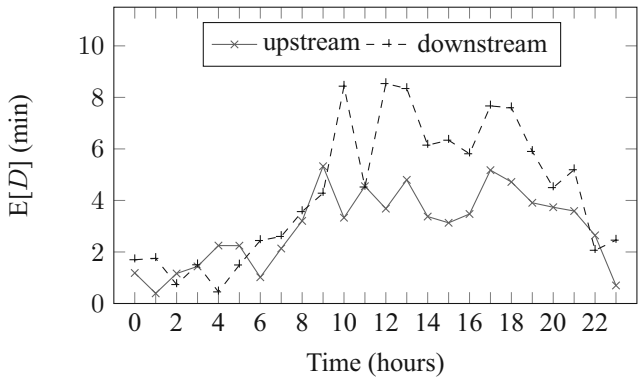


Figure 8: Average delay times during weekdays for base case scenario.

are manually operated to maximize the throughput. The operators always try to empty the waiting queues completely in one go to avoid that vessels need to perform multiple departure and stopping manoeuvres (once for each green-red cycle). Vessels that arrive at a non-empty queue or red light enter the queue at the tail and leave the queue according to a FIFO policy. It is assumed that the spacing between vessels in the queue is equal to 2m and increases to at least 30m for moving vessels. It is further assumed that vessels are moving with a uniform speed along a certain arc and that speed changes are immediate. In order to avoid nuisance waves, a speed limit of 5 kph is set in the entire working zone. Finally, a lower speed is also assumed for vessels coming from one of the docks to take into account the time that is needed to perform turning manoeuvres. A snapshot of the simulation model is given in Figure 6.

Performance measures

To assess the overall impact of the new traffic conditions, the following performance measures are considered relevant:

- The vessels' delays D at the traffic lights.
- The length L (in ship meters) of each queue.

RESULTS AND ANALYSES

In this section, we analyse the system for different scenarios. We first look at what happens when the arrivals exactly correspond with the reference period (approx. 2800 vessels). Figure 7 depicts the delay times for the first week of this base case scenario. The maximum delay for this time period are respectively equal to 25 and 71 minutes for upstream and downstream traffic. Most vessels do not experience any waiting and the average delays are respectively equal to 3.0 and 4.2 minutes. Obviously, the vessels' delays depend on the arrival time. Figure 8 shows for each moment of the day the delay that a vessel may expect. It can be seen that the average delays during daytime are approximately 4 and 7 minutes for respectively upstream and downstream traffic, while less than 2 minutes during night time.

Next, we generate data using a non-homogeneous Poisson process with time-varying rate $\alpha\lambda_k^{\text{ref}}(t)$ as given by Equation (3). For each scenario, a simulation time of 124 weekdays is used to estimate the performance measures of which 4 weekdays are used as warm-up period. Figure 9 depicts the distribution of the queue length for upstream traffic. It can be seen that 90% of the time the queue is shorter than 100m (log Prob[$L > 100$]= -1). The maximum queue lengths that we encountered during the 4-month simulations were less than 800m for $\alpha \leq 1.2$. Given these results, we thus do not expect any problems regarding the space constraint (queue space ≈800m) for upstream traffic in the harbour when the arrival intensity increases less than 20% compared to our reference set. For higher arrival intensities, the operator may need to give priority to upstream traffic to avoid a crowded queue during peak hours.

Finally, Figures 10 and 11 respectively present the delay time distributions for upstream and downstream traffic. It can be seen that approximately 10% of the vessels (log Prob[$D > t$]= -1) have a delay of more than 20 minutes and less than 1% a delay longer than 40 minutes. Long delays are more common for downstream traffic. This can be explained by the fact that a higher priority is given to upstream traffic because of space constraints.

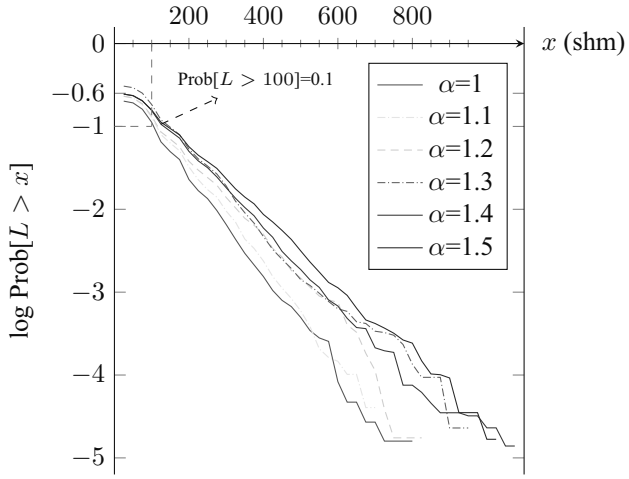


Figure 9: Distribution of the queue length for upstream traffic in function of the arrival intensity.

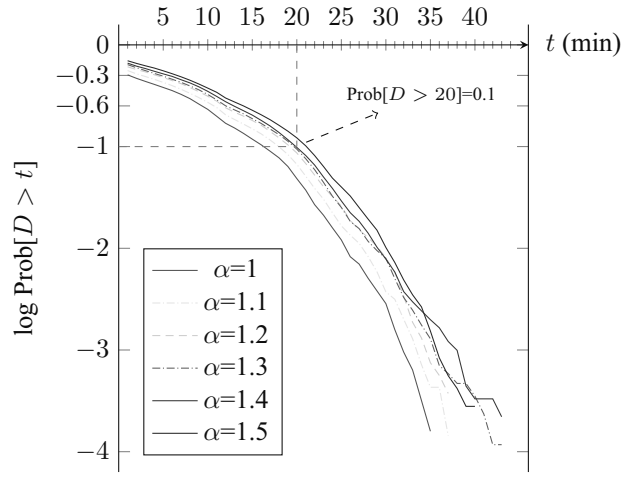


Figure 10: Distribution of the delays for upstream traffic in function of the arrival intensity .

CONCLUSIONS

A stochastic simulation model was created to assess the overall impact of implementing a one-way policy in the Albert Canal. Due to construction works, only part of the canal will be available for vessel traffic. The inputs of the simulation model were found by performing statistical analysis on real Automatic Identification System (AIS) data. The main performance measures include the vessels' waiting times and the queue lengths. Several scenarios were investigated with varying arrival intensities.

REFERENCES

2018. *The Oosterweel link*. Retrieved from: `https://www.oosterweelverbinding.be/oosterweel-link`.

Almaz A.; Or I.; and Ozbas B., 2006. *Investigation of the transit maritime traffic in the Strait of Istanbul through*

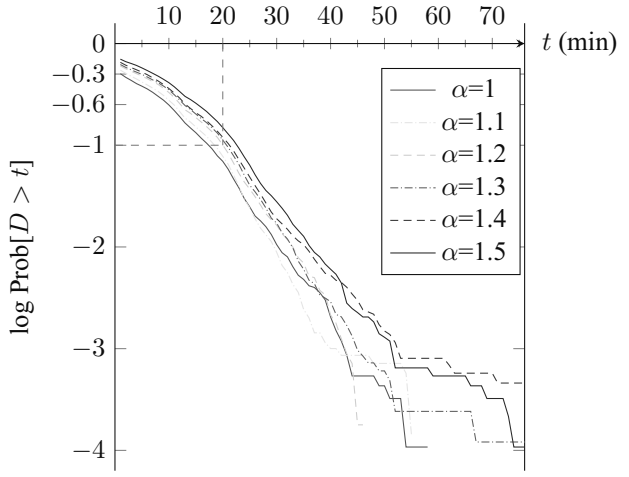


Figure 11: Distribution of the delays for downstream traffic in function of the arrival intensity.

simulation modeling and scenario analysis. *International Journal of Simulation*, 7, no. 7, 1–9.

Golkar J.; Shekhar A.; and Buddhavarapu S., 1998. *Panama canal simulation model*. In *Simulation Conference Proceedings, 1998. Winter*. IEEE, vol. 2, 1229–1237.

Köse E.; Başar E.; Demirci E.; Güneroğlu A.; and Erkebay Ş., 2003. *Simulation of marine traffic in Istanbul Strait*. *Simulation Modelling Practice and Theory*, 11, no. 7-8, 597–608.

Merrick J.R.; Van Dorp J.R.; Blackford J.P.; Shaw G.L.; Harrald J.; and Mazzuchi T.A., 2003. *A traffic density analysis of proposed ferry service expansion in San Francisco Bay using a maritime simulation model*. *Reliability Engineering & System Safety*, 81, no. 2, 119–132.

Özbaş B. and Or I., 2007. *Analysis and control of maritime transit traffic through the İstanbul Channel: a simulation approach*. *Central European Journal of Operations Research*, 15, no. 3, 235–252.

Thiers G.F. and Janssens G.K., 1998. *A port simulation model as a permanent decision instrument*. *Simulation*, 71, no. 2, 117–125.

A SIMULATION DRIVEN BRANCH AND BOUND OPTIMIZER FOR PLANNING CHARGING INFRASTRUCTURES

Hubert Büchter
Fraunhofer-Institute for Material Flow and Logistics IML
Joseph-von-Fraunhofer-Str. 2-4
44227 Dortmund
Germany
E-mail: hubert.buechter@iml.fraunhofer.de

Elisabeth Pöter
Fraunhofer-Institute for Material Flow and Logistics IML
Joseph-von-Fraunhofer-Str. 2-4
44227 Dortmund
Germany
E-mail: elisabeth.poeter@iml.fraunhofer.de

KEYWORDS

Optimization, Simulation, Transportation, Urban affairs, Object-oriented

ABSTRACT

Electric buses in public transport are increasingly being put into operation. Although batteries with large capacities are available, at least one charging point must exist. Especially buses with smaller batteries need a well-designed charging infrastructure. Planning an optimized charging infrastructure is an elaborate and time-consuming optimization process. In the end, no bus is allowed to give up its service due to the lack of energy, which is checked by simulation. A successful simulation for the maximal layout of a charging network is a prerequisite for the existence of at least one solution. Under this condition, an optimizer minimizes the number of charging stations with respect to a successful simulation and a minimal value of a given objective function.
The developed optimizer, based on a Branch and Bound algorithm, solves this integer problem with a specific heuristic. This optimizer will be used as a central component of a planning tool for bus companies, energy providers and urban planners.

INTRODUCTION

The operation of electric vehicles in public transport reduces environmental impacts. Running a fleet of buses requires a well-designed charging infrastructure, which has to take the requirements for the electricity network into account. Big batteries with an overnight charging or medium sized batteries with a rapid charging concept produce unwanted power peaks. An infrastructure, which supports the power grid, requires an appropriate amount of charging stations with the aim to spread the load in space and in time.

Many publications about the planning problem for charging infrastructures deal with non-public urban transport, e.g. (Frade et al. 2011; Kuchshaus et al. 2012; Chen et al. 2013). An overview of optimal planning for charging stations is in (Zheng et al. 2014). That work deals with stochastic models with the focus on the grid and does not meet the specific characteristic of public transportation, which is timetable driven. The complexity of the planning process for charging infrastructure is analyzed in (Lam et al. 2014) and it shows that it is NP complete. Due to this result, any planning algorithm of practical relevance is an approximation. (Olsen and Kliewer 2016) extend a vehicle-scheduling problem by charging models with different battery characteristics. They find that in a very often used operation range a linear approximation of the battery characteristic is sufficient. (Buechter and Naumann 2016a) and (Buechter and Naumann 2016b) used this approach in an optimizer that is based on linear programming and simulation. This optimizer solves the problem for linear behavior of all components but with a huge demand of memory. It is limited to about 100 buses and 1000 bus stops, which is too less for many applications.
(Espinouse et al. 2000) introduce a simulator driven optimizer based on branch and bound, which is very specific for job sequencing and vehicle dispatching. In our problem, there is no need for sequencing but many potential solutions are not feasible.
The central focus of this paper is on public transport with electrically driven buses. The charging infrastructure must comprise bus stops, which are equipped with charging devices. The task is to select those bus stops, which must become charging stations.
This work is part of the project 'MENDEL' (Trumpold et al. 2017) that is funded by the German government.

Figure 1 shows an example with six bus stops and two buses. Bus stop 1 is already a charging station and bus stop 5 is excluded from upgraded for charging by the planner. The other bus stops {2,3,4,6} are candidates for additional charging stations.

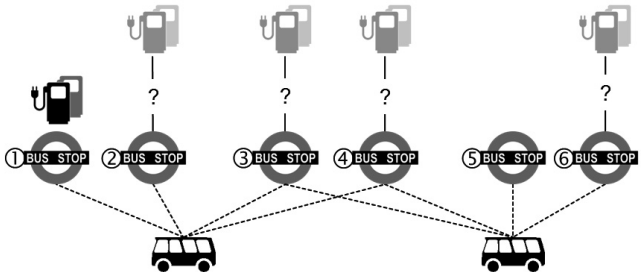


Figure 1: A Tiny Planning Problem

It is obvious that at least one stop from {3,4,6} must belong to the solution set in order to charge the second bus. The planning algorithm calculates the best choice of bus stops, which guarantee a continuous operation of all buses. Figure

2 shows a possible result for this example. Only one additional charging station at bus stop 4 is required.
The currently implemented planning process minimizes the costs for investment. An extension for other e.g. technical aspects like minimal time for overlapping charging of different buses is possible.

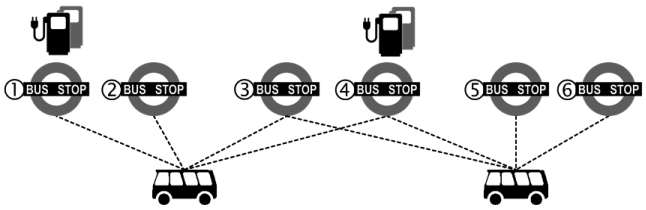


Figure 2: Solution for the Tiny Planning Problem

The optimizer is based on a Branch and Bound algorithm with a heuristic also useable for solving analogous problems.

This paper is structured as follows: The first section introduces the problem and describes the objects and their interaction and properties. Based on this, the following sections describe the simulation and optimization models. Finally, we present some results and give an outlook for the further work.

PROBLEM DESCRIPTION

The *optimization model* as well as the *simulation model* requires a *system model*. The relevant physical objects, which the system model takes into consideration, are *buses*, *ways*, *bus stops* and *charging stations* (1).

$$\left.\begin{array}{rlll} \text{Buses} & i \in \mathbb{B}, & \mathbb{B} & = \{1, \dots, nbBuses\} \\ \text{Ways} & j \in \mathbb{W}, & \mathbb{W} & = \{1, \dots, nbWays\} \\ \text{Bus Stops} & k \in \mathbb{S}, & \mathbb{S} & = \{1, \dots, nbStops\} \\ \text{Charging Stations} & & \mathbb{C} & \subseteq \mathbb{S} \end{array}\right\} \quad (1)$$

The transportation network consists of bus stops $\mathbb{S}$ and ways $\mathbb{W}$. A way is a sequence of roads connecting two bus stops. The corresponding mathematical model is a directed graph $N = (\mathbb{S}; \mathbb{W})$. The nodes $\mathbb{S}$ represent the bus stops and the edges $\mathbb{W}$ represent the ways each connecting two bus stops. Each way is attributed with length l and average values for drive time t_D and energy consumption e^-.

Buses can charge at those bus stops on their routes, which are equipped with a charging device. There are some limits for available power, power transfer and battery energy (Table 1).

Table 1: Power and Energy Limits for Buses and Charging Stations

Bus		Charging Station	
$\boldsymbol{pb_{max}}$	Max. charging power	$\boldsymbol{ps_{max}}$	Max. power offer
$\boldsymbol{p_{base}}$	Base power consumption		
$\boldsymbol{e_{min}}$	Min. battery energy		
$\boldsymbol{e_{max}}$	Max. battery energy		

Batteries have limited charging power. This limit can be individual for each bus ($\boldsymbol{pb_{max}}$). The charging devices have a maximal output power. This limit can be individual for each charging station ($\boldsymbol{ps_{max}}$).
If a bus stop is equipped with a charging device, it becomes a *charging station* $\mathbb{C} \subseteq \mathbb{S}$. The estimation of $\mathbb{C}$ is part of the optimization model, which is introduced in the following section. The system model also describes the charging behavior of the batteries depending on specific characteristics.

According to their schedule buses drive on ways from one bus stop to the next. Details of the ways, such as junctions and traffic lights, are not taken into consideration. All buses operate cyclic over time. At the end of a cycle all buses are at their start position again. The cycle time is one day typically. At the beginning of a cycle, the batteries have an *initial charge* $\boldsymbol{e_{ini}}$ and to the end, they must still have a *residual charge* of $\boldsymbol{e_{trm}}$ that is often smaller than the initial charge. This energy difference $\Delta\boldsymbol{e} = \boldsymbol{e_{trm}} - \boldsymbol{e_{ini}}$ can fed at a depot during the operating pause, but this is not considered here. For a complete cycle, the buses need an amount of energy that consists of the traction energy and the base energy. The *traction energy consumption* for each way and for each bus is given in a matrix $\boldsymbol{Q} \in \mathbb{R}_0^{+|\mathbb{B}|\times|\mathbb{W}|}$. Additional to the traction energy there is a *base load* $\boldsymbol{pb_{base}} \in \mathbb{R}_0^{+|\mathbb{B}|}$ for e.g. air condition, lightning and heating as a power request for each bus. Figure 3 shows a typical graph for the battery charge over time.

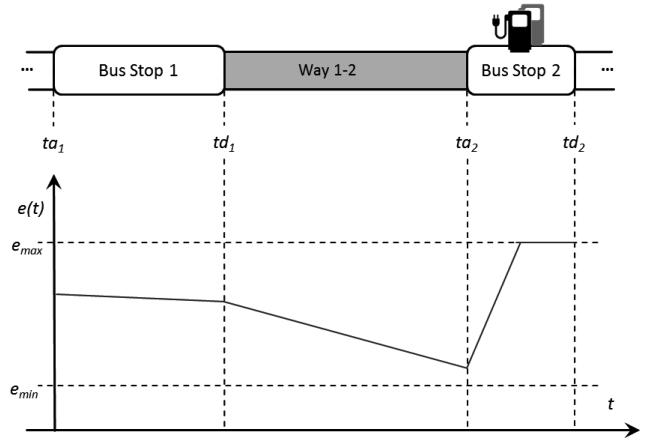


Figure 3: Energy for One Bus Arriving at Bus Stop 1 at ta_1 Starting at t_{d_1} and Driving over Way 1-2 to Charging Station at Bus Stop 2

The buses' charging controllers may make full use of the energy offer or even in part. Less charging power is preferable to reduce power peaks on the grid and to save battery life.

SIMULATION MODEL

The simulation model is deterministic and works on a set of charging stations $\mathbb{C}$ and the maximum power $\boldsymbol{ps_{max}}$, which they can deliver. At the beginning of simulation, all buses are placed at their initial bus stop, the battery energy is set to the initial value e_{ini} and the simulation clock is set to the start time t_0.

The simulator runs event driven. There are two types of events: Departure from a bus stop and arrival at a bus stop. When an event occurs, the charging states of the batteries are updated and some data is collected for later evaluation. The energy of the batteries must always be in a range between e_{min} and e_{max}.

There are two different simulation modes: *Full simulation* and *embedded simulation*. Full simulation collects many data while running and it does not stop until all trips will have terminated. If the battery's charge of one or more buses falls below e_{min} and even if one or more batteries are temporarily in a negative charging state, which is physically impossible, the simulation process continues. The purpose is to find bottlenecks manually e.g. by analyzing some plots of battery charge versus time or battery charge versus number of charging stations visually.

The embedded simulation acts strictly. If at least one bus's energy falls below e_{min} then the simulation stops and returns a degree of progress. This degree in a range $[0,1]$ is a measure of the feasibility of the set of charging stations. The optimizer uses this mode internally thus reducing the computation time significantly.

When a bus arrives at a bus stop the energy consumption for the drive from the previous bus stop is to calculate and to subtract from the current bus's energy e. For a more sophisticated model it is not only a subtraction but it also follows a specific nonlinear battery characteristic. The precision of the model is up to the requirements. The energy demand depends on the type of the bus, the driven way and the drive time for this way. The consumed base energy is given by $p_{base} \cdot t_{drive}$.

When a bus leaves a bus stop, the battery energy is reduced by the base energy which is $p_{base} \cdot t_{stop}$. If the bus stop is a charging station then the calculation of the power flow is a bit more complicated. Arriving at those bus stops the charging process starts if the bus's battery energy has fallen below a given value and if the stop time is greater than the set-up time for connection plus the set-down time for disconnection. During the stay at a charging station the charging device delivers the base power and the charging power. The transfer power $p_{transfer}$ is the minimum of the power offer p_{stop} from the charger and the maximum allowed charging power p_{max} of the battery and the base power p_{base} (2).

$$\left.\begin{aligned} p_{transfer} &= p_{charge} + p_{base} \\ p_{transfer} &= \min(p_{stop},\, p_{max} + p_{base}) \\ p_{charge} &= \min(p_{stop} - p_{base},\, p_{max}) \end{aligned}\right\} \quad (2)$$

The power flow p_{charge} for charging can be negative if the charger does not deliver enough power to satisfy the base power demand.

SIMULATION BASED OPTIMIZER

The aim of the optimizer is to estimate from a given set Y of bus stops a subset X of charging stations which minimizes the value of the *objective function* $c(X)$ under the constraint $f(X) = 1$. See (3) for the problem definition.

Let the function $f(X)$ express a value for the *feasibility* of the set X. The solution is feasible if $f(X) = 1$. Otherwise, $f(X)$ is a measure for the benefit of X. The better the feasibility the more suitable is the vector X.

The condition for the existence of a solution is the feasibility of the origin set (4)

$$X \subseteq Y \quad \begin{cases} c(X) \to \min & c \in [0,\infty] \\ f(X) = 1 & f \in [0,1] \end{cases} \quad (3)$$

$$f(Y) = 1 \quad (4)$$

$$U \subseteq V \subseteq Y \quad \begin{cases} c(U) \le c(V) \\ f(U) \le f(V) \end{cases} \quad (5)$$

The objective function reflects very often the economic costs. However, other aspects like urban planning factors, risks or technical efforts could be considered too. Both, the objective function $c(X)$ and the simulator function $f(X)$ are monotonic (5) which is given in the context of the planning problem. If more charging devices are installed then the costs increase but also the feasibility of the system increases.

There are some specific requirements for the infrastructure planning process. Some charging stations may already exist and some bus stops cannot be equipped with charging devices. A second requirement is a short run time for the optimizing process. The problem is NP-hard and without any heuristics, the run time of any implementation would be unacceptable. The following paragraphs present a solution for these problem specific requirements.

A simple solution is to set the costs for existing charging stations E to zero and those for the excluded bus stops F to infinity. However, in order to reduce the solution space it is better to give the set E and the set F in advance. These sets are defined in (6) and (7) shows the influence to the resulting set X.

$$F \subseteq Y, \quad E \subseteq Y, \quad F \cap E = \emptyset \quad (6)$$

$$X \subseteq (Y \backslash F) \cup E \quad (7)$$

The reduction of the problem size has a significant influence on the run time. It is applicable to many real bus systems because many bus stops are not suitable to become a charging station.

The optimizer is based on the Branch and Bound algorithm (Algorithm 1). The lower bound is given by an objective function for invest for feasible sets of charging stations, which are selected. Real world problems cannot be optimized in an acceptable run time without any heuristic.

Algorithm 1: Branch and Bound principle

Input: candidates Y ; startCandidates $\subseteq Y$
Output: $X \subseteq Y$; minimal costs $minCosts$

```
1:  minCosts ← ∞;
2:  X ← ∅
3:  initialize queue q
4:  q.put(startCandidates)
5:  while not q.empty()
6:      candidate ← q.get()
7:      currentCosts ← c(candidates)
8:      if currentCosts < minCosts
9:          minCosts ← currentCosts
10:         X ← candidate
11:     elseif
12:         q.put(nextCandidates(candidates))
13:     endif
14: endwhile
15: return (X, minCosts)
```

It depends on the access strategy of the queue how the solution set develops. Three basic branching strategies are very common:

- *First In First Out*: The search tree is developed in the breadth first (BFS). The implementation of the queue has the semantic of a sequential queue (FIFO).
- *Last In First Out*: The search tree is developed in the depth first (DFS). The implementation of the queue has the semantic of a stack (LIFO). This strategy is often implemented by recursion, which uses implicitly the call stack of the runtime environment.
- *First In Best Out*: The search tree follows always the best candidates first (BestFS). The implementation of the queue has the semantic of a priority queue.

The optimizer for the planning process uses a combination of all three basic strategies. It starts with a breadth search and puts the results in a priority queue with a limited capacity k. This reduces the search space significantly but with the risk for losing the solution for the global optimum. The priority is calculated as the relation between the feasibility and the costs of the current selected subset. This is a measurement for the use costs relation, which finally represents the selection strategy.
The next step is a recursive invocation of the optimizer for all entries stored in the priority queue, which corresponds to a depth search (Algorithm 2). If a solution exists, this approach finds always a first result with an upper bound of run time. The maximal number N of nodes to develop for a first solution for a cardinality n of the set of bus stops is

$$N = 1 + n \cdot \frac{(n+1)}{2} \tag{8}$$

The complexity for finding a first solution is $\mathcal{O}(n^2)$. Longer run time can improve the result.

Algorithm 2: Optimizer with the specific heuristic

```
Input:  function simulate (S)
        function costs (S)
Output: Optimal set of charging stations X
X ← { }
minCosts ← ∞
function opt(E)
    priorityQueue pq
    // breadth-first search
    for all i ∉ E
        E ← E ∪ {i}
        c ← costs(E)
        if c < minCosts
            f ← simulate(E)
            if f = 1        // the solution is feasible
                X ← E
                minCosts ← c
            else
                pq.put( f/c , E)
            endif
        endif
        E = E \ {i}
    endfor
    // depth-first search
    while not pq.isempty
        opt(pq.get())
    endwhile
endfunction
```

SOFTWARE ARCHITECTURE

The implementation follows the Object Oriented paradigm and comprises three packages:

(1) The problem independent package *optimize*: This package contains interfaces for the simulator and for the objective function and an implementation of the branch and bound optimizer. Up to here there is neither problem depending code nor data types.
(2) The project specific package *mendel*: This package implements the problem depending simulator and objective function. A specific constructor extends the optimizer.
(3) The package for the application layer *isp* for the infrastructure planning: This package includes a graphical user interface and a server for remote access to the planning process as a service.

Figure 4 shows a class diagram for inheritance and associations between classes.

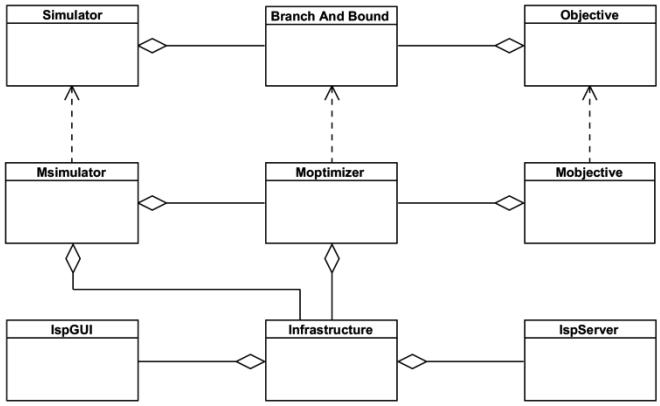


Figure 4: Class Diagram for the Planning System

The Branch and Bound optimizer runs in an own thread which starts a scalable number of additional threads to delegate subtasks when a new branch is generated.
This structure is important for applying the theoretical idea to other optimizing problems within the same problem class. Some callbacks allow following the optimizing process, which is very useful for first tests of new applications.

RESULTS

The introduced algorithms are implemented in pure Java. For first tests, artificial data form a model generator were sufficient. Next results came from investigations with bus operator companies in Germany and in Poland. Current work is on a medium sized model for the city of Braunschweig, Germany. This model consists of 1153 bus stops, 1758 ways and 230 buses.
All optimization were done on a Windows 7 System running on an Intel(R) Xenon(R) Hardware at 2.4 GHz and 8 Mbyte of Memory.
Figure 5 shows the number of analyzed nodes per second and the number of simulator calls per second. The plot of the graph begins, when the first solution was found. At this point, a lower bound is available. This lower bound reduces the search space and the number of simulator calls. The number of processed nodes per second increases.

In the beginning and after finding a good solution, the run time of the simulator is very often short because of an abort due to the lack of energy.

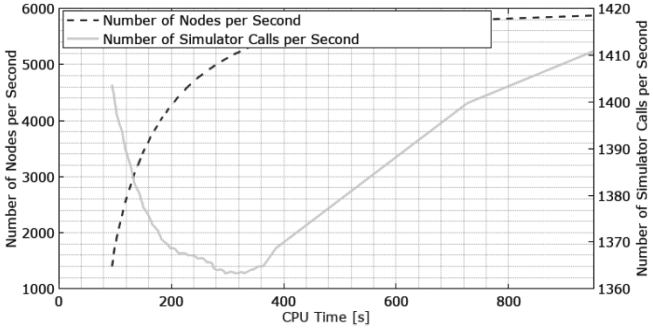


Figure 5: Number of Analyzed Nodes per Second and Number of Simulation Runs per Second versus CPU Time.

Figure 6 shows the number of charging stations and the costs versus CPU time. Both values have the same progress because the cost for building charging stations at different locations do not differ very much in this example. After approximately 360 s, a solution is available which does not improve significantly when the optimization continues.

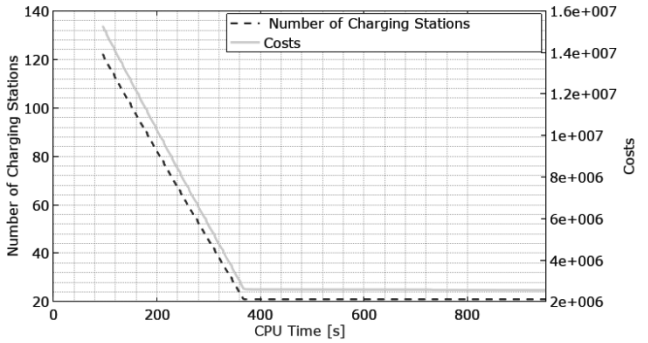


Figure 6: Number of Charging Stations and Costs versus CPU Time

This approach solves the planning problem for charging infrastructures. It runs quick and delivers results for practical use. Nevertheless, some improvements could be made to the optimizer. A dynamic length of the priority queue could keep the solution space small and investigate more candidates of interest. All nodes with feasibility to costs relation (see Algorithm 2, line 15) which are better or close to the already queued best candidates will be put into the queue. An adaptive algorithm should control this dynamic behavior. Another task is parallelizing the optimizer in order to reduce the computing time and to use more than one processor kernel.
The limits of earlier solutions (Buechter and Naumann 2016a) and (Buechter and Naumann 2016b) are expanded. Now, applications with of more than 1000 buses and more than 10000 bus stops can be solved.

CONCLUSIONS

The problem of planning a charging infrastructure for electrical driven buses is solved by a simulation driven branch and bound algorithm. The simulation controls the selection process and the lower bound is given by an objective function.

For further development, charging on the road with energy supply by pantographs or inductive energy transfer could be integrated in the simulator. As an advantage of the chosen software structure, this does not affect the optimizer. The optimizer only works on sets without any problem specific semantics. The simulator then works with a set of ways instead a set of bus stops.

REFERENCES

Buechter, H. and Naumann, S.. 2016a. “Optimizing Charging Infrastructure for Electrically Driven Buses.” In: European Simulation and Modelling Conference ESM 2016, pages 384–386.

Buechter, H. and Naumann, S. 2016b. “A Hybrid Planning Method for Charging Infrastructure for Electrically Driven Buses in Public Transportation.” In: Grzegorz Sierpiński (Editor) Intelligent transport systems and travel behavior. 13th Scientific and Technical Conference "Transport Systems. Theory and Practice 2016" Katowice, Poland, September 19-21, 2016 : selected papers., Cham: Springer (Advances in intelligent systems and computing, Volume 505), pages 175-185.

Chen, T. D., Kockelman, K. M., Moby Khan, et al. 2013. “The electric vehicle charging station location problem: A parking-based assignment method for Seattle.” In: Transportation Research Board 92nd Annual Meeting, volume 340, 2013, pages 13-1254

Enriquez, M. 2011, "A modular software framework supporting simulation-driven optimization techniques", In: Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA 30th, 2011, pages 7E3-1-7E3-10.

Espinouse, M. L.; Lacomme, P.; Moukrim, A.; Tchernev, N. 2000. "Branch and bound approach based on simulation for the job, input sequencing and device dispatching in FMS" In: IFAK Management and Control of Production and Logistics, Conference Proceedings, Grenoble, pages. 729–734.

Frade, I.; Ribeiro, A.; Goncalves, G.; Antunes, A.P. 2011. “Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal.” In: Transportation Research Record, Journal of the Transportation Research Board, 2011, pages 91-98.

Kuchshaus, V.; Bioly, S.; Klumpp, M. 2012. “Deducting E-Mobility Loading Station Locations from City Parking Data.” In: 17th International Working Seminar on Production Economics, Conference Proceedings, Innsbruck 20.-24.02.2012, Bd. 2: Innsbruck, pages. 81–92.

Lam, A. Y. S.; Leung, Y.-W.; Chu, X. 2014. “Electric vehicle charging station placement: Formulation, complexity, and solutions.” In IEEE Transactions on Smart Grid (5), pages 2846-2856.

Olsen, N., Kliewer, N. 2016, “Electric Vehicle Scheduling - A study on charging modeling for electric vehicles.” In: Operations Research Proceedings, 2016, pages 82–102.

Trumpold, J.; Oertel, R.; Weißer, D.; Schaefer M.; Maier, P; Naumann, S.; Büchter, H., 2017. "MENDEL: Minimum load of electrical networks caused by charging operations of electric buses", In: EVS30 International Battery, Hybrid and Fuel Cell Electric Vihicle Symposium, 9.-11. October 2017, Stuttgart.

Zheng, Y.; Dong, Z. Y.; Meng, K.; Zhao J. H.; Qiu, J., 2014. “Electric Vehicle Battery Charging/Swap Stations in Distribution Systems: Comparison Study and Optimal Planning” In: IEEE Transactions on Power Systems, Vol. 29, No. 1, January 2014, pages 221-229.

INVENTORY TRACKING

STOCHASTIC OPTIMIZATION OF A LARGE-SCALE INVENTORY-ROUTING PROBLEM WITH TRANSSHIPMENT THROUGH INTRODUCTION OF EFFECTIVE SIMULATION STEPS

Wouter Lefever
Department Industrial Systems Engineering, and Product Design, Ghent University
Technologiepark 903, 9052 Zwijnaarde
Belgium
email: wouter.lefever@ugent.be,

El-Houssaine Aghezzaf
Department Industrial Systems Engineering, and Product Design, Ghent University
Technologiepark 903, 9052 Zwijnaarde
Belgium
email: elhoussaine.aghezzaf@ugent.be

Khaled Hadj-Hamou
DISP
INSA Lyon
21 avenue Jean Capelle, 69621 Villeurbanne
France
email: khaled.hadj-hamou@insa-lyon.fr

KEYWORDS
Inventory-Routing, Stochasticity, Large-scale Optimization, Simulation

ABSTRACT

In this paper, we consider the Inventory-Routing Problem with Transshipment (IRPT) under stochastic demand. Traditional methods that optimize the decisions of stochastic problems often approximate the probability distribution by a set of scenarios. Although for small instances this approach often results in good quality solutions, the computational requirements make it unsuited for the optimization of large-scale problems. We investigate how such a sample average approximation method can be adjusted so that large instances of the stochastic IRPT can be solved within reasonable time. For this purpose we intersperse the optimization steps with a simulation phase that eliminates uninteresting solutions. We also develop a sequential simulation procedure to effectively select the optimal solution in the final stage of the algorithm. The experimental results show that the adjusted sample average approximation algorithm is able to solve instances with up to 35 retailers within reasonable time.

INTRODUCTION

Currently, one of the most investigated supply chain strategies is Vendor-Manged Inventory (VMI). In a VMI system the supplier is in charge of both the distribution of the products and the inventory management at the retailers warehouse. By doing so he can combine deliveries to multiple retailers and optimize the overall distribution costs. The only restriction he needs to respect is to make sure no stock-outs occur at each retailer.
The mathematical model that optimizes the decisions of the supplier in a VMI system is called the Inventory-Routing Problem (IRP). One of the major assumptions in the theoretical work on the IRP is that the demand is fixed and perfectly predictable. In comprehensive survey on the industrial aspects of the IRP Andersson et al. (2010) point out that this assumption is almost never satisfied in a real-life context. They indicate the need for more flexible models that include stochasticity and uncertain elements.
In a recent paper Lefever et al. (2018) demonstrate that the Inventory-Routing Problem with Transshipment (IRPT) is a flexible variant of the IRP that has a lot of potential when stochasticity is included. By adding transshipments the costs of stochastic problems is drastically reduced. One of the limitations of this work is that the presented algorithm requires the solving of several stochastic problems. These problems are known to be very hard to solve. Consequently, the stochastic optimization of large problems becomes computationally intractable. In this paper, we investigate how the algorithm of Lefever et al. (2018) can be adapted for large-scale IRPTs.

LITERATURE REVIEW

Within the field of inventory-routing the issue of uncertain demand is often resolved by introducing flexibility or robustness within the distribution plan. Gaur and Fisher (2004) consider the IRP in a supermarket chain context. The demands vary hourly. By clustering stores they are able to optimize the routing and inventory decisions over a week-long period. Solyalı et al. (2012) investigate an IRP with stochastic demands in which backlogging is allowed. They propose two robust MIP formulations for the problem. In the first one the problem is reformulated using the Price Of Robustness framework Bertsimas and Sim (2004). In the second one they use modified demand values to create robust solutions. For

both models a branch-and-cut (B&C) algorithm is presented.

In this paper flexibility is introduced by allowing transshipments. The study of transshipments within inventory-routing is a relatively new area of research. To the best of our knowledge, Coelho et al. (2012) are the first to investigate transshipments in the IRP. In their study, they consider four variants of the problem: the IRP with transshipment (IRPT) and the IRP without transshipment (IRP), under an order-up-to or an maximum level replenishment policy. To solve these problems they propose a heuristic based on the adaptive large neighborhood search (ALNS) framework presented by Pisinger and Ropke (2007).

The first Mixed Integer Program (MIP) to solve the IRPT is presented by Coelho and Laporte (2013). The authors propose branch-and-cut algorithm which includes a solution improvement procedure that approximates the cost of a new solution resulting from vertex removals and/or insertions. The MIP of Coelho and Laporte (2013) is improved by Lefever et al. (2017). They develop two new sets of valid inequalities for the problem which greatly strengthen the linear relaxation of the problem. Also, the routing component is reformulated by exploiting the possible presence of direct shipments in the optimal solution.

The flexibility of the IRPT is demonstrated in Coelho et al. (2014). In this paper the Dynamic and Stochastic IRP (DSIRP) is investigated. The authors consider a rolling planning horizon in which the retailer demands are revealed gradually. Four different policies are investigated: reactive or proactive (using demand forecasts) and with or without transshipments. Results show that using demand forecasts and transshipments greatly improves the solutions quality.

Chrysochoou and Ziliaskopoulos (2015) study the IRPT under stochastic demand. They develop an L-Shaped Algorithm for this problem and propose a set of new valid inequalities that enhance the computational process. The experimental results are conducted for scenario sets consisting of 3 and 5 scenarios. As the size of the scenario sets is very limited, the resulting solutions are very scenario-dependent.

Lefever et al. (2018) extend this work by introducing the concept of a skeleton solution. In their stochastic optimization algorithm, they fix the skeleton based on a small scenario set and demonstrate that the remaining problem is a linear optimization problem. Hence, the other critical parameters can be optimized using a much larger scenario set. To ensure that the right skeleton is found, they replicate the procedure multiple times with different scenario sets.

In this paper, we will extend the work of Lefever et al. (2018) for large-scale IRPTs by introducing simulation steps.

PROBLEM DESCRIPTION

Coelho and Laporte (2013) present a model for the IRPT using the following binary variables: x_{ij}^p is equal to 1 iff retailer j immediately follows retailer i on the route of the supplier's vehicle in period p and y_i^p is equal to 1 iff either the depot $(i = 0)$ or retailer i is visited in period p.

$$\min \sum_{i\in\mathcal{V}}\sum_{p\in\mathcal{P}} h_i I_i^p + \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V},i<j}\sum_{p\in\mathcal{P}} c_{ij}x_{ij}^p + \sum_{i\in\mathcal{R}\cup\{0\}}\sum_{j\in\mathcal{V}'}\sum_{p\in\mathcal{P}} b_{ij}w_{ij}^p \quad (1)$$

$$\text{s.t. } I_0^p = I_0^{p-1} + r^p - \sum_{i\in\mathcal{V}'} q_i^p - \sum_{i\in\mathcal{V}'} w_{0i}^p \qquad p\in\mathcal{P} \quad (2)$$

$$I_i^p = I_i^{p-1} + q_i^p + \sum_{j\in\mathcal{R}\cup\{0\}} w_{ji}^p - \sum_{j\in\mathcal{V}'} w_{ij}^p - d_i^p \qquad p\in\mathcal{P},\ i\in\mathcal{V}' \quad (3)$$

$$I_i^p \le C_i \qquad p\in\mathcal{P},\ i\in\mathcal{V}' \quad (4)$$

$$q_i^p \le C_i - I_i^{p-1} \qquad p\in\mathcal{P},\ i\in\mathcal{V}' \quad (5)$$

$$q_i^p \le C_i y_i^p \qquad p\in\mathcal{P},\ i\in\mathcal{V}' \quad (6)$$

$$\sum_{i\in\mathcal{V}'} q_i^p \le Q y_0^p \qquad p\in\mathcal{P} \quad (7)$$

$$\sum_{j\in\mathcal{V}',i<j} x_{ij}^p + \sum_{j\in\mathcal{V}',i>j} x_{ji}^p = 2y_i^p \qquad p\in\mathcal{P},\ i\in\mathcal{V} \quad (8)$$

$$\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S},i<j} x_{ij}^p \le \sum_{i\in\mathcal{S}\setminus\{m\}} y_i^p \qquad \mathcal{S}\subseteq\mathcal{V}',\ p\in\mathcal{P},\ m\in\mathcal{S} \quad (9)$$

$$x_{0j}^p \in \{0,1,2\} \qquad j\in\mathcal{V}',\ p\in\mathcal{P} \quad (10)$$

$$x_{ij}^p \in \{0,1\} \qquad i,j\in\mathcal{V}',\ p\in\mathcal{P} \quad (11)$$

$$y_i^p \in \{0,1\} \qquad i\in\mathcal{V},\ p\in\mathcal{P} \quad (12)$$

$$q_i^p \ge 0 \qquad i\in\mathcal{V}',\ p\in\mathcal{P} \quad (13)$$

$$I_i^p \ge 0 \qquad p\in\mathcal{P},\ i\in\mathcal{V} \quad (14)$$

$$w_{ij}^p \ge 0 \qquad i\in\mathcal{R}\cup\{0\}, j\in\mathcal{V}',\ p\in\mathcal{P} \quad (15)$$

In the IRPT, the total cost to be minimized is the sum of inventory holding costs at the supplier and at the retailers, the routing costs for the supplier's vehicle and the transshipment costs (1). Constraints (2) and (3) are inventory balance constraints for the supplier and the retailers. Constraints (4) limit the inventories of the retailer by their maximum capacity. Constraints (5)-(6) define the quantities delivered. If retailers i is not visited in period p, then constraints (6) enforce that the quantity delivered to it will be zero. Otherwise, if retailer i is visited in period p, then constraints (6) limit the quantity delivered to the retailer's inventory holding capacity, and this bound is tightened by constraints (5), making it impossible to deliver more than what would fill this capacity. Constraints (7) state that the vehicle capacity is not exceeded. Constraints (8-9) guarantee that a feasible route is determined to visit all retailers served in period p. Finally, constraints (10-15) enforce integrality and non-negativity conditions on the variables.

THE STOCHASTIC IRPT

In Lefever et al. (2018) a two-stage recourse program for the stochastic IRPT is defined. In this model the the timing and

size of the deliveries of the supplier are optimized based on the probability distribution of the uncertain demand instead of on a forecast. These are called the first-stage decisions (x, y, q). In the second phase the exact demand is revealed. Based on this information it is determined whether additional deliveries, performed by a subcontractor, are needed. Also, the inventory management decisions are optimized.
The stochastic optimization problem can be written as:

$$\min \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}, i<j} \sum_{p\in\mathcal{P}} c_{ij} x_{ij}^p + \sum_{n=1}^{N} \frac{1}{N} G(x, y, q, n) \tag{16}$$

subject to $(6) - (13)$

In this problem the uncertain demand is approximated by N scenarios. Therefore, the above problem is often referred to as the sample average approximation (SAA) problem. For one particular scenario n the second-stage problem can be written as:

$$G(x, y, q, n) = \min \sum_{i\in\mathcal{V}} \sum_{p\in\mathcal{P}} h_i I_i^p + \sum_{i\in\mathcal{R}\cup\{0\}} \sum_{j\in\mathcal{V}'} \sum_{p\in\mathcal{P}} b_{ij} w_{ij}^p \tag{17}$$

$$\text{s.t. } I_i^p = I_i^{p-1} + q_i^p + \sum_{j\in\mathcal{R}\cup\{0\}} w_{ji}^p - \sum_{j\in\mathcal{V}'} w_{ij}^p - d_i^t(n) \quad p \in \mathcal{P},\ i \in \mathcal{V}' \tag{18}$$

$$(2), (4), (5), (14), (15)$$

Lefever et al. (2018) present a procedure that solves the SAA-problem consisting of three steps. First the routing skeleton (x, y) is determined based on a small scenario set. Secondly, the optimal quantity values q are determined based on a a large scenario set. Finally the cost of the first-stage solution (x, y, q) is estimated by using an evaluation scenario set. To increase the probability that the best routing skeleton is found, this procedure is repeated M times.
Although for small instances this procedure is very effective, for large instances the procedure becomes computationally very demanding. In the next sections, we will discuss how the algorithm can be adapted so that the best stochastic solution can be found in reasonable time. First, we will discuss the framework of SAA-Algorithm for large instances. Then, we discuss two simulation procedures to accelerate the algorithm and to ensure that the best stochastic solution is found. Finally, we present the new SAA-Algorithm.

Framework of the SAA-Algorithm

In the SAA-algorithm of Lefever et al. (2018) the steps are executed consecutively. A disadvantage of this approach is that computation time may be spent to obtain results that are already known. If two sets lead to the same routing skeleton, then their quantity values and estimated cost will be the same. We propose to decompose the SAA-algorithm into three phases. First we generate M stochastic solutions based on a small scenario set. Then, we optimize the quantity values q of the $M' \leq M$ unique routing skeletons. Finally, we select the best of the M' stochastic solutions (x, y, q).
By rearranging the components of the SAA-Algorithm we avoid spending computation time on identical tasks. The gain in computation time will be determined by phases two and three. If the number of unique routing skeletons M' is much smaller than M, then the second and third phase will be more effective than the M replications of steps two and three in the SAA-Algorithm of Lefever et al. (2018).

Selection of the Best Routing Skeletons

The number of unique routing skeletons M' depends heavily two aspects of the stochastic problem: the size of the instance (the number of retailers and the planning horizon) and the number of scenarios in the scenario set. In table 1 an example is given in which 100 stochastic solutions are created for different instance sizes and different scenario set sizes. For every combination the number of unique routing skeletons is given.

Table 1: The Number of Unique Routing Skeletons.

		Scenario Set Size		
		1	5	10
#Retailers	15	17	7	4
	25	37	11	6
	35	46	16	8

Because the stochastic IRPT is very complex to solve to optimality for larger instances, we will use just one scenario in the initial scenario set N. In general this will result in more routing skeletons. This is also confirmed by the results of table 1. Some of these routing skeletons might perform well for one particular scenario, but are poor solutions in the stochastic context. Therefore, we propose to add a simulation step to eliminate these routing skeletons between the generation of the M stochastic solutions and the rest of the procedure.
In the simulation step we will estimate the cost interval of each routing skeleton: For a particular routing skeleton (x, y) the simulation of the first-stage solution (x, y, q) results in an upper bound of the cost of the optimal stochastic solution: $\hat{g}(x, y, q) \geq \hat{g}(x, y, q^*)$. On the other hand, the simulation of the routing skeleton (x, y) in which the q values are not fixed, but optimized for every sample, results in a lower bound on the cost of the optimal stochastic solution: $\hat{g}(x, y, q^*) \geq \hat{g}(x, y)$. For every routing skeleton we obtain thus an interval $[\hat{g}(x, y); \hat{g}(x, y, q)]$ in which the cost of the optimal stochastic solution will lie. An example of these intervals is given in figure 1.
Based on the intervals $[\hat{g}(x, y); \hat{g}(x, y, q)]$ we can elimi-

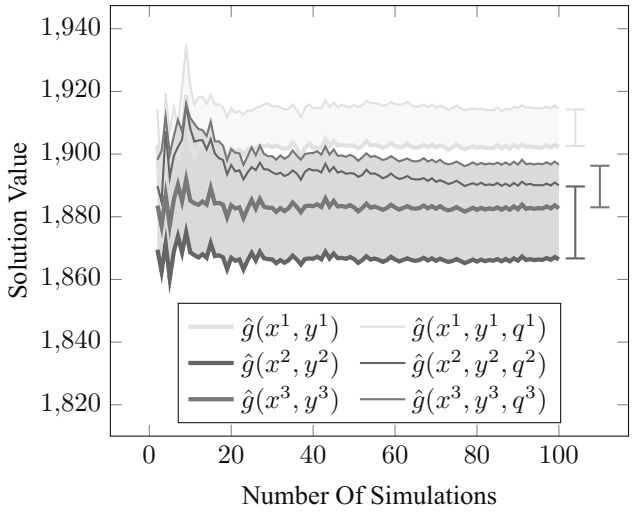


Figure 1: The Cost Intervals of Three Routing Skeletons

nate on the routing skeletons (x_j, y_j) of which the lower bound of the interval is higher than the lowest upper bound, $\hat{g}(x_j, y_j) > \min_i\{\hat{g}(x_i, y_i, q_i)\}$. After this elimination, we can continue the procedure with the remaining $M'' \leq M'$ routing skeletons.

For the example of figure 1, it is clear that after simulation over 100 samples the first routing skeleton can never produce the optimal first-stage solution. Thus, this routing skeleton will be eliminated for the rest of the procedure.

Selection of the Optimal Solution

For the M'' routing skeletons that cannot be eliminated, we optimize the quantity values q using a scenario set of N' scenarios. In the SAA-Algorithm of Lefever et al. (2018) the cost of the remaining M'' first-stage solutions is estimated using an evaluation scenario set of N'' scenarios. The main question in this approach is how large should N'' be so that it is likely that the best solution is actually chosen: A small scenario set N'' has a higher chance of selecting a sub-optimal solution, while a large scenario set will be computationally more demanding.

To face these two issues, we propose a solution procedure that dynamically eliminates solutions. In our solution procedure we will sequentially simulate the cost of every first-stage solution. For every solution the α-confidence interval can be calculated. We use the Student's t distribution as the critical value t^* for our confidence interval:

$$\left(\overline{g(x, y, q^*)} - t^* \frac{s}{\sqrt{n}}, \overline{g(x, y, q^*)} + t^* \frac{s}{\sqrt{n}}\right) \tag{19}$$

If the lower bound of one confidence interval exceeds the upper bound of another confidence interval during the simulation, we will eliminate this solution from further simulation. An example of our approach is shown in figure 2. The 95%-confidence intervals of the four solutions are indicated. At iteration 134, the confidence interval of the fourth solution does no longer overlap with the confidence interval of the first solution. Hence, this solution can be eliminated for the rest of the simulation. Similarly, the third solution and the second solution can be eliminated after 379 and 639 iterations.

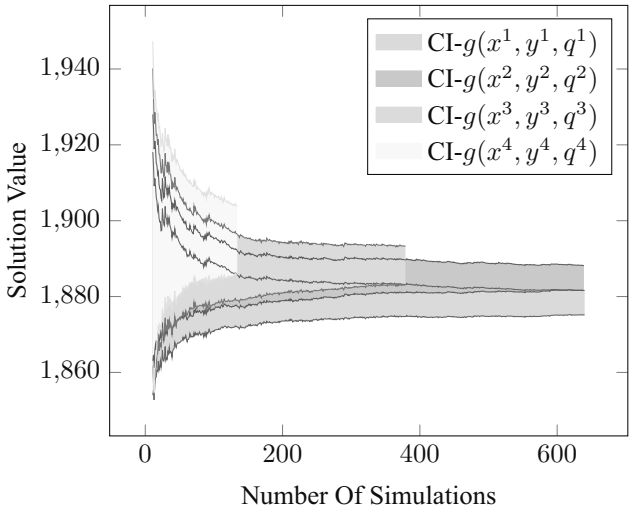


Figure 2: Selection Procedure of the Best Solution

By using the α-confidence intervals and eliminating uninteresting solutions early in the procedure, we are able to find the best solution with a high probability and limit the computational demands of the procedure.

Proposed SAA-Algorithm

We will now formally state the proposed algorithm.

Algorithm 1 SAA-Algorithm for solving the SIRPT

Step 0. For $m = 1, \ldots, M$: Solve the SAA-problem for a random generated scenario.

Step 1. Among the M' unique routing skeletons, select the $M'' \leq M'$ best skeletons based on their simulated cost interval $[\hat{g_k}(x, y); \hat{g_k}(x, y, q)]$.

Step 2. Optimize the quantity values q of the M'' best routing scenarios using scenario set N.

Step 3. Sequentially simulate the M'' first-stage solutions (x, y, q) and eliminate until one solution remains.

To reduce the variance in the simulation steps, Step 1 and Step 3, we use Common Random Numbers and Antithetic Variates.

EXPERIMENTAL RESULTS

To evaluate the proposed SAA-Algorithm we conducted experiments on a benchmark set of 35 instances (Archetti et al. (2007)). In Table 2 we compare the performance of the presented SAA-Algorithm with the solution procedure of Lefever et al. (2018). The cost of the obtained stochastic solution and the computation time are given.

Table 2: Comparison between the two SAA-Algorithms

Instance	Lefever et al. (2018)		SAA-Algorithm	
	Value	Time (s)	Value	Time (s)
abs05	770.79	18.07	770.81	3.92
abs10	1629.51	175.73	1629.52	145.18
abs15	1910.94	838.02	1910.92	251.79

The solution procedure of Lefever et al. (2018) and the presented SAA-Algorithm find the same optimal solutions. The small differences in solution values originate from the simulation of the cost of the optimal solution. In terms of computation time, the presented algorithm is much faster than the procedure of Lefever et al. (2018).
The effectiveness of the algorithm is further demonstrated in table 3. In the second and third column the number of unique routing skeletons '#URS' and the number of eliminated routing skeletons '#ERS' are reported. In the fourth, fifth and sixth column we respectively report the total solution time 'Time (s)', the simulation time of step 2 'ST_1 (s)' and the average optimization time of step 4 'AOT (s)'.

Table 3: Detailed Performance Results

Instance	#URS	#ERS	Time (s)	ST_1 (s)	AOT (s)
abs05	3.8	1.8	3.92	0.72	0.30
abs10	10.4	2.2	145.18	5.91	4.58
abs15	18.4	5.8	251.79	14.42	5.63
abs20	19.6	7.2	630.32	29.64	13.63
abs25	37.6	9.6	1845.80	93.96	18.88
abs30	39.4	9.4	2746.42	241.78	23.68
abs35	45.2	11.2	5353.61	245.39	34.32

We observe that the number of unique skeletons increases rapidly with the size of the instance. Therefore, it is of great importance to eliminate the uninteresting routing skeletons early in the procedure. By simulating the cost interval of the routing skeletons on a small sample we are able to eliminate 20% or more of the routing skeletons for every instance. This considerably accelerates the procedure since we do not need to optimize the quantity values of these solutions and they are not considered in the final selection procedure.

CONCLUSION

In this paper, we presented an effective solution for the optimization of a large-scale stochastic Inventory-Routing Problem with Transshipment. In our solution method we use two simulation steps to reduce the computational burden of a sample average approximation algorithm. The computational results demonstrate the effectiveness of our solution procedure for both small and large instances.
This work opens up to a number of meaningful extensions. One of the drawbacks of this paper is the generation of routing skeletons using exact optimization techniques. Heuristically addressing this problem would significantly accelerate the procedure. Furthermore, a comparison between different modeling approaches of the SIRPT would be interesting. In this paper the routing and quantity decisions are chosen as first-stage decision. However, quantity decisions may also be delayed as trucks are often loaded shortly before leaving the depot. Such a comparison might provide useful insights for logistics managers.

REFERENCES

Andersson H.; Hoff A.; Christiansen M.; Hasle G.; and Løkketangen A., 2010. *Industrial aspects and literature survey: Combined inventory management and routing*. *Computers & Operations Research*, 37, no. 9, 1515–1536.

Archetti C.; Bertazzi L.; Laporte G.; and Speranza M.G., 2007. *A branch-and-cut algorithm for a vendor-managed inventory-routing problem*. *Transportation Science*, 41, no. 3, 382–391.

Bertsimas D. and Sim M., 2004. *The price of robustness*. *Operations Research*, 52, no. 1, 35–53.

Chrysochoou E. and Ziliaskopoulos A., 2015. *Stochastic Inventory Routing Problem with Transshipment Recourse Action*. *SMMSO 2015*, 17.

Coelho L.C.; Cordeau J.F.; and Laporte G., 2012. *The inventory-routing problem with transshipment*. *Computers & Operations Research*, 39, no. 11, 2537–2548.

Coelho L.C.; Cordeau J.F.; and Laporte G., 2014. *Heuristics for dynamic and stochastic inventory-routing*. *Computers & Operations Research*, 52, 55–67.

Coelho L.C. and Laporte G., 2013. *The exact solution of several classes of inventory-routing problems*. *Computers & Operations Research*, 40, no. 2, 558–565.

Gaur V. and Fisher M.L., 2004. *A periodic inventory routing problem at a supermarket chain*. *Operations Research*, 52, no. 6, 813–822.

Lefever W.; Aghezzaf E.; and Hadj-Hamou K., 2017. *An Improved Formulation for the Inventory-Routing Problem with Transshipment*. *Submitted to Computers & Operations Research*.

Lefever W.; Aghezzaf E.; and Hadj-Hamou K., 2018. *An effective sample average approximation algorithm for the Inventory-Routing Problem with Transshipment with stochastic demand.* In *ILS2018*.

Pisinger D. and Ropke S., 2007. *A general heuristic for vehicle routing problems*. *Computers & operations research*, 34, no. 8, 2403–2435.

Solyalı O.; Cordeau J.F.; and Laporte G., 2012. *Robust inventory routing under demand uncertainty*. *Transportation Science*, 46, no. 3, 327–340.

THE USE OF SEMI-VARIANCE FOR SAFETY INVENTORY DETERMINATION IN CASE OF UNCERTAIN COMPOUND POISSON DISTRIBUTED DEMAND

Katrien Ramaekers
Gerrit K. Janssens

Research Group Logistics, Faculty of Business Economics
Universiteit Hasselt – campus Diepenbeek
Agoralaan 1, B-3590 Diepenbeek, Belgium

KEYWORDS
Inventory management, safety inventory, semi-variance

ABSTRACT

An inventory system containing uncertainty in demand during lead-time requires to determine a safety inventory for re-ordering. Many textbooks on inventory control propose to use a Normal distribution for describing the demand during lead-time. Based on the knowledge of the standard deviation and on the distribution assumption, the safety inventory is calculated, given a prescribed customer service level. In case the real distribution is different from the Normal distribution, errors in the obtained service level may occur and, by this, also in the incurred cost. In other disciplines other risk measures than the standard deviation have been used. One of these is the concept of semi-variance. This research investigates whether the use of semi-variance is a valid alternative for the standard deviation in the determination of safety stock. The investigation is tested on compound Poisson distributed lead-time demand.

INTRODUCTION

Inventory management is confronted with various types of risk: uncertainty in demand, uncertainty in delivery time, uncertainty in quality and quantity of the goods delivered, and deterioration of perishable goods, amongst others. A literature study by Tang and Musa (2011) show that most academic articles focus on risk in material flow. They also state that there is a lack of models analysing the risk associated with information flows. Many academic articles deal with traditional inventory control models studying these models under particular conditions or incorporating additional consideration into the established models (Williams and Tokar, 2008). Berman, Krass and Tajbaksh (2011) investigate inventory pooling in a multi-location network, more specifically the sensitivity of the cost reduction to the variability in demand and to the number of locations being pooled.

Companies wish to offer a high service level to their customers. This means, when an order arrives, they want to be able to deliver the goods from stock without delay.

To protect them from this risk, companies hold extra inventory, called safety stock. They have to make decisions on the level of safety stock. A trade-off exists because higher levels of safety stock increase the service level to the customers but also increase the cost of holding that stock.

Consider the case of uncertainty in demand during lead-time. Most textbooks, but also most research in inventory management, base the determination of level of safety stock – given a predefined service level – on the standard deviation of the demand distribution as a risk measure.

Inspiration for alternatives may be found in the world of actuarial mathematics. The mathematics of non-life insurance deal with risk of claims. The role of a safety stock against variability in demand in inventory management is taken over by a premium in the insurance business to compensate for a set of claims. Premium calculation principles make use of a measure of risk. However many more principles are used in actuarial mathematics compared to those used in inventory management. Next to the standard deviation principle, some examples of other principles are: the expected value principle, the maximal loss principle, the semi-variance principle, the zero utility principle, the Swiss premium calculation principle, the Orlicz principle and the Esscher principle (Goovaerts, de Vylder and Haezendonck (1984). In this research the focus lies on the semi-variance principle for reasons to be explained further on.

The demand during lead-time (DDLT) mostly is stochastic, so the companies take the option of holding safety stock. The company decides on the level of safety inventory based on the probability of an out-of-stock event or on the expected number of units short while being out-of-stock. For this decision, the probability distribution of the DDLT needs to be known. Most textbooks and software assume that the DDLT follows a Normal distribution. An estimate of the mean and variance allows for the determination of the safety inventory, given the risk that the company would like to accept.

But in reality the DDLT does not always have the characteristics of a normal distribution: the distribution is not always unimodal and it is not always symmetric. This fact might lead to wrong and costly decisions. This paper investigates whether the choice of the 'standard deviation principle' is part of the problem leading to wrong decisions and whether another principle, the semi-variance principle, is more robust or not towards deviations from the shape of the Normal distribution.

In this study a compound Poisson distribution is chosen as an experimental basis. Compound Poisson distribution have appeared many times in literature as will be explained in the literature review section. Further it will be explained that, if the Poisson process is compounded with a uniform distribution, a wide range of distributions showing different values of skewness and kurtosis measures can be obtained. In such a way the value of the semi-variance principle can be investigated for various shapes of distributions but within the same family of distributions.

LITERATURE REVIEW

Several studies have shown that the shape of the demand distribution during lead time has an important influence on this decision and might lead to either wrong decisions in terms of service levels or will lead to higher costs than expected. It is not the intention of this paper to give a review on studies which investigate the consequences of Normal distribution assumptions while they are not valid. However, a few papers will be highlighted as they are of direct relevance to this study.
Lau and Zaki (1982) note that mean and variance are not sufficient for safety stock calculation, but also skewness and kurtosis should be accounted for. Bartezzaghi et al. (1999) show a significant impact of the shape of the demand distribution on the service level, based on a large set of experiments. Furthermore, Käki et al. (2013) show the impact of the demand distribution shape on replenishment, based on experiments with qualitative shape characteristics (normal, positively skewed, negatively skewed, and bimodal).
Sometimes, it is even an unreasonable assumption that the demand obeys a known distribution. In such a case, for with some agricultural products, an inventory replenishment policy has been proposed on the mean of the distribution only (Chen et al., 2016). In Janssens and Ramaekers (2011) an approach has been developed to obtain the reorder point based on the knowledge of the range, mean and variance of the demand distribution only, which is the same information as required for the use of the normal distribution (as many times used in commercial software).
In this study it is assumed that the demand follows a compound Poisson distribution. The use of such a distribution is not new and rather well established. Important early papers on inventory management using the compound Poisson distribution as a distribution for demand during lead-time are Archibald and Silver (1978), Federgruen, Groenevelt and Tijms (1983), Richards (1975) and Thompstone and Silver (1975).

PROBLEM FORMULATION

The intention of this study is to investigate whether the semi-variance principle offers a valid alternative for the standard deviation principle for calculating the safety stock and under which conditions it is or it is not. In this section first the concept of semi-variance is explained and then more elaboration is done on the shape of a compound Poisson distribution. The former section is, for the greater part, based on Ramaekers, Merkuryeva and Janssens (2017) but repeated here for easy reference and notations.

The Concept of Semi-Variance

The demand during lead time (DDLT) is expressed as a random variable X which follows a probability distribution with density $f(x)$. The cumulative distribution of X is written as $F(x) = Prob\{ X <= x\}$. The expected value of the DDLT is written as:

$$\mu_X = \int_{-\infty}^{+\infty} x \, dF(x)$$

and its variance as:

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 dF(x)$$

The variance may be split up in two parts defined as:

$$\sigma_X^2 = \sigma_X^{2-} + \sigma_X^{2+}$$

with

$$\sigma_X^{2-} = \int_{-\infty}^{\mu_X} (x - \mu_X)^2 dF(x)$$

and

$$\sigma_X^{2+} = \int_{\mu_X}^{+\infty} (x - \mu_X)^2 dF(x)$$

The formula (5) is called the *(positive) semi-variance* which also might be rewritten as

$$\sigma_X^{2+} = E[max(0, X - \mu_X)^2]$$

Formula (6) mostly will be estimated by means of the sample semi-variance, defined as

$$s_X^{2+} = \sum_{i=1}^{n} \frac{\left(max(x_i - \bar{x}, 0)\right)^2}{n}$$

where

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

The mean is estimated in the classical sense as the average value of the demand during a number of time periods. Once the mean is estimated the semi-variance of the DDLT can be estimated.
The reorder point R can be expressed as

$$R = \mu_X + k \, \sigma_X^+$$

with $k > 0$ a safety factor. In this study it is assumed that the right-hand side of the distribution corresponds to the tail of a normal distribution. Therefore the value of k in formula (9) can be determined by $= k_{Normal} * \sqrt{2}$, where k_{Normal} is the safety factor as obtained from the normal distribution. In practice the reorder point R is calculated as:

$$R = \bar{x} + k_{Normal} * \sqrt{2} * s_X^+$$

The Shape of a Compound Poisson Distribution

The demand process is described as demands that arrive at random epochs. The epochs may follow a Poisson

process, a renewal process, or a counting process. When the demand size at any of those epochs is an independent identically distributed (i.i.d.) random variable, the arrival process is called 'compound'. In this study the demand arrival process is a Poisson process and the distribution of the demand size follows a uniform distribution on a finite interval $[a,b]$ with $a < b$.

Consider a fixed lead time over which demand arises for a product. Let $D_1, D_2, D_3, ..$ denote the sizes of the demand which are i.i.d. random variables with a common distribution D, specified by a density function f_D with finite mean μ_D and variance σ^2_D. Furthermore let N denote the number of demands occurring in a lead-time and the distribution of N is Poisson with intensity parameter λ. Then

$$X = \sum_{n=1}^{N} D_n$$

denotes the total demand occurring during a lead time. The distribution of X is compound Poisson with parameter λ and demand size density function f_D. The mean and variance of the compound Poisson distribution are:

$$\mu = \lambda \mu_D$$

and

$$\sigma^2 = \lambda(\sigma_D^2 + \mu_D^2)$$

For the purposes of this study, some central moments of the compound Poisson distribution are required. Let the density function of the uniform distribution with support on [a,b] be

$$f(x) = \frac{1}{(b-a)} for\ a \le x \le b, 0\ else$$

The second, third and fourth central moment of the Poisson distribution compounded with this uniform distribution are:

$$\mu_2' = \lambda \frac{b^3 - a^3}{3(b-a)}$$

$$\mu_3' = \lambda \frac{b^4 - a^4}{4(b-a)}$$

$$\mu_4' = \lambda \frac{b^5 - a^5}{5(b-a)} + 3\lambda^2 \left[\frac{b^3 - a^3}{3(b-a)}\right]^2$$

Given a mean and variance of a distribution, its normalized third and fourth moments can be represented as a single point on, what is called, a Pearson chart (Kendall and Stuart, 1977). The Pearson distribution family can satisfactorily represent a wide range of observed data including Normal, Beta and gamma distributions. The Pearson two-dimensional chart indicates a range of distributions in terms of an asymmetry characteristic β_1 and a kurtosis characteristic β_2.

The asymmetry and kurtosis characteristics required for the chart are defined as:

$$\beta_1 = \frac{(\mu_3')^2}{(\mu_3')^3}$$

$$\beta_2 = \frac{\mu_4'}{(\mu_2')^2}$$

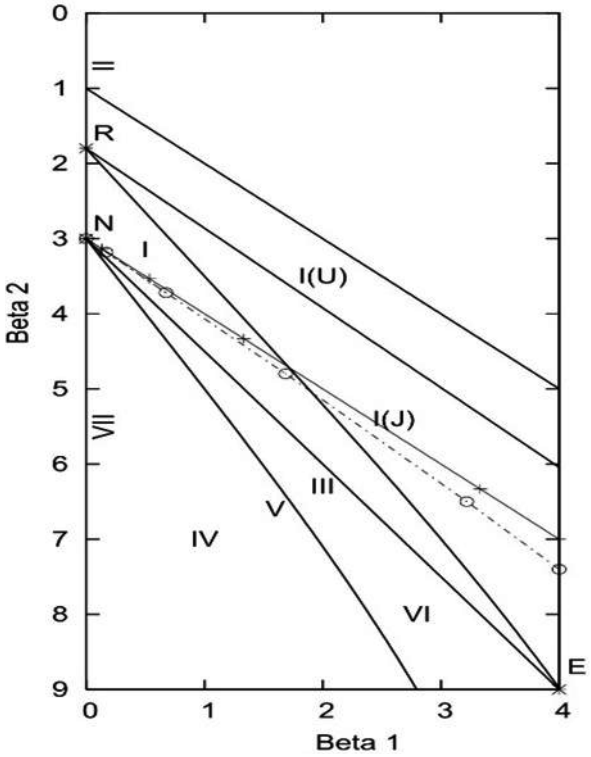


Figure 1: The Pearson chart, including information on the compound Poisson distribution

The Pearson chart is shown in Figure 1. The following types of distributions can be distinguished on the graph:

- Type I: Beta distribution. Three areas of type I can be distinguished based on the shape of the Beta-distribution: a unimodal Beta-distribution (I), a J-shaped Beta-distribution (I(J)) and a U-shaped Beta-distribution (I(U));
- Type II-line: special type of type I, restricted to symmetric distributions;
- Type III-line: Gamma or Chi-square distribution;
- Type IV: Cauchy distribution;
- Type V-line: reciprocal of Gamma or Chi-square distribution;
- Type VI: F-distribution;
- Type VII-line: Student's t-distribution;
- Normal distribution point N: (0, 3);
- Uniform distribution point R: (0, 1.8);
- Exponential distribution point E: (4, 9).

The chart in Figure 1 shows all lines, which form the boundaries of the various types of Pearson-distributions. The area filled with the compound distributions under study (i.e. compounded with the uniform distribution) is shown

between a line marked with the symbol + and a line marked with the symbol *o* .
For several combinations of mean and variance, the parameters β_1 and β_2 are calculated, which, for every combination, leads to a single point. It could be observed that, when the mean is high with respect to the variance, the values of β_1 (resp. β_2) are close to 0 (resp. 3), which means that the demand distribution is close to the Normal distribution. The smaller the mean and/or the greater the variance, the more β_1 and β_2 move away from the Normal values. The total demand moves towards the shape of a unimodal Beta-distribution. If the mean decreases further relative to the variance, β_1 and β_2 further increase and the shape of the distribution moves towards a J-shaped Beta-distribution.
In general, it can be stated that, when μ^2/σ^2 exceeds 15, the total demand distribution is close to the Normal distribution. When μ^2/σ^2 lies between 0.65 and 15, the unimodal Beta-distribution describes closest the total demand and, when μ^2/σ^2 is below 0.65, the total demand is close to a J-shaped Beta-distribution.

SOLUTION METHOD

Remind the research question: is the semi-variance principle a valid alternative for the standard deviation principle to define the re-order point in case it is expected that the Normal distribution assumption for the DDLT may lead to significant errors ?
The re-order points of both calculation methods shall be compared not only among themselves but also with the 'exact' re-order point obtained from the compound Poisson distribution. The latter distribution is not known in practice but can be controlled in the experiment. The quality of both methods might be different for various levels of customer service, therefore the experiment needs to include several customer levels. A comparison is made, for various service levels ranging from 90% till 99% of three ways of calculating the re-order point: (1) based on the standard deviation; (2) based on the semi-variance; (3) based on the complete knowledge of the compound distribution.
Assume mean and standard deviation of the DDLT are given or estimated. The first way requires consultation of the standard Normal distribution table. The second and third ways require simulation. For obtaining the value in the third way, a large sample of the compound distribution are generated from which the 90% till 99% are obtained, as an approximation for the real values. For the second, the value of the semi-variance (as defined in equation (5)) is approximated by

$$\sum_{i=index(\mu_D)}^{Ns} (x_i - \mu_D)^2 . \frac{1}{Ns}$$

where *Ns* is the size of the large sample and *index(μ_D)* is the ranked index (between 1 and *Ns*) of the mean value.
When the mean and variance of the aggregated demand are given as M and V, then the lower and upper bound of the uniform distribution can be expressed in their values as:

$$a = \frac{\frac{2M}{\lambda} - \sqrt{\frac{12V}{\lambda} - \frac{12M^2}{\lambda^2}}}{2}$$

$$b = \frac{\frac{2M}{\lambda} + \sqrt{\frac{12V}{\lambda} - \frac{12M^2}{\lambda^2}}}{2}$$

From the knowledge that a and b ≥ 0 and that $M^2 \leq E[X^2]$, a valid lower and upper bound for the value of λ can be determined as:

$$\frac{M^2}{V} \leq \lambda \leq \frac{4M^2}{3V}$$

NUMERICAL ILLUSTRATION

The theory is illustrated by a number of experiments. The demand during lead time has two characteristics: the mean value is kept constant at 100, while the variance ranges from 2000 till 32000. In such a way the variation coefficient changes, as well as the ranges $[a,b]$ of the uniform distribution and, with this, the measures of asymmetry and kurtosis. Table 1 shows the basic data for five experiments. Next to the mean and variance also the square of the variation coefficient (VC^2) and the lower and upper bounds for the order arrival intensity (λ^L and λ^U) are mentioned.

Exp.No.	*Mean (μ)*	*Variance (σ^2)*	VC^2	λ^L	λ^U
1	100	2000	5.0000	5.0000	6.6667
2	100	4000	2.5000	2.5000	3.3333
3	100	8000	1.2500	1.2500	1.6667
4	100	16000	0.6250	0.6250	0.8333
5	100	32000	0.3125	0.3125	0.4167

Table 1: Basic data for the experimental design

The middle value for the order arrival intensity is used in the experiments and mentioned as λ in Table 2. Given this value, the corresponding values of a and b are calculated as well as the values for the asymmetry (β_1) and kurtosis (β_2) measures.

Exp.No.	λ	*a*	*b*	β_1	β_2
1	5.8333	5.0210	29.2647	0.2429	3.2215
2	2.9167	10.0420	58.5294	0.4858	3.4430
3	1.4583	20.0841	117.0588	0.9716	3.8860
4	0.7292	40.1682	234.1175	1.9431	4.7718
5	0.3646	80.3364	468.2350	4.5341	7.1344

Table 2: Experimental values for the Compound Poisson distributions

To approximate the real Compound Poisson distribution a sample of size 1000 is generated using the following procedure:

Procedure Compound_Poisson;

```
Generate Number of Orders (Poisson distribution);
Cumulative_Size := 0;
While Number_of_Orders > 0
 do
        Generate Order_Size (Uniform distribution);
        Cumul_Size := Cumul_size + Order_Size;
        Number_of_Orders := Number_of_Orders – 1
 enddo;
```

An occurrence of the Demand During Lead Time is returned in the variable Cumul_Size.

Experiment number	*Fitted Distribution*
1	258 x Beta(2.56,4.07)
2	379 x Beta(1.52,4.33)
3	430 x Beta(0.72,2.49)
4	702 x Beta(0.25,1.68)
5	1150 x Beta(0.189,1.99)

Table 3: Best fit distributions for the Compound Poisson distributions

In order to show how the Compound distributions look like, the empirical distribution has been fitted to a number of standard distributions, as available in the Arena Input Analyzer. In each of the five experiments, the best fit is made by a Beta distribution. The estimated parameters for those Beta distributions are shown in Table 3.
Figures 2, 3 and 4 show the levels of safety stock in function of the service level for the experiments 1, 3 and 5. Each figure shows three curves depending on the way how the level of safety stock is calculated: the Normal approach using the variance (or standard deviation), the Normal approach using the semi-variance and the approximation of the real Compound Poisson distribution.

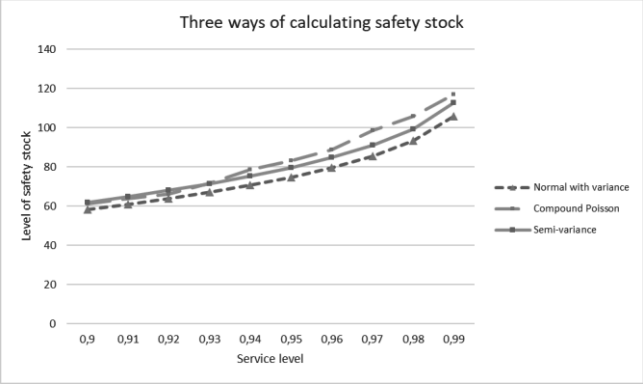


Figure 2: Level of safety stock in function of the service level (Experiment 1)

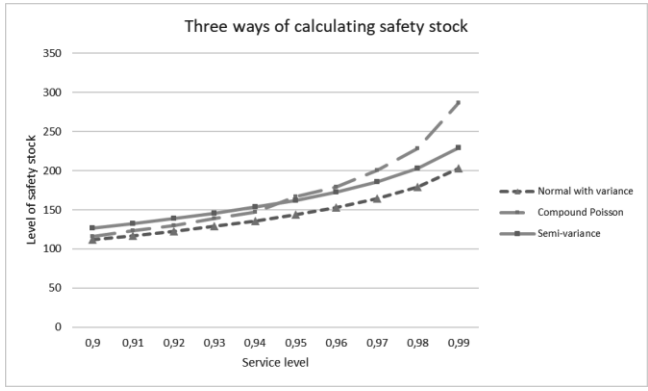


Figure 3: Level of safety stock in function of the service level (Experiment 3)

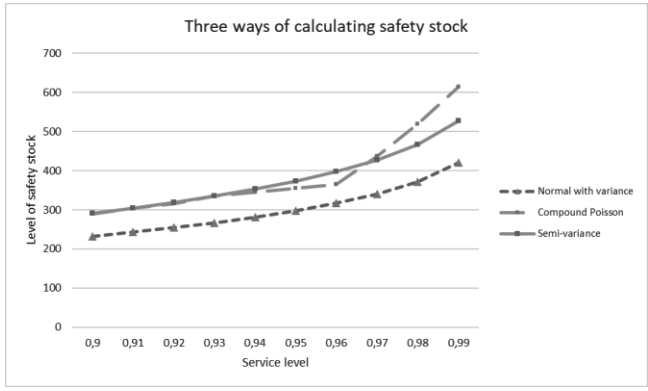


Figure 4: Level of safety stock in function of the service level (Experiment 5)

From the figures it can be learnt that the use of the semi-variance leads to a higher level of safety stock compared to the classical approach of using the standard deviation or variance. The difference between both levels increases with increasing value of the variance.
In general it can be stated that, for this type of distribution, the Normal approach with use of the semi-variance offers better results than the classical approach, when compared to the 'real' values (as obtained by the approximation). The classical approach; however, offers better results for the high variance experiments and the lower service levels (in the range under study).

CONCLUSIONS

The performance of the method of determining safety stock levels, based on the variance and on a safety factor based on the Normal distribution, depends on the shape of the real but many times unknown demand distribution during lead-time. This research paper has introduced the new concept of 'semi-variance' in this operational context. Preliminary results show that the new concept shows improvement over the classical approach but not under all conditions. While the compound Poisson distribution offers some flexibility in shape, the distributions are positively-skewed (or right skewed). So, intuitively it could be expected that the use of semi-variance should lead to higher levels of safety stock. Further investigation is required to investigate the cases

with a left-skewed distribution. The compound Poisson distribution, however, cannot serve for this.

ACKNOWLEDGEMENTS

The help of dr. Hanne Pollaris (UHasselt) in producing the figures is highly appreciated.

REFERENCES

Archibald, B. & Silver, E. (1978). (s,S) Policies under continuous review and discrete compound Poisson demands. *Management Science*, 24, 899-909.

Bartezzaghi, E., Verganti, R., & Zotteri, G. (1999). Measuring the impact of asymmetric demand distributions on inventories. *International Journal of Production Economics*, 60-61, 395-404.

Berman, O., Krass, D. & Tajbaksh, M.M. (2011). On the benefits of risk pooling in inventory management. *Production and Operations Management,* 20(1), 57-71.

Chen, W., Li, J., & Jin, X. (2016). The replenishment policy of agri-products with stochastic demand in integrated agricultural supply chains. *Expert Systems with Applications*, 48, 55-66.

Federgruen, A., Groenevelt, H. & Tijms, H.C. (1983). Coordinated replenishments in a multi-item inventory system with compound Poisson Demands, *Management Science*, 30(3), 344-357.

Goovaerts, M.J., de Vylder F. & Haezendonck, J. (1984). *Insurance Premiums: Theory and Applications,* North-Holland, Amsterdam.

Janssens, G.K, & Ramaekers, K. (2011). A linear programming formulation for an inventory management decision problem with a service constraint. *Expert Systems with Applications*, 38, 7929-7934.

Käki, A., Salo, A., & Talluri, S. (2013). Impact of the shape of demand distribution in decision models for operations management. *Computers in Industry*, 64, 765-775.

Kendall, M. & Stuart, A. (1977). *The Advanced Theory of Staistics – Volume I: Distribution Theory.* Charles Griffin, London.

Lau, H.-S., & Zaki, A. (1982). The sensitivity of inventory decisions to the shape of lead time-demand distribution. *IIE Transactions*, 14, 265-271.

Naddor, E. (1978). Sensitivity to distributions in inventory systems. *Management Science*, 24, 1769-1772.

Ramaekers K. & Janssens, G.K. (2007), On the choice of a demand distribution for inventory management models. *European Journal of Industrial Engineering*, 2(4), 479-491.

Ramaekers, K., Merkuryeva, G. & Janssens, G.K. (2017). The concept of semi-variance as a tool for safety inventory decisions in case of uncertain demand. *Proc. of the 2017 European Simulation and Modelling Conference (ESM'2017)*, Lisbon, Portugal, 201-205.

Richards, F.R. (1975). Comments on the distribution of inventory position in a continuous-review (s,S) inventory system, *Operations Research*, 23, 366-371.

Tang, O. & Musa, S.N. (2011). Identifying risk issues and research advancements in supply chain risk management. *International Journal of Production Economics,* 133, 25-34.

Thompstone, R. & Silver, E. (1975). A coordinated inventory control system for compound Poisson demand and zero lead time. *International Journal of Production Research*, 13, 581-602.

Williams, B.D. & Tokar, T. (2008). A review of inventory management research in major logistics journals. *International Journal of Logistics Management*, 19(2), 212-232.

AUTHOR LISTING

AUTHOR LISTING