Advances in Intelligent Systems and Computing 523

Jerzy Świątek
Zofia Wilimowska
Leszek Borzemski
Adam Grzech *Editors*

Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III

Springer

Advances in Intelligent Systems and Computing

Volume 523

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

More information about this series at http://www.springer.com/series/11156

Jerzy Świątek · Zofia Wilimowska
Leszek Borzemski · Adam Grzech
Editors

Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III

Springer

Editors
Jerzy Świątek
Department of Computer Science
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

Leszek Borzemski
Department of Computer Science
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

Zofia Wilimowska
Department of Management Systems
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

Adam Grzech
Department of Computer Science
Faculty of Computer Science
and Management
Wrocław University of Technology
Wrocław
Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-46588-3 ISBN 978-3-319-46589-0 (eBook)
DOI 10.1007/978-3-319-46589-0

Library of Congress Control Number: 2016951674

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This four-volume set of books contains the proceedings of the 37th International Conference on Information Systems Architecture and Technology, or ISAT 2016 for short, held during September 18–20, 2016 in Karpacz, Poland. The conference was organized by the Department of Management Systems and the Department of Computer Science, Wrocław University of Science and Technology, Poland.

The International Conference on Information Systems Architecture and Technology has been organized by the Wrocław University of Science and Technology since 1970s. The purpose of the ISAT conference is to discuss a state of the art of the information systems concepts and applications as well as the architectures and technologies supporting modern information systems. Contemporary organizations seem to be knowledge-based organizations and in connection with that information becomes the most important resource. Knowledge management is the process through which organizations generate value from their intellectual and knowledge-based assets. It is a management philosophy, which combines good practice in purposeful information management with a culture of organizational learning, in order to improve business performance. The computers are able to collect and select the information can make some statistics, and so on, but decisions have to make managers basing on their experience and taking into consideration computer support. An improvement of decision-making process is possible to be assured by analytical process supporting. Applying some analytical techniques, such as computer simulation, expert systems, genetic algorithms, can improve quality of managerial information.

One of the conference's aims is also to consider an impact of the knowledge, information, computing, and the communication managing technologies of the organization functionality scope as well as the enterprise information systems design, implementation and maintenance processes taking into the account various methodological, technological, and technical aspects. It is also devoted to the information systems concepts and applications supporting exchange of goods and services by using different business models and exploiting opportunities offered by Internet-based electronic business and commerce solutions.

ISAT is a forum for specialized disciplinary research, as well as on interdisciplinary studies that aims to present original contributions and to discuss different subjects of today's information systems planning, designing, development, and implementation. The event is addressed to the scientific community, people involved in variety of topics related to information, management, computer and communication systems, and to people involved in the development of business information systems and business computer applications.

This year, we received more than 110 papers form about 10 countries. Each paper was reviewed by at least two members of Program Committee or independent reviewers. Only 86 best papers were selected for oral presentation and publication in the 37th International Conference on Information Systems Architecture and Technology proceedings.

The book is divided into four volumes which splits papers into areas: Managing Complex Planning Environments, Systems Analysis and Modeling, Modeling of financial and Investment decisions, Risk Management, Project Management, Logistics and Market, Artificial Intelligence, Knowledge Based Management, Web Systems, Computer Networks and Distributed Computing, High Performance Computing, Cloud Computing, Multi-agent Systems, Internet of Things, Mobile Systems, Service Oriented Architecture Systems, Knowledge Discovery and Data Mining, Quality of Service, E-Business Systems.

We would like to thank the Program Committee and external reviewers, who were essential for reviewing the papers and ensuring a high standard of the ISAT 2016 Conference and its proceedings. We thank the authors, presenters, and participants of ISAT 2016, and without them the conference would not have taken place. Finally, we thank the organizing team for their efforts during this and previous years which have led to a successful conclusion of the conference.

Wrocław, Poland
September 2016

Jerzy Świątek
Zofia Wilimowska
Leszek Borzemski
Adam Grzech

ISAT 2016 Conference Organization

General Chair

Zofia Wilimowska, Poland

Program Co-Chairs

Zofia Wilimowska, Poland
Leszek Borzemski, Poland
Adam Grzech, Poland
Jerzy Świątek, Poland

Local Organizing Committee

Zofia Wilimowska, Chair
Leszek Borzemski, Vice-Chair
Adam Grzech, Vice-Chair
Jerzy Świątek, Vice-Chair
Arkadiusz Górski, Technical Editor, Conference Secretary
Anna Czarnecka, Technical Editor, Website Administrator
Agnieszka Parkitna, Conference Secretary
Anna Kamińska, Conference Secretary
Michał Kowalski, Technical Coordinator
Ziemowit Nowak, Website Support
Mariusz Fraś, Website Support

International Program Committee

Witold Abramowicz, Poland
Dhiya Al-Jumeily, UK
Iosif Androulidakis, Greece
Patricia Anthony, New Zealand
Zbigniew Banaszak, Poland
Elena Benderskaya, Russia
Leszek Borzemski, Poland
Janos Botzheim, Japan
Patrice Boursier, France
Wojciech Cellary, Poland
Haruna Chiroma, Malaysia
Edward Chlebus, Poland
Gloria Crisan, Romania
Marilia Curado, Portugal
Zhaohong Deng, China
Małgorzata Dolińska, Poland
Ewa Dudek-Dyduch, Poland
El-Sayed El-Alfy, Saudi Arabia
Naoki Fukuta, Japan
Piotr Gawkowski, Poland
Manuel Graña, Spain
Wiesław Grudzewski, Poland
Adam Grzech, Poland
Katsuhiro Honda, Japan
Marian Hopej, Poland
Zbigniew Huzar, Poland
Natthakan Iam-On, Thailand
Biju Issac, UK
Arun Iyengar, USA
Jürgen Jasperneite, Germany
Janusz Kacprzyk, Poland
Henryk Kaproń, Poland
Yannis L. Karnavas, Greece
Ryszard Knosala, Poland
Zdzisław Kowalczuk, Poland
Binod Kumar, India
Jan Kwiatkowski, Poland
Antonio Latorre, Spain
Gang Li, Australia
José M. Merigó Lindahl, Chile
Jose M. Luna, Spain
Emilio Luque, Spain
Sofian Maabout, France

Zygmunt Mazur, Poland
Pedro Medeiros, Portugal
Toshiro Minami, Japan
Marian Molasy, Poland
Zbigniew Nahorski, Poland
Kazumi Nakamatsu, Japan
Peter Nielsen, Denmark
Tadashi Nomoto, Japan
Cezary Orłowski, Poland
Sandeep Pachpande, India
Michele Pagano, Italy
George Papakostas, Greece
Zdzisław Papir, Poland
Marek Pawlak, Poland
Jan Platoš, Czech Republic
Tomasz Popławski, Poland
Edward Radosiński, Poland
Dolores I. Rexachs, Spain
José S. Reyes, Spain
Leszek Rutkowski, Poland
Gerald Schaefer, UK
Habib Shah, Malaysia
Jeng Shyang, Taiwan
Anna Sikora, Spain
Małgorzata Sterna, Poland
Janusz Stokłosa, Poland
Remo Suppi, Spain
Edward Szczerbicki, Australia
Jerzy Świątek, Poland
Eugeniusz Toczyłowski, Poland
Elpida Tzafestas, Greece
José R. Villar, Spain
Bay Vo, Vietnam
Hongzhi Wang, China
Leon S.I. Wang, Taiwan
Jan Werewka, Poland
Thomas Wielicki, USA
Zofia Wilimowska, Poland
Bernd Wolfinger, Germany
Józef Woźniak, Poland
Roman Wyrzykowski, Poland
Jaroslav Zendulka, Czech Republic
Bernard Ženko, Slovenia

ISAT 2016 Reviewers

Patricia Anthony, New Zealand
Zbigniew Antoni Banaszak, Poland
Elena Benderskaya, Russian Federation
Grzegorz Bocewicz, Poland
Leszek Borzemski, Poland
Jerzy Brzeziński, Poland
Wojciech Cellary, Poland
Krzysztof Cetnarowicz, Poland
Haruna Chiroma, Malaysia
Witold Chmielarz, Poland
Grzegorz Chodak, Poland
Robert Ryszard Chodorek, Poland
Kazimierz Choroś, Poland
Andrzej Chydziński, Poland
Gloria Cerasela Crisan, Romania
Mariusz Czekała, Poland
Pedro D. Medeiros, Portugal
Aldona Dereń, Poland
Grzegorz Dobrowolski, Poland
Ewa Dudek-Dyduch, Poland
Mariusz Fraś, Poland
Naoki Fukuta, Japan
Krzysztof Goczyła, Poland
Arkadiusz Górski, Poland
Manuel Grana, Spain
Jerzy Grobelny, Poland
Adam Grzech, Poland
Bogumila Hnatkowska, Poland
Maciej Hojda, Poland
Zbigniew Huzar, Poland
Natthakan Iam-On, Thailand
Przemysław Ignaciuk, Poland
Jerzy Józefczyk, Poland
Krzysztof Juszczyszyn, Poland
Adam Kasperski, Poland
Włodzimierz Kasprzak, Poland
Grzegorz Kołaczek, Poland
Zdzislaw Kowalczuk, Poland
Andrzej Kozik, Poland
Dorota Kuchta, Poland
Lumír Kulhánek, Czech Republic

Halina Kwaśnicka, Poland
Jan Kwiatkowski, Poland
Wojciech Lorkiewicz, Poland
Jose M. Luna, Spain
Lech Madeyski, Poland
Zbigniew Malara, Poland
Rafał Michalski, Poland
Zbigniew Nahorski, Poland
Jerzy Nawrocki, Poland
Peter Nielsen, Denmark
Tadashi Nomoto, Japan
Andrzej Nowak, Poland
Krzysztof Nowicki, Poland
Cezary Orłowski, Poland
Donat Orski, Poland
Piotr Pacyna, Poland
Michele Pagano, Italy
Agnieszka Parkitna, Poland
Marek Pawlak, Poland
Willy Picard, Poland
Jan Platoš, Czech Republic
Łukasz Popławski, Poland
Dolores Rexachs, Spain
Radosław Rudek, Poland
Jarogniew Rykowski, Poland
José Santos, Spain
Danuta Seretna-Sałamaj, Poland
Anna Sikora, Poland
Maciej Stasiak, Poland
Malgorzata Sterna, Poland
Janusz Stokłosa, Poland
Grażyna Suchacka, Poland
Remo Suppi, Spain
Joanna Szczepańska, Poland
Edward Szczerbicki, Poland
Paweł Świątek, Poland
Jerzy Świątek, Poland
Halina Tarasiuk, Poland
Kamila Urbańska, Poland
José R. Villar, Spain
Krzysztof Walczak, Poland
Zofia Wilimowska, Poland
Marek Wilimowski, Poland

Bernd Wolfinger, Germany
Jozef Wożniak, Poland
Roman Wyrzykowski, Poland
Jaroslav Zendulka, Czech Republic
Bernard Ženko, Slovenia
Maciej Zięba, Poland

Contents

Part I Modelling and Identification

Applying Hidden Markov Models to Visual Activity Analysis for Simple Digital Control Panel Operations 3
Jerzy Grobelny and Rafał Michalski

Real-Time Sliding Mode Observer Estimator Integration in Hybrid Electric Vehicles Battery Management Systems 15
Nicolae Tudoroiu, Liana Elefterie, Elena-Roxana Tudoroiu, Wilhelm Kecs, Maria Dobritoiu and Nicolae Ilias

Recursive Identification of Nonlinear Aerodynamic Characteristics of Air-Plane 29
Jacek Pieniążek and Piotr Cieciński

Actor Model Approach to Control Stability of an Articulated Vehicle 41
Kornel Warwas and Krzysztof Augustynek

Part II Image Processing

Automatic Detection of Nerves in Confocal Corneal Images with Orientation-Based Edge Merging 55
Adam Brzeski

Methods of Digital Hilbert Optics in the Analysis and Objects' Recognition 65
Adam Sudol

Light-Reflection Analysis Method for 3D Surface Damage Identification 79
Michał Turek and Dariusz Pałka

Part III Operation Research Applications

Applications of Operations Research and Intelligent Techniques in the Health Systems 91
Marek Lubicz

A Simulation Model of Aircraft Ground Handling: Case Study of the Wroclaw Airport Terminal 109
Artur Kierzkowski and Tomasz Kisiel

Analysis of the Impact of Changes in the Size of the Waste Stream on the Process of Manual Sorting of Waste 127
Robert Giel and Marcin Plewa

The Extension of User Story Template Structure with an Assessment Question Based on the Kano Model 137
Grażyna Hołodnik-Janczura

Part IV Manufacturing Systems

Comparison of Discrete Rate Modeling and Discrete Event Simulation. Methodological and Performance Aspects 153
Jacek Zabawa and Edward Radosiński

Simulation Tool for Effective Tasks Subcontracting in Manufacturing Networks 165
Joanna Gąbka

Distribution of Inversions and the Power of the τ- Kendall's Test 175
Mariusz Czekala and Agnieszka Bukietyńska

Part V Network and Transport Systems

Reduction of Congestion in Transport Networks with a Fractal Structure 189
Grzegorz Bocewicz, Zbigniew Banaszak and Izabela Nielsen

Modelling of Switching Networks with Multiservice Traffic by the IPGBMT Method 203
Mariusz Głąbowski and Michał Dominik Stasiak

Time to Buffer Overflow in a Finite-Capacity Queueing Model with Setup and Closedown Times 215
Wojciech M. Kempa and Iwona Paprocka

The Multiple Criteria Optimization Problem of Joint Matching Carpoolers and Common Route Planning 225
Grzegorz Filcek and Jacek Żak

The Impact of Structure Network Parameters on Consumers Behavior: A Cellular Automata Model 237
Agnieszka Kowalska-Styczeń

Author Index 249

Part I
Modelling and Identification

Applying Hidden Markov Models to Visual Activity Analysis for Simple Digital Control Panel Operations

Jerzy Grobelny and Rafał Michalski

Abstract The paper presents an application of Hidden Markov Models (HMM) to fixations' sequences analysis. The examination concerns eye tracking data gathered during performing simple comparison and decision tasks for four versions of plain control panels. The panels displayed the target and current velocity either on a digital or analog (clock-face) speedometers. Subjects were to decide whether increase or decrease the current speed by pressing the appropriate button. The obtained results suggest that females, generally exhibit different covert attention patterns than men. Moreover, the article demonstrates the estimated four HMM with three hidden states for every examined control panels variant and provides discussion of the outcomes.

Keywords Ergonomics · Control panel design · Human visual behavior · Human-computer interaction

1 Introduction

The human visual activity may be analyzed by the eye tracking investigations. The visual scanpath typically registered by such systems consists of a sequence of fixations and saccades. Generally, remaining for a longer time within a specific, constraint area is considered to be a fixation while rapid, ballistic changes in eye position are called saccades. It is widely believed, and assumed that the visual processing takes place when fixating whereas during saccadic jumps the information extraction is suppressed. By examining a series of fixations and saccades one

J. Grobelny · R. Michalski (✉)
Wrocław University of Science and Technology, Wrocław, Poland
e-mail: rafal.michalski@pwr.edu.pl
URL: http://www.RafalMichalski.com

J. Grobelny
e-mail: jerzy.grobelny@pwr.edu.pl
URL: http://www.JerzyGrobelny.com

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_1

can infer about the attention shifting process that accompanies visual task executions.

Ellis and Stark [6] suggested that scanpaths can be modeled by a stochastic first-order Markov process, where the fixation position depends on the location of the previous fixation. In recent years more and more popular becomes the extension of this approach to Hidden Markov Models. The interest in this type of analysis is directly related to latest findings from visual processing psychology that differentiated between the overt and covert attention [8]. Observable eye movements are obviously associated with overt attention while hidden states from HMM are coupled with shifting covert attention. HMM tools allow for finding stochastic relations between observations and hidden states and, thus, are very useful in broadening the basic knowledge about attentional visual processes.

Liechty et al. [11], for instance, showed in their study, that registered during visual analysis of printed advertisements saccades' lengths may be explained by means of the two-states HMM. These states, according to authors reflect two states of covert attention—local and global. In the work of Hayashi [10], in turn, it was demonstrated that the HMM states estimated from pilots' fixations can be associated with tasks performed during plane landing. The number of hidden states obtained from HMM indicates that pilots organize the covert attention changing process differently, depending on their experience. A similar approach applied by Chuk et al. [4] to the analysis of the face recognition task yielded various visual activity strategies among individuals. The authors identified two specific attention patterns—holistic and analytic. Experiments conducted by Simola et al. [17] revealed that it is possible to discover particular, common patterns of attention used during textual information search. In this case, the HMM scanpaths investigation provided evidence for existing three consecutive hidden states defined as scanning, reading and the answer.

The problem of identification of various tasks performed by humans based on their visual activity registered in a scanpath form is subject to investigation of many researchers. The importance of this issue results, among other things, from its possible practical application in the domain of intelligent interactive systems. In such a perspective the HMM may play a significant role. For example, Haji-Abolhassani and Clark [9] studied visual observations tasks of black and white pictures and showed using HMM and clustering methods that it is possible to detect the visual task type given the specific eye ball movements data. Furthermore, Courtemanche et al. [5] proposed a framework allowing for identification of tasks having hierarchical structures in human-computer interaction processes. This was modeled by a Layered version of HMM.

The present study is focused on the visual activity observed during performing easy human-machine interaction tasks. The analyses, carried out in the HMM perspective aim at attempting to answer fundamental questions inspired by the abovementioned research. First, what types of the visual strategies and attention shifts patterns are employed by subjects while operating control panels and how do they depend on the interface design? Secondly, what are the qualitative differences

(if any) between subjects' visual activities? The latter issue was examined between males and females.

The remainder of the article is organized as follows. At first, the idea of the HMM is shortly described, then the control panel experiment is overviewed. Next, HMM simulation results are demonstrated and discussed.

2 Hidden Markov Models Overview

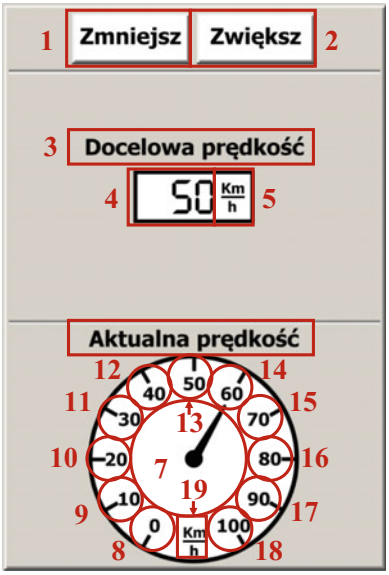
An HMM is usually specified by the following components (see e.g. [15]): (1) a set of N states in a model: $\boldsymbol{S} = \{s_1, s_2, \ldots, s_N\}$, (2) a group of M observation symbols for each state called also the vocabulary or alphabet: $\boldsymbol{V} = \{v_1, v_2, \ldots, v_M\}$, (3) a states' transition probability matrix which specifies the probability of moving from state i to state j: $\boldsymbol{A} = \{a_{ij}\}$, where $a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i]$; $1 \leq i, j \leq N$; q_t is the state at time t, (4) The observation symbol likelihoods distribution in state j, also called emission probabilities matrix: $\boldsymbol{B} = \{b_j(k)\}$, where $b_j(k) = P[v_k \text{ at t} \mid q_t = S_j]$; $1 \leq j \leq N$, $1 \leq k \leq M$, (5) the Markov chain starting probability distribution $\boldsymbol{\pi} = \{\pi_i\}$, where $\pi_i = P[q_1 = S_i]$, $1 \leq i \leq N$. In short, the full HMM model requires the specification of states, symbols, and three sets of probabilities distributions: $\lambda = (A, B, \pi)$. Such a model can generate the sequence of observations: $O = O_1 O_2 \ldots O_T$, where every O_t is one of the V symbols, and T is the observation sequence length. In the present study, a set of (A, B, π) are being sought given the observation sequences O, assumed number of states N and a defined vocabulary V.

3 Control Panels Experiment Overview

The current paper Hidden Markov modeling is based on the eye tracking results from an experimental study presented by Michalski [12]. Therefore, only a very brief overview of this investigation is provided in this section.

The visual activity data came from 33, young subjects (15 women, and 18 men) and were registered by a SMI RED500 stationary eye tracker. The participants performed simple comparison and decision making tasks regarding the digital control panels displayed by JavaScript in a web browser. Specifically, they were to compare target and current velocities, decide whether to increase or decrease the current value to get the desired one, and click with a mouse on the proper button.

The designed stimuli differed in the speedometer types. Digital and analog (clock-face) speedometer versions and types of displayed information produced four experimental conditions: (1) **AA**—target and current velocities presented on analog speedometers, (2) **AD**—the target demonstrated on an analog while the current on a digital version, (3) **DA**—the target on digital and the current on analog, and finally both velocities presented on digital speedometers (4) **DD**. Figure 1 shows one (DA) of the four control panel types examined in this study, with



1. **D** (Decrease)
2. **I** (Increase)

3. **TD-Txt** (Target Digital)
4. **TD-V** (Velocity Value)
5. **TD-Kmh**

6. **CA_Txt** (Current Analog)
7. **CA_Cen** (Center)

8. **CA-V0**
9. **CA-V10**
10. **CA-V20**
11. **CA-V30**
12. **CA-V40**
13. **CA-V50**
14. **CA-V60**
15. **CA-V70**
16. **CA-V80**
17. **CA-V90**
18. **CA-V100**

19. **CA_Kmh**

Fig. 1 Areas of interests defined for one of the digital control panel variants. The AOIs and their abbreviations presented next to the panel were analogically created for the remaining three speedometer type combinations

superimposed areas of interests used for the present study modeling purposes. Subjects performed six trials per each speedometer type combination factor (STC), they resulted from combining three target velocity values (20, 50 and 80 km/h), and the required response type (Increase, or Decrease). Each subject tested all types of control panels presented randomly.

4 Modeling Visual Activity by Hidden Markov Models

4.1 Preparation of Eye Tracking Data for HMM

The idea of preparing the data for HMM was similar to the one used in a work of Hayashi [10]. However, instead of using the whole instruments as the HMM symbols, the present study employs much smaller panel components for this purpose. The identified areas of interests for one of the examined panel versions (DA) are graphically illustrated in Fig. 1. The AOIs and their abbreviations presented next to the panel were analogically created for the remaining three speedometer type combinations The sequences of fixations observed on these salient panel components towards which subjects direct their attention were used to train the HMM. The data were exported by a SMI BeGaze software Ver. 3.6, and processed in the STATISTICA ver. 12.

4.2 HMM Estimation

The initial, transition, and emission probabilities for first order, discrete time, hidden Markov models were estimated using the Baum-Welch algorithm [2] implemented by Murphy [13]. The maximum number of iterations in this algorithm was set at 1000 and the convergence threshold was specified at 0.0001. All the estimations and calculations were performed in a MatLab (7.11.0) R2010b package.

4.3 Gender and Speedometer Effects in HMM

A simulation experiment was performed to verify how many hidden states will best model the four versions of analyzed control panels. Since the findings regarding a number of subjects' visual behavior characteristics while performing the experimental task presented by Michalski [12] suggest that females probably apply qualitatively different visual strategies than males, the models were examined separately for men and women. Overall 32 conditions were investigated involving four possible hidden states (from 3 to 6), four control panel types (AA, AD, DA, DD), and the gender factor: 4 (States) $\times$ 4 (STC) $\times$ 2 (Gender).

As the estimated HMM parameters may depend on their starting values, each time for every estimated model 100 simulations were performed with randomly chosen initial states, transitions', and emissions' probabilities. The models were assessed by means of the Akaike's [1] Information Criterion (AIC), and Schwarz's [16] Bayesian Information Criterion (BIC). The obtained mean (standard deviations in brackets) and minimum values for AIC, BIC along with log-likelihoods are put together in Table 1. The smaller values of AIC and BIC signify more accurate models.

The evaluation of obtained models in these simulations according the AIC and BIC criteria is ambiguous. The AIC values suggest that the gender factor significantly differentiates the attentional process dynamics recorded during executing experimental tasks. The data show that generally the smaller number of hidden states is characteristics of females than males. Almost for all examined control panel interfaces, the more accurate models required one state less for females than males. Given the tiny differences between the two smallest AIC values in two cases that is, AA and DD for females and models with a smaller number of hidden states were chosen due to the parsimony principle.

Hayashi [10] analyzing the pilots eye movement data during landing found that their visual behavior described by the optimal number of hidden states depended on their flying experience. Models with a bigger number of hidden states better fit to the experimental data for more experienced pilots. The author explains the fact by a more detailed analysis of instruments which leads to a more detailed structure of the mental attention model. Although in our experimental tasks the experience is not directly connected with gender, but it seems that designed control panel operations

Table 1 The hidden Markov model simulation results. Standard deviations in brackets

No.	STC	Gender	No. of states	AIC		BIC		Log-likelihood	
				Mean	Min	Mean	Min	Mean	Min
1	AA	Male	3	4320 (33)	4265	4784 (33)	**4728**	−2061 (16)	−2106
2	AA	Male	4	4314 (33)	4261	4950 (33)	4897	−2021 (17)	−2056
3	AA	Male	**5**	4318 (34)	**4256**	5137 (34)	5075	−1984 (17)	−2033
4	AA	Male	6	4326 (31)	4271	5337 (31)	5282	−1947 (16)	−1989
5	AA	Female	3	3906 (44)	3862	4360 (44)	**4316**	−1854 (22)	−1911
6	AA	Female	**4**	3831 (52)	**3785**	4455 (52)	4409	−1780 (26)	−1870
7	AA	Female	5	3843 (44)	**3785**	4646 (44)	4588	−1747 (22)	−1815
8	AA	Female	6	3859 (34)	3808	4850 (34)	4798	−1714 (17)	−1772
9	AD	Male	3	3802 (21)	3777	4127 (21)	**4102**	−1832 (10)	−1870
10	AD	Male	4	3781 (22)	3756	4234 (22)	4209	−1795 (11)	−1841
11	AD	Male	5	3766 (27)	3735	4355 (27)	4324	−1758 (14)	−1803
12	AD	Male	**6**	3768 (24)	**3725**	4504 (24)	4460	−1728 (12)	−1763
13	AD	Female	3	3122 (25)	3090	3436 (25)	**3403**	−1492 (13)	−1546
14	AD	Female	4	3077 (44)	3032	3513 (44)	3468	−1443 (22)	−1492
15	AD	Female	**5**	3056 (33)	**3020**	3624 (33)	3587	−1403 (16)	−1452
16	AD	Female	6	3064 (26)	3027	3773 (26)	3735	−1376 (13)	−1438
17	DA	Male	3	2913 (33)	2889	3211 (33)	**3187**	−1391 (16)	−1436
18	DA	Male	4	2883 (27)	2853	3298 (27)	3267	−1350 (14)	−1396
19	DA	Male	**5**	2877 (29)	**2842**	3418 (29)	3383	−1318 (14)	−1364
20	DA	Male	6	2889 (22)	2860	3565 (22)	3536	−1294 (11)	−1329
21	DA	Female	3	2346 (35)	2333	2631 (35)	**2619**	−1107 (18)	−1209
22	DA	Female	**4**	2312 (32)	**2280**	2709 (32)	2677	−1064 (16)	−1103
23	DA	Female	5	2309 (25)	2287	2828 (25)	2805	−1035 (12)	−1101
24	DA	Female	6	2332 (14)	2309	2980 (14)	2957	−1016 (7)	−1056
25	DD	Male	3	1889 (22)	1859	2044 (22)	**2014**	−908 (11)	−928
26	DD	Male	4	1868 (22)	1851	2093 (22)	2076	−882 (11)	−914
27	DD	Male	**5**	1860 (19)	**1840**	2162 (19)	2142	−860 (9)	−886
28	DD	Male	6	1868 (14)	1849	2256 (14)	2238	−844 (7)	−874
29	DD	Female	3	1513 (34)	1483	1661 (34)	**1631**	−721 (17)	−743
30	DD	Female	**4**	1478 (26)	**1464**	1691 (26)	1677	−687 (13)	−725
31	DD	Female	5	1479 (15)	**1463**	1767 (15)	1750	−669 (8)	−719
32	DD	Female	6	1493 (12)	1475	1863 (12)	1845	−656 (6)	−682

may be associated with car driving or computer simulation games—and these are rather closer to men. The above finding may provide a possible explanation of the somewhat unexpected result described in Michalski [12]. The analysis of effectiveness showed there that women committed more than twice as many mistakes (7.8 %) while performing trials than men (3.7 %, $\chi^2 = 6.1$, $p = 0.013$) which is in contrast with many research (e.g. [3]). In light of the present study findings, this phenomenon could probably result from different covert attention distribution for

females and males and worse mental representation of the studied problem in women. The suggested problems with correctly mapping the experimental task to females' internal mental representations were additionally visible in markedly longer fixations (Michalski [12]) associated with more difficult tasks.

Analyzing simulation experiment results, one can also notice that the optimum number of hidden states (according to AIC) across the investigated control panel types is quite similar and amounts to 5 for men and 4 for females. Only for the Analog—Digital option models with a bigger number of states (6 states for males and 5 for women) are more accurate. This, in turn, sheds some more light on another surprising result from the analysis provided in Michalski [12] regarding the statistically significantly worst mean operation times for the AD versus DA.

It could be predicted that having identical combination of speedometer types presenting the same type of physical value (velocity), the efficiency results would be comparable. However, as it can be observed from a different number of desired hidden states in demonstrated HMM, these versions of control panels are not equivalent. It seems that attentional patterns necessary to process target and current velocity values are considerably different. This discrepancy probably underlain the differences between AD versus DA registered for task completion times, saccade counts, as well as scanpath lengths.

The AIC and BIC indicators penalize the log-likelihood in such a way that the model assessment takes the maximum likelihood principle when the number of model parameters to be estimated is as small as possible. In the BIC criterion the penalization effect is more severe than in AIC which can be seen in Table 1. According to this criterion the suggested best model for all examined conditions includes only three hidden states. The search for three states HMMs depending on the control panel design is an attempt to answer the second research question of this study.

4.4 Analysis of HM Models for Speedometer Type Combination

Analogically as in the examination described in the previous section a series of 100 HMM simulations were performed for all four speedometer type combinations assuming three hidden states. From the obtained group of models, the ones with the smallest log-likelihood value were chosen and presented in Table 2. The initial states probabilities are displayed in the second row, while the three further rows contain the transition probabilities distribution. The remaining rows include the emission matrix probabilities.

The HM models (A, B, π) estimated from the fixation sequences registered for all subjects examining the given control panel interface variant exhibit far-reaching stability in attention shifting strategy between these three states. The outcomes seem to be quite similar irrespective of the panel construction. The analysis of

Table 2 The three states hidden Markov models for all speedometer type combinations

(a) AA	S1	S2	S3	(b) AD	S1	S2	S3	(c) DA	S1	S2	S3	(d) DA	S1	S2	S3
π	0.03	0.84	0.13	**π**	0.01	0.74	0.25	**π**	0.04	0.85	0.11	**Π**	0.02	0.81	0.17
S1	0.91	0.00	0.37	**S1**	0.84	0.02	0.28	**S1**	0.95	0.01	0.41	**S1**	0.88	0.07	0.40
S2	0.09	0.68	0.00	**S2**	0.14	0.67	0.03	**S2**	0.03	0.60	0.00	**S2**	0.10	0.60	0.03
S3	0.00	0.32	0.63	**S3**	0.02	0.31	0.69	**S3**	0.02	0.39	0.59	**S3**	0.02	0.33	0.57
TA-Txt	0.13	0.26	0.00	TA-Txt	0.10	0.31	0.00	TD-Txt	0.12	0.42	0.00	TD-Txt	0.14	0.44	0.00
TA-Cen	0.12	0.30	0.03	TA-Cen	0.04	0.30	0.15	TD-V	0.20	0.45	0.00	TD-V	0.19	0.47	0.02
TA-V0	0.01	0.01	0.00	TA-V0	0.00	0.01	0.00	TD-Kmh	0.01	0.01	0.00	TD-Kmh	0.00	0.03	0.00
TA-V10	0.01	0.00	0.03	TA-V10	0.00	0.01	0.01	CA-Txt	0.02	0.01	0.27	CD-Txt	0.00	0.00	0.56
TA-V20	0.01	0.03	0.06	TA-V20	0.01	0.04	0.06	CA-Cen	0.00	0.08	0.27	CD-V	0.00	0.04	0.41
TA-V30	0.01	0.03	0.01	TA-V30	0.02	0.02	0.00	CA-V0	0.00	0.00	0.00	CD-Kmh	0.00	0.00	0.01
TA-V40	0.01	0.05	0.00	TA-V40	0.01	0.06	0.00	CA-V10	0.00	0.00	0.05	D	0.32	0.02	0.00
TA-V50	0.05	0.10	0.01	TA-V50	0.03	0.12	0.03	CA-V20	0.00	0.00	0.03	I	0.35	0.00	0.00
TA-V60	0.00	0.00	0.00	TA-V60	0.02	0.01	0.00	CA-V30	0.00	0.00	0.09				
TA-V70	0.02	0.02	0.01	TA-V70	0.01	0.01	0.00	CA-V40	0.00	0.00	0.10				
TA-V80	0.03	0.04	0.00	TA-V80	0.00	0.02	0.05	CA-V50	0.00	0.01	0.05				
TA-V90	0.00	0.01	0.01	TA-V90	0.00	0.00	0.01	CA-V60	0.04	0.01	0.02				
TA-V100	0.02	0.01	0.00	TA-V100	0.00	0.00	0.01	CA-V70	0.00	0.00	0.06				
TA-Kmh	0.00	0.02	0.00	TA-Kmh	0.00	0.03	0.00	CA-V80	0.00	0.00	0.01				
CA-Txt	0.00	0.00	0.24	CD-Txt	0.02	0.00	0.34	CA-V90	0.01	0.00	0.04				
CA-Cen	0.00	0.04	0.21	CD-V	0.00	0.01	0.32	CA-V100	0.00	0.00	0.01				
CA-V10	0.00	0.00	0.04	CA-Kmh	0.01	0.00	0.01	D	0.30	0.00	0.00				
CA-V20	0.00	0.00	0.03	D	0.44	0.04	0.00	I	0.30	0.00	0.00				
CA-V30	0.00	0.00	0.07	I	0.30	0.01	0.02								
CA-V40	0.00	0.01	0.07												

(continued)

Table 2 (continued)

(a) AA	S1	S2	S3	(b) AD	S1	S2	S3	(c) DA	S1	S2	S3	(d) DA	S1	S2	S3
CA-V50	0.00	0.02	0.03												
CA-V60	0.01	0.00	0.03												
CA-V70	0.00	0.00	0.05												
CA-V80	0.00	0.00	0.01												
CA-V90	0.00	0.00	0.04												
CA-V100	0.00	0.00	0.01												
CA_Kmh	0.00	0.00	0.00												
D	0.31	0.04	0.00												
I	0.27	0.01	0.00												

models' parameters allow for relating the obtained hidden states with the executed tasks. Examining, for example, the DA control panel from Fig. 1 and the HMM for this variant presented in Table 2(c) one can easily notice that the most probable starting state is S_2 ($p = 0.85$). The biggest emission probabilities associated with this state indicate that it is strongly related with fixating on the target velocity. Similar values of probabilities were estimated for the target caption and value (TD-Txt, TD–V): 0.42 and 0.45 respectively. Thus, the S_2 can be specified as the ***Target identification state***. The S_3 state generates the biggest emission probabilities for CA-Txt and CA-Cen symbols (0.27) that is, the areas presenting the current velocity caption and the analog speedometer indicator. As a result of this, the S_3 can be called ***Current state analysis***. The last state S_1 with maximum probabilities values (0.30) for D and I (Decrease and Increase) buttons observations can be described as ***Decision and execution***. Comparable patterns of probabilities distributions for $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{\pi}$ are visible in all versions of investigated control panels, thus similar process of identifying state's meaning can also be replicated for the remaining variants.

Relatively high values in diagonals of the states' transmission matrix A. as well as similar emission probabilities for symbols CA-Txt and CA-Cen as well as for TD-Txt and TD-V suggest the existence of a hierarchical states' structure. For instance, while performing the ***Target identification***, it is highly probable that subjects would stay in the same state in the next step. In such a case, the sub-goals of the S_2 could be described as the ***Indicator location*** and ***Reading the velocity value***. Thus, it may be assumed that participants were moving their eyes according to the spot-light principle [14] at first, and when they came across the clock-face type of speedometer they changed their visual behavior to typical to the zoom-lens model (Eriksen and James [7]). This finding additionally supports similar conjectures put forward in Michalski [12] and confirms that subjects may apply complex attention shifting strategies even for such relatively simple experimental tasks. Moreover, the whole process may probably be an interesting research direction and possibly modeled by means of Layered HMM (Courtemanche [5]).

5 Conclusions

The current study results are promising. The human visual behavior analysis in the HMM perspective provides opportunities of discovering basic attention shifting strategies while operating the control panel. The performed simulations revealed the usefulness of the HMM in discovering attention states characteristic of the human-machine interaction. The obtained models have logical interpretation in categories of real experimental subtasks. What is more, the interpretations appear to be independent of the interface design details which may result in practical applications involving automatic identification of tasks performed by operators in human-machine systems.

Furthermore, such a possibility can be a basis of developing intelligent systems for interactive products and industrial machinery. The findings regarding gender differences in attentional processes can be used for examining individual discrepancies among human-machine operators. A particular area of practical applications may regard e.g. the qualifications or training effectiveness assessment.

Generally, in contrast to classic, static and deterministic scanpath analysis which is based solely on the overt attention observations, the identification of the hidden states has its origin in physiological and psychological processes related with the covert attention. Thus, the ability to look inside the human visual processing is invaluable in extending the knowledge about the human-machine interaction. This would certainly lead to better understanding of the nature of communication with interactive products.

Acknowledgments The work was partially financially supported by Polish National Science Centre Grant No. 2011/03/B/ST8/06238. The eye tracking data were recorded by the system made available by the Laboratory of Information Systems Quality of Use which is a part of a BIBLIOTECH project cofounded by the European Union through the European Regional Development Fund under the Operational Programme Innovative Economy 2007–2013.

References

1. Akaike, H.: Information theory as an extension of the maximum likelihood theory. In: Petrov, B.N., Csaki, F. (eds.) Second International Symposium on Information Theory, pp. 267–281. Akademiai Kiado, Budapest (1973)
2. Baum, L.E.: An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In: Shisha, O. (ed.) Proceedings of the 3rd Symposium on Inequalities, University of California, Los Angeles, pp. 1–8 (1972)
3. Blatter, K., Graw, P., Munch, M., Knoblauch, V., Wirz-Justice, A., Cajochen, C.: Gender and age differences in psychomotor vigilance performance under differential sleep pressure conditions. Behav. Brain Res. **168**(2), 312–317 (2006). doi:10.1016/j.bbr.2005.11.018
4. Chuk, T., Chan, A.B., Hsiao, J.H.: Understanding eye movements in face recognition using hidden Markov models. J. Vis. **14**(11), 1–14 (2014). doi:10.1167/14.11.8
5. Courtemanche, F., Aïmeur, E., Dufresne, A., Najjar, M., Mpondo, F.: Activity recognition using eye-gaze movements and traditional interactions. Interact. Comput. **23**(3), 202–213 (2011). doi:10.1016/j.intcom.2011.02.008
6. Ellis, S.R., Stark, L.: Statistical dependency in visual scanning. Hum. Factors J. Hum. Factors Ergon. Soc. **28**(4), 421–438 (1986). doi:10.1177/001872088602800405
7. Eriksen, C.W., James, J.D.S.: Visual attention within and around the field of focal attention: a zoom lens model. Percept. Psychophys. **40**(4), 225–240 (1986). doi:10.3758/BF03211502
8. Findlay, J.M., Gilchrist, I.D.: Active Vision. The Psychology of Looking and Seeing. Oxford University Press, New York (2003)
9. Haji-Abolhassani, A., Clark, J.J.: An inverse Yarbus process: predicting observers' task from eye movement patterns. Vis. Res. **103**, 127–142 (2014). doi:10.1016/j.visres.2014.08.014
10. Hayashi, M.: Hidden Markov Models to identify pilot instrument scanning and attention patterns. In: IEEE International Conference on Systems, Man and Cybernetics, 2003, vol. 3, pp. 2889–2896 (2003). doi:10.1109/ICSMC.2003.1244330

11. Liechty, J., Pieters, R., Wedel, M.: Global and local covert visual attention: Evidence from a bayesian hidden markov model. Psychometrika **68**(4), 519–541 (2003). doi:10.1007/BF02295608
12. Michalski, R.: Information presentation compatibility in the simple digital control panel design—eye-tracking study. In: European Network Intelligence Conference—ENIC 2016, 5–7 September, Wroclaw, Poland (2016)
13. Murphy, K.: Hidden Markov Model (HMM) Toolbox for Matlab (1998, 2005). www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html
14. Posner, M.I., Snyder, C.R., Davidson, B.J.: Attention and the detection of signals. J. Exp. Psychol. Gen. **109**(2), 160–174 (1980). doi:10.1037/0096-3445.109.2.160
15. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989). doi:10.1109/5.18626
16. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978). doi:10.1214/aos/1176344136
17. Simola, J., Salojärvi, J., Kojo, I.: Using hidden Markov model to uncover processing states from eye movements in information search tasks. Cogn. Syst. Res. **9**(4), 237–251 (2008). doi:10.1016/j.cogsys.2008.01.002

Real-Time Sliding Mode Observer Estimator Integration in Hybrid Electric Vehicles Battery Management Systems

Nicolae Tudoroiu, Liana Elefterie, Elena-Roxana Tudoroiu, Wilhelm Kecs, Maria Dobritoiu and Nicolae Ilias

Abstract In this paper we develop and implement a real-time sliding mode observer estimator (SMOE) for state-of-charge (SOC) and for current fault in Li-Ion batteries packs integrated in the battery management systems (BMS) structure of hybrid electric vehicles (HEVs). The estimation of SOC is critical in automotive industry for successful marketing of both electric vehicles (EVs) and hybrid electric vehicles (HEVs). Gradual capacity reduction and performance decay can be evaluated rigorously based on the current knowledge of rechargeable battery technology, and consequently is required a rigorous monitoring and a tight control of the SOC level, necessary for increasing the operating batteries lifetime. The novelty of this paper is that the proposed estimator structure can be also tailored to estimate the SOC and the possible faults that could occur inside of the batteries of different chemistry by augmenting the dimension of the model states, according to the number of estimated battery faults. The preliminary results obtained in this research are encouraging and reveal the effectiveness of the real-time implementation of the proposed estimator in a MATLAB/SIMULINK programming simulation environment.

N. Tudoroiu (✉)
Engineering Technologies, John Abbott College, Sainte-Anne-de-Bellevue, Quebec, Canada
e-mail: nicolae.tudoroiu@johnabbott.qc.ca

L. Elefterie
Spiru Haret University, Faculty of Economic Sciences, Constanta, Romania
e-mail: elefterie.liana@spiruharet.ro

E.-R. Tudoroiu · W. Kecs · M. Dobritoiu
Faculty of Sciences, University of Petrosani, Petrosani, Romania
e-mail: tudelena@mail.com

W. Kecs
e-mail: wwkecs@yahoo.com

M. Dobritoiu
e-mail: mariadobritoiu@yahoo.com

N. Ilias
Mechanical and Electrical Engineering, University of Petrosani, Petrosani, Romania
e-mail: iliasnic@yahoo.com

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_2

Keywords Sliding mode observer estimator · State-of-charge estimation · Fault detection and estimation · Battery management system · Hybrid electric vehicles

1 Introduction

The state-of-charge (SOC) of a battery is defined as its available capacity expressed as a percentage of its rated capacity. More specifically the SOC is the remaining capacity of the battery affected by its load current and temperature [1]. The battery SOC estimation is essential in automotive industry for both electric vehicles (EVs) and hybrid electric vehicles (HEVs). Furthermore the SOC of a battery, representing a critical condition parameter for battery management system (BMS) [1–3]. Also an accurate estimate of the battery SOC is the main issue for its healthy and safe operation no matter its chemistry. Today the most advanced and promising technology in batteries field is the production of nickel-metal hydride (Ni-MH) and lithium-ion (Li-Ion) batteries [1–3, 4]. The both batteries are rechargeable, and consequently they are of a great prospective for automotive industry to be used on a large scale to plug-in hybrid electric vehicles (PHEVs), hybrid electric vehicles (HEVs), and battery electric vehicles (BEVs). By now it was proved that the Ni-MH batteries lead the competition for a lot of automotive industry applications in electric and hybrid electric vehicles since they are quite inexpensive and have a high specific energy, high specific power, long cycle life, and also no poisonous heavy metals [1–3]. The main disadvantage of these batteries is that they have to be often completely discharged in order to avoid the "memory" effect that reduces the battery's life [1–3]. They also produce more heat during the charging cycles, and specially in operating conditions at heavy loads the heat increases drastically and thus can reduce significantly the battery's life. In contrast, Li-Ion batteries are more expensive, but lighter in weight and they have a small size, thus easy to be incorporated inside the vehicle structure. It is worth to mention that they are not affected by a "memory" effect so they should not be completely discharged in order to keep up the battery for a long time; therefore they do not require any maintenance [1–3]. Moreover Li-Ion batteries can be stored for a long time by keeping intact their charge. By reason of their great potential to store higher specific energy and energy density, the implementation of Li-Ion batteries is expected to grow very fast in EVs, mainly in PHEVs and BEVs [1–3, 4]. They perform better if they never are fully charged or discharged, therefore an accurate on-board SOC estimate has to be strictly controlled or a protection switch has to be installed; therefore the equipment complexity and its cost increase. Also at extreme temperatures Li-Ion batteries do not perform well, so a cooling and heating system is required to be installed in order to increase their life's time [1–3, 4]. The Li-Ion batteries are capable to provide enough power to boost up the acceleration and to increase the energy efficiency through on-board battery energy storage [1–3, 4]. The gradual capacity attenuation and performance degradation of the Li-Ion battery during its time operation can be evaluated strictly based on the current knowledge of rechargeable battery technology, and consequently an

accurate estimate and a tight control of the SOC level in order to increase the operating battery's lifetime are required [1–3]. In this paper we investigate new directions in order to develop the most accurate SOC estimator combined with the most suitable fault detection, isolation and reconstruction (FDIR) technique. In second section we choose the most suitable Li-Ion battery model required to develop in third section a robust sliding mode observer estimator (SMOE) able to estimate with accuracy the SOC of a high power Li-Ion battery SOC for a suburban small HEVs. Therefore the purpose of this paper is to built a robust, accurate SOC estimator, easy to be implemented in real-time, as a viable alternative to Kalman filter estimation techniques [3, 5, 6] developed until now in the literature. The proposed SOC estimator is based on a simple generic battery model described in a state-space representation by two differential first-order state equations. For simulation purpose, in order to show its effectiveness the generic battery model is tailored on a particular Li-Ion battery chemistry.

2 The Li-Ion Battery Model

Until now in the literature are reported three developed fundamental types of battery models, particularly the experimental, electrochemical and electric circuit-based [1–3, 5, 4]. The first two models seems to be inappropriate to represent the cell dynamics of battery packs from state-of-charge (SOC) estimation viewpoint, compared to the electric circuit-based models that are very useful to represent the electrical characteristics of the batteries [2, 3, 4]. The simplest electric model, known as the Thevenin model, consists of an ideal voltage source in series with an internal battery cell resistance [4]. A battery cell is the smallest unit connected in parallel or in series to form one module. A module is then connected in a parallel or series configuration to form one battery pack, as is integrated in the vehicle [1–3, 4]. The voltage measured between the battery pack terminals when a load is applied is called terminal voltage or the measured output of the battery model, and the voltage measured between the battery pack or cell terminals when no load is applied is called open-circuit voltage (OCV) [1–3, 4]. The input of the model is the charging or discharging current. Battery charge rate, denoted by C-rate, describes the rate at which the battery is charged or discharged relative to its maximum capacity [1–3, 4]. A 1C rate means that the applied discharge current will discharge a fully charged battery in 1 h. For example a battery with a capacity of 6 Ah, this equals to a 6 A discharge current. A 5C rate for this battery would be $6 \times 5 = 30$ A, and a C/4 rate would be $6/4 = 1.5$ A. The identification of the electric circuit parameters (the values of the resistances and capacitors) for the model is based on a quite complex technique called impedance spectroscopy [1–3, 4]. Shepherd developed a model described by a first order differential equation that capture the electrochemical behavior of a battery in terms of the measured terminal voltage and discharge current, OCV, internal resistance, and SOC [2, 3, 4]; it is suitable for discharge as well as for charge, and also for SOC estimation. Compared to other models, the

Shepherd model is much more interesting but causes algebraic loop instability in the closed-loop simulation of modular models [2, 3, 4]. Generic battery models that are described by a nonlinear first-order differential state equation with only SOC as a state variable are discussed in [2, 3, 5, 4]. These models are very similar to Shepherd's but don't generate algebraic loop instability in closed-loop simulations [1–3, 4]. This simple model using only SOC as a state variable is capable to reproduce precisely the manufacturer's OCV characteristic curves of the battery under investigation versus its state of charge (SOC). Typically, the values of the battery model parameters are obtained by using a nonlinear least squares curve fitting method based on the nonlinear battery discharging characteristic curves OCV versus SOC for a particular battery chemistry, as is shown in [1–3, 5, 4]. For simulation purpose, as case study we choose a Li-Ion battery model, such as the one developed in [3, 5, 4] that is a simple electric circuit, known as the 1st order RC model consisting of an open circuit voltage (OCV) in series with an internal resistance R_i and one parallel RC circuit [4]. In Figs. 1 and 2 we show the schematic diagrams of the electrical circuit battery model built in a National Instruments MULTISIM

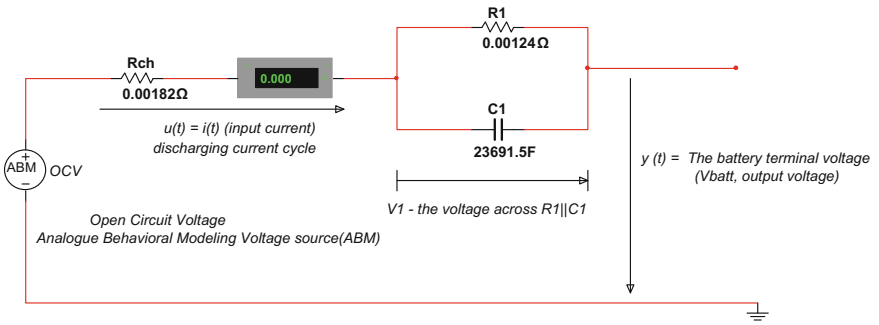


Fig. 1 Schematic diagram of the RC 1st order battery model for charging cycle (in MULTISIM 11 editor)

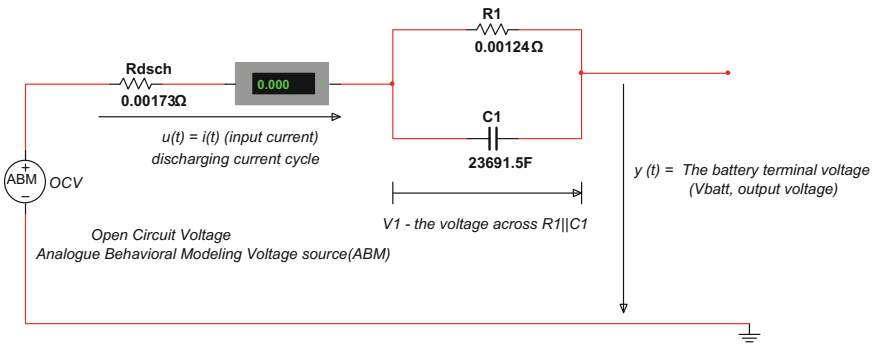


Fig. 2 Schematic diagram of the RC 1st order battery model for discharging cycle (in MULTISIM 11 editor)

11 editor that captures accurately the electrochemical behavior of a battery in terms of the measured battery terminal voltage $y(t)$, charge or discharge current $u(t)$, OCV, battery internal resistance (R_{ch}, R_{dsch}), and SOC. Also this model is suitable for implementation in real-time the SOC sliding mode estimator. The both diagrams are the same, only the values of the battery internal resistance are different, according to charging (R_{ch}) or discharging cycle (R_{dsch}). The overall battery model can be described in discrete time by two states, first one is the voltage across the parallel group (R_1, C_1) of the components, and the second one is the battery SOC. The third equation is the output equation that relates the terminal battery voltage to the both states, and the input charging or discharging current.

The state and the input-state-output equations can be written as follows:

$$x_{1,k+1} = \left(1 - \frac{T_s}{R_{1C_1}}\right)x_{1,k} + \frac{T_s}{C_1}u_k \tag{1}$$

$$x_{2,k+1} = x_{2,k} - \left(\frac{\eta_i}{Q_{nom}}T_s\right)u_k \tag{2}$$

$$y_k = OCV\left(x_{2,k}\right) - x_{1,k} - R_i u_k \tag{3}$$

where

- T_s is the sampling time in seconds
- η_i is the columbic efficiency for charging and discharging cycle, $\eta_i = 0.98$ for charging cycle, and $\eta_i = 1$ for a discharging cycle
- Q_{nom} is the nominal capacity of the battery
- $x_{1,k} = V_1(k)$, $x_{2,k} = SOC(k)$
- k is the discrete time, e.g. $t_k = k \times T_s$
- $u_k = i(k)$ is the input current, if $i(k) < 0$ is a charging cycle, and if $i(k) > 0$ is a discharging cycle.
- R_i is the battery internal resistance, for a charging cycle $R_i = R_{ch}$, and for a discharging cycle $R_i = R_{dsch}$
- R_1C_1 is the polarization time constant of the parallel circuit $(R_{1,}C_1)$
- y_k is the battery terminal output voltage

The battery terminal voltage y_k may be predicted based on the battery SOC; the most accurate formulation is using a combination of three models Shepherd, Unnewehr universal, and Nernst models, as in [3, 5, 4] for $OCV\left(x_{2,k}\right)$:

$$OCV\left(x_{2,k}\right) = K_0 - \frac{K_1}{x_{2,k}} - K_2 x_{2,k} + K_3 \ln\left(x_{2,k}\right) + K_4 \ln\left(|1 - x_{2,k}|\right) \tag{4}$$

The model parameters $(K_0, K_1, K_2, K_3, K_4)$ are chosen to fit the model to the manufacture's data by using a least squares curve fitting identification method OCV = f(SOC), as shown in [2, 3, 5, 4], where the OCV curve is assumed to be the average of the charge and discharge curves taken at low currents rates from fully

charged to fully discharged battery. Using low charging and discharging currents can be minimized the cell dynamics. A simple offline (batch) processing method for parameters calculation can be carried out as in [3, 5, 4]. For a high power Li-Ion battery of 6 Ah and nominal voltage 3.6 V olts the rated capacity Q is Amps hours (1C) and its nominal capacity is defined as $Q_{nom} = 0.8 \times Q = 4.8$ Ah. The OCV charging and discharging curves related to the SOC are shown in Figs. 3 and 4, respectively. The values of the parameters that fit the model can be found in [4]. Based on these settings of the model parameters we will build in the next section the sliding mode observer estimator (SMOE) in order to estimate accurately the Li-Ion battery SOC, and extensive simulations will be carried out in a MATLAB/SIMULINK programming environment:

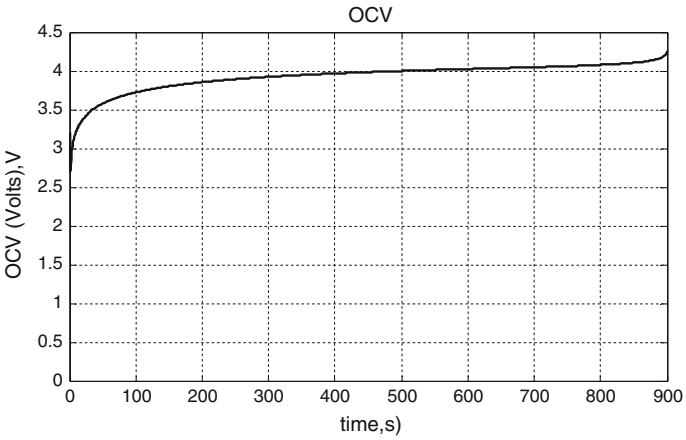


Fig. 3 The OCV charging curve at 5C rate (30 A constant current)

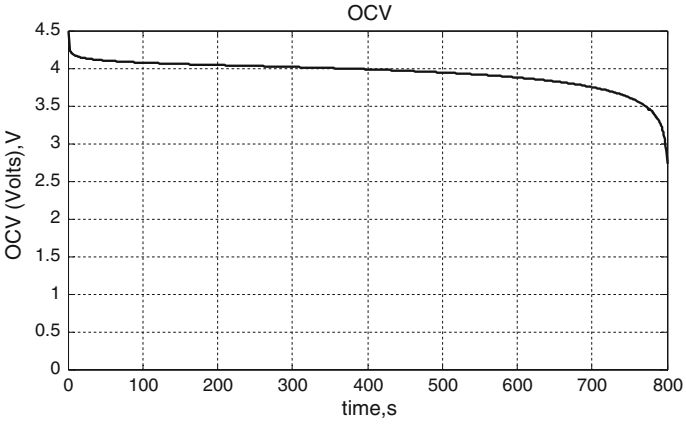


Fig. 4 The OCV discharging curve at 5C rate (30 A constant current)

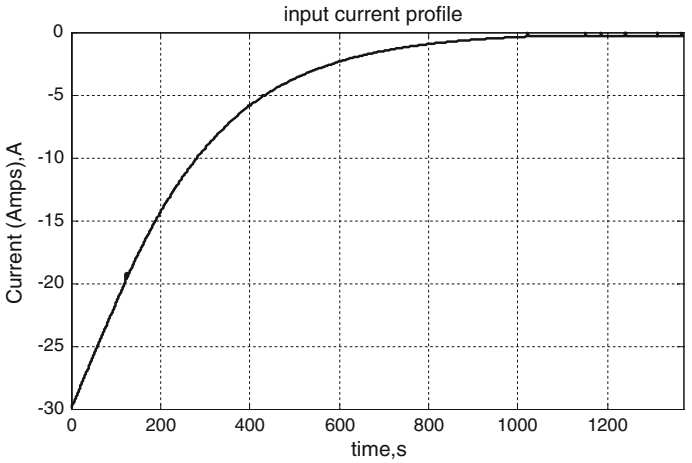


Fig. 5 Advisor EPA UDDS driving cycle current profile (Advisor-2 software package free download open-source)

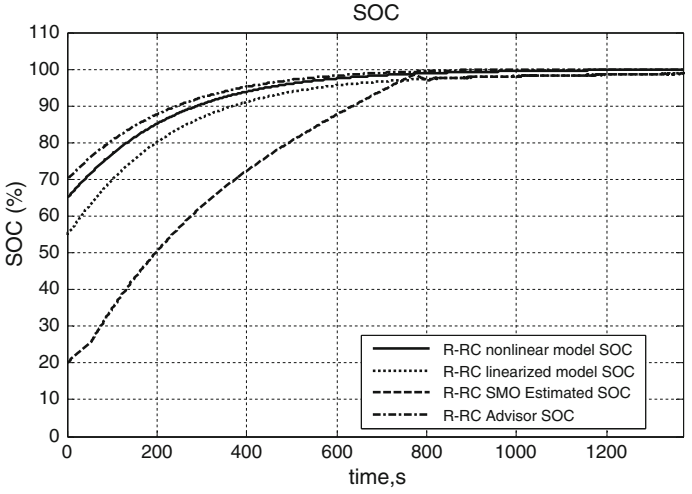


Fig. 6 Li-Ion 6Ah SAFT Battery—SOC curves for R-RC model (nonlinear, linearized, SMO estimator, and ADVISOR-2) for L = 1, M = 0.0001

$$K_0 = 4.23, K_1 = 3.8e-5, K_2 = 0.24, K_3 = 0.22, K_4 = -0.04, R_{ch} = 0.00182\,\Omega,$$
$$R_{dsch} = 0.00173\,\Omega, \mathrm{C}_1 = 23691.5\,\mathrm{F}, \mathrm{R}_1 = 0.00124\,\Omega.$$

Also in our simulations we set the sampling time T_s to be 1 s.

In order to analyze the behavior of the our Li-Ion battery model selection for different driving conditions such as urban, suburban and highway, some different

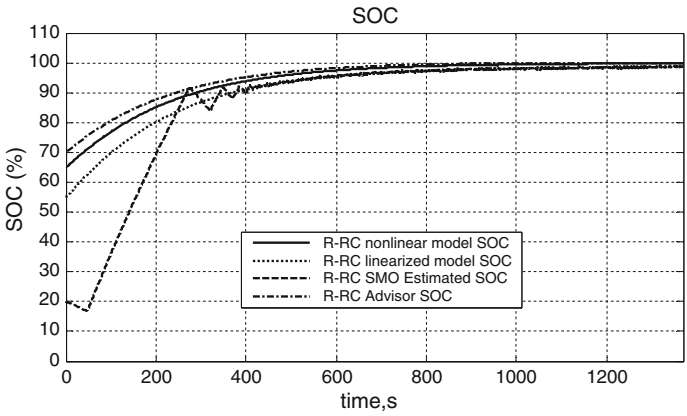


Fig. 7 Li-Ion 6Ah SAFT Battery—SOC curves for R-RC model (nonlinear, linearized, SMO estimator, and ADVISOR-2) for L = 5, M = 0.0001

current profile tests are used. Also for comparison purpose we use the results of the tests on a suburban small RWD' hybrid electric vehicle under standard initial conditions in an Advanced Vehicle Simulator (ADVISOR) environment, developed by US National Renewable Energy Laboratory (NREL), as in [1–3, 5, 4]. Among different driving cycle current profiles provided by the ADVISOR US Environmental Protection Agency we choose for our case study an Urban Dynamometer Driving Schedule (UDDS) current profile as is shown in Fig. 5. The corresponding SOC curves to this current profile are shown in Fig. 6 with some

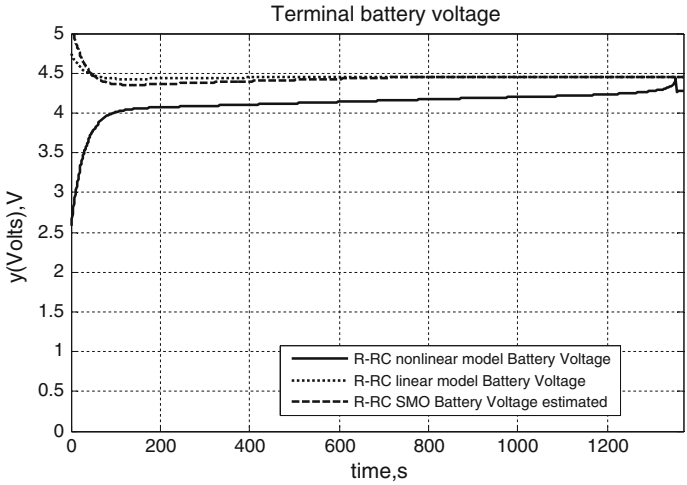


Fig. 8 The battery terminal voltage for UDDS current profile for L = 5, M = 0.0001

details given in the next section. Also the corresponding terminal voltages of the chosen Li-Ion battery that uses a charging UDDS current profile can be seen in the Figs. 7 and 8.

3 The Sliding Mode Observer Estimator

A combination of sliding mode methods and an observer provide the ability to generate a sliding motion on the error between the measured plant output and the observer output in order to ensure that a sliding mode observer (SMO) produces a set of state estimates precisely matching with the actual output of the plant [7]. Also the analysis of the average value of the applied observer injected signal, known as equivalent injection signal, contains valuable information about the disparity between the observer model and the actual plant [7]. In order to design a SMO estimator (SMOE) of Li-Ion SOC battery we follow the same design procedure used in [7]. The corresponding continuous dynamic description to the discrete time battery model given by (1)–(4) can be arranged in the following matrix state space representation:

$$\frac{dx}{dt} = A_{n\times n}x + B_{n\times m}u \tag{5}$$

$$y = C_{p\times n}x + D_{p\times n}u \tag{6}$$

where $n = 2$ represents the number of states, $m = 2$ is the number of inputs, and $p = 1$ is the number of outputs.

The state vector $x(t)$ is designated by $x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$, the input vector is $u(t) = \begin{bmatrix} i(t) \\ 1(t) \end{bmatrix}$ ($i(t)-$ input battery current profile, $1(t)-$ step input signal), and the battery output voltage $y(t)$ is related to SOC and OCV by a linear matrix equation. The matrices $C_{p\times n}$, $D_{p\times n}$ are obtained by linearizing the Eq. (4) around an operating point, assuming in our case study $SOC_0 = 0.6$. Following the procedure described in [7], the matrices triplet (A, B, C) is converted into canonical form (A_c, B_c, C_c), by using a nonsingular state transformation matrix

$$z = T_c x = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, z_1 \in R^{n-p}, z_2 \in R^p,\ T_c = \begin{bmatrix} N_c^T \\ C \end{bmatrix},\ N_c \in R^{n\times(n-p)} \tag{7}$$

where the column of the matrix N_C spans the null space of C [7], such that

$$A_c = T_c A T_c^{-1}, B_c = T_c B, C_c = C T_c^{-1} = [0 C_2] \tag{8}$$

A robust observer exists for the canonical battery model if $A_{c,11} < 0$ [7], so if it is stable. The sliding mode observer dynamics is attached to this canonical form and can be described by the following two equations [7]:

$$\frac{d\hat{z}_1}{dt} = A_{c,11}\hat{z}_1(t) + A_{c,12}\hat{z}_2(t) + B_{c,1}u(t) + L\vartheta \tag{9}$$

$$\frac{d\hat{z}_2}{dt} = A_{c,21}\hat{z}_1(t) + A_{c,22}\hat{z}_2(t) + B_{c,2}u(t) - \vartheta \tag{10}$$

where $\hat{z}_1(t), \hat{z}_2(t)$ represent the estimates of the battery model states in the canonical form, and L is the linear observer matrix gain. The value of L is given by imposing the spectrum of the matrix $A_{c,11} + LA_{c,21}$ to lie in C_- [7], i.e.

$$A_{c,11} + LA_{c,21} < 0 \tag{11}$$

The vector ϑ can be seen as a nonlinear observer vector switching gain given by:

$$\vartheta_i = Msgn\left(\hat{z}_{2,i} - z_{2,i}\right), M \in R_+, i = 1, \ldots, p \tag{12}$$

where the gain coefficient M is a very useful sliding mode observer tuning parameter needed to increase the battery SOC estimate accuracy. Similar, the linear gain L, is a second tuning SMOE parameter that offers a new freedom degree for SMOE to increase the estimate accuracy and to decrease drastically the detection time of possible faults that could occur inside the battery cells. Extensive MATLAB simulations will be done in the next two subsections with the aim to prove the effectiveness of the proposed sliding mode observer estimator in real-time implementation of Li-Ion battery SOC estimation and fault detection.

3.1 *Sliding Mode Observer Estimator—SOC Estimation Simulation Results*

In this case study we present the simulation results for a high power Li-Ion cell battery 6Ah and 3.6 V nominal voltage developed by SAFT America for small suburban HEVs. This company is part of the U.S. Advanced Battery Consortium, as U.S. Partnership for New Generation of Vehicles programs [2]. Based on the experimental test data, the National Renewable Energy Laboratory (NREL) developed a MATLAB resistive equivalent circuit battery model in order to be compared to a SAFT 2-capacitance battery [2]. For purpose comparison of the sliding mode predictions was used the ADVISOR simulator for different operating conditions provided by two driving cycles tests, US06 and EPA UDDS [2]. Inspired by the usefulness of the ADVISOR simulator and the results obtained in this area we prove the effectiveness of our proposed estimator design strategy using an

EPA UDDS driving cycle current profile integrated in Advisor 2 software package (open-source free download since 2014). This driving cycle current profile is used as a charging current test for the chosen R-RC first order battery model. Then the performance of the battery model and sliding mode observer estimator can be compared to those obtained by the ADVISOR simulator for different Li-Ion HEVs batteries with similar characteristics. The ADVISOR EPA UDDS driving cycle current profile is shown in Fig. 5 and it seems to be a smooth exponential charging current cycle. The continuous battery model dynamics is described in a state-space representation by the matrices quadruplet (A, B, C, D) that appear in the state Eqs. (5)–(6) given by:

$$A = \begin{bmatrix} -\frac{1}{R_1 C_1} & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} \frac{1}{C_1} & 0 \\ -\eta/Q_{nom} & 0 \end{bmatrix}, C = \begin{bmatrix} -1 \\ C_2 \end{bmatrix}, D = [-R_i \; K_0] \tag{13}$$

$$T_c = \begin{bmatrix} C_2 & 0 \\ 1 & 1 \end{bmatrix}, C_2 = K_1/x_0^2 - K_2 + K_3/x_0 - K_4/(1 - x_0), x_0 = 0.6 \tag{14}$$

$$K_0 = 4.23, K_1 = 3.8e - 5, K_2 = 0.24, K_3 = 0.22, K_4 = -0.04 \tag{15}$$

$$R_{ch} = 0.00182\,\Omega, R_{dsch} = 0.00173\,\Omega, \mathrm{C}_1 = 23691.5\,\mathrm{F}, \mathrm{R}_1 = 0.00124\,\Omega \tag{16}$$

The Li-Ion battery SOC evolution curves during the charging cycle are shown in Fig. 6, for first order R-RC nonlinear model, for its linearized dynamics and SMO estimator with the gains settings L = 1, M = 0.0001; also for comparison purpose is shown on the same graph the SOC obtained by the ADVISOR simulator for the same driving current profile. In Fig. 7 we represent the same curves for observer gains tuned to L = 5, M = 0.0001 that reveal a fast estimation with some small

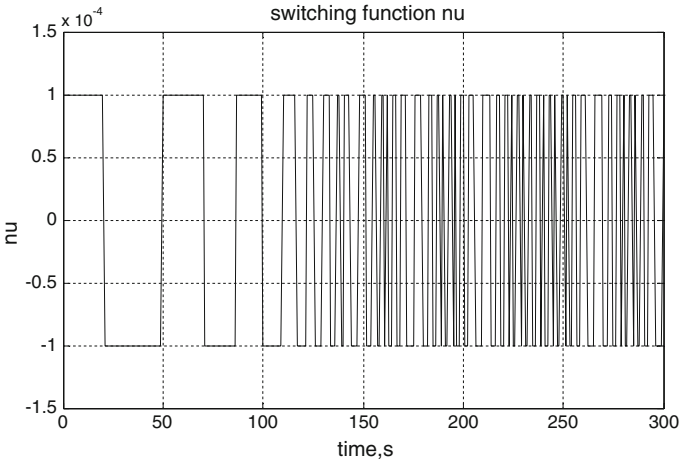


Fig. 9 Switching nonlinear observer gain for L = 1, M = 0.0001

oscillations around the SOC steady-state value, and sliding mode observer robustness to a big change in the initial SOC value as a first guess starting point value ($SOC_0 = 0.2$ instead to $SOC_0 = 0.7$ as is used in ADVISOR). In Fig. 8 is shown the terminal battery voltages for nonlinear battery model, for its linearized model, and for SMO estimator, when the observer gains are set to L = 5, M = 0.0001. Again we can see an accurate estimation of the terminal battery voltage when SMO estimator is used. The evolution of switching nonlinear observer gain is shown in Fig. 9.

3.2 *Sliding Mode Observer Estimator—Battery Cell Current Fault Detection Simulation Results*

The sliding mode observer estimator developed in this research can be also used for fault detection purpose. To prove its effectiveness in fault detection we consider the particular case of an intermittent fault injection in the battery cells current profile, as is shown in Fig. 10. A zero faulty current that simulates an open cell circuit in a battery pack is injected after 100 s and it will be removed after 100 samples. So the window length of the persistent injected fault is only 100 s.

The sliding mode observer estimator reacts very fast to this current. The current fault is detected in some few seconds, as is shown in Fig. 11. In this figure is shown the terminal battery voltage residual generated by the SMOE.

In the future work we will develop some applications of the proposed SMOE to detect and estimate the severity of the faults, by augmenting the model dynamics state-space with a new dimension given by the following fault parameter equation [5]:

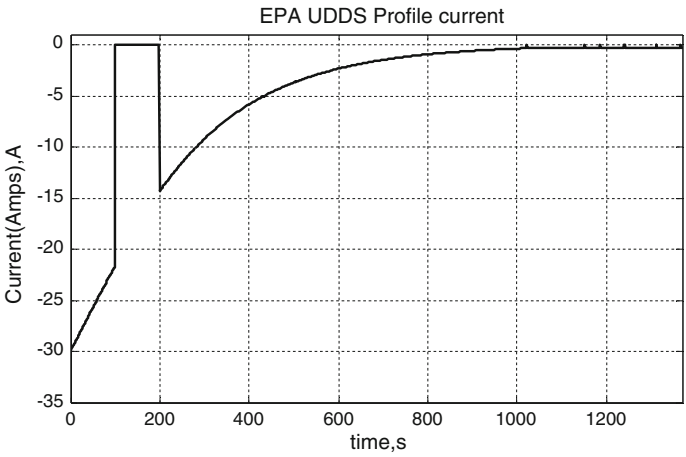


Fig. 10 Intermittent current fault in SAFT battery cell injected between iterations 100 and 200

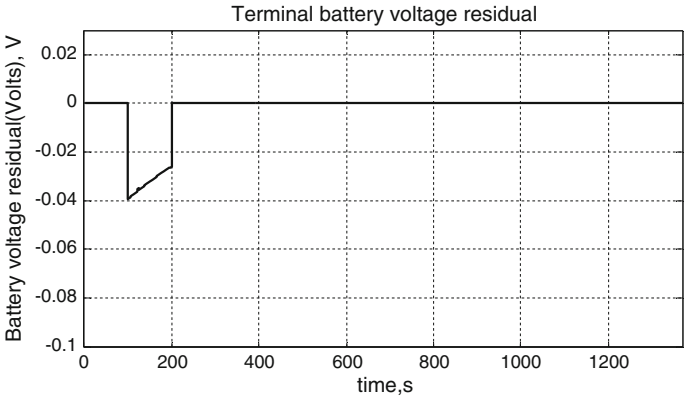


Fig. 11 SMO Fault detection (SAFT battery cell voltage residual)

$$\Theta_{k+1} = \Theta_k + \nu_k \tag{17}$$

where Θ_k represent discrete time fault parameter description, and ν_k is a small level noise. Now the SMO estimator has three states and the procedure to estimate the states will be the same to those used for two states, as in [7]. The big challenge in this development is the capability of the augmented estimator to estimate the severity of the faults.

4 Conclusion

The novelty of this paper is the implementation in real time of a robust Sliding Mode Observer estimator capable to estimate with high accuracy the Li-Ion battery SOC based on a linearised model without disturbance uncertainties, and also to detect the faults inside the battery cells. The proposed SMO estimation strategy seems to be simple, easy to be implemented in real-time, and of lower computations complexity, compared to Kalman filter estimation techniques. The proposed SMO estimator will be very useful in our research to solve problems of fault detection and isolation (FDI) that could appear in the BMS of HEVs and EVs.

References

1. Young, K., Wang, C. Le Yi Wang, L.Y., Strunz, K.: Electric vehicle battery technologies—chapter 12. In: Garcia-Valle, R., Peças Lopez, J. A. (eds.) Electric Vehicle Integration into Modern Power Networks 2013, vol. IX, pp. 16–56. Springer, Berlin (2013)
2. Johnson, V.H.: Battery performance models in ADVISOR. J. Power Sources **110**, 321–329 (2001)

3. Plett, G.L.: Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs—part 2. Modeling and identification. J. Power Sources **134**(2), 262–276 (2004)
4. Farag, M.: Lithium-Ion batteries. In: Modeling and State of Charge Estimation, Master Thesis, McMaster University, Hamilton, Ontario, Canada, 169 (2013)
5. Plett, G.L.: Extended Kalman filtering for battery management systems of LiPB-Based HEV battery packs—part 3. State and parameter estimation. J. Power Sources **134**(2), 277–292 (2004)
6. Tudoroiu, N., Khorasani, K.: Satellite fault diagnosis using a bank of interacting Kalman filters. J. IEEE Trans. Aerosp. Electron. Syst. **43**(4), 1334–1350 (2007)
7. Eduards, C., Spurgeon, S.K., Patton, R.J.: Sliding mode observers for fault detection and isolation. Pergamon, Automatica **36**, 541–553 (2000)

Recursive Identification of Nonlinear Aerodynamic Characteristics of Air-Plane

Jacek Pieniążek and Piotr Cieciński

Abstract The subject of the paper concerns the method of on-line identification, especially considering the form of equation of air-plane motion. The main idea is using the table representation of the function. The proposed approximation of the data obtained during the experiment is the basis of the recursive algorithm, which can find the best approximation in the area of excitation. The test results shows the importance of proper design of the stimulation signals. This signal should be sum of low-frequency set-point component and high-frequency part. For the purpose of presentation identification of the c_m aerodynamic coefficient of an air-plane is presented.

Keywords Recursive identification · Air-plane aerodynamic · Non-linear identification

1 Introduction

During the design process both the airframe and the automatic system which control the behavior of the flying vehicle the proper dynamical model of this vehicle is required. The creation of the perfect model of the motion of the flying vehicle is almost impossible because of the lack of good description of aerodynamic phenomena and cross dependencies among the various state variables. During the process of the modelling of the flying object motion the various methods are used, such as handbook methods, computational fluid dynamics methods, research in the wind tunnel and in-flight tests. Computational methods enable the anticipation of behavior of the object during the design process also before the construction of the vehicle. Final test of the model of the object motion are the real structure research in the wind tunnel or in-flight. The best results can be achieved during the wind tunnel

J. Pieniążek (✉) · P. Cieciński
Rzeszów University of Technology, Rzeszów, Poland
e-mail: jp@prz.edu.pl

P. Cieciński
e-mail: pciecins@prz.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_3

tests of the real size vehicle or its model, but the high cost of these tests is the reason of the research conducted on the models made in smaller scale. The necessity of the transformations of the results of experiment conducted in the scale to the full size leads to committing of the additional errors, which does not enable of usage every advantages of research in the wind tunnel [1].

2 Air-Plane Mathematical Model

An air-plane rigid body equations of motion have general form given by (1) with gravity forces (2) and aerodynamic phenomena model (3) [1].

$$\begin{aligned}
X_a + X_g &= m\left(\overset{\circ}{u} + qw - rv\right)\\
Y_a + Y_g &= m\left(\overset{\circ}{v} + ru - pw\right)\\
Z_a + Z_g &= m\left(\overset{\circ}{w} + pv - qu\right)\\
L &= \overset{\circ}{p} I_{xx} + qr(I_{zz} - I_{yy}) + (pq + \overset{\circ}{r})I_{xz}\\
M &= \overset{\circ}{q} I_{yy} + rp(I_{xx} - I_{zz}) + (p^2 - r^2)I_{xz}\\
N &= \overset{\circ}{r} I_{zz} + pq(I_{yy} - I_{xx}) + (qr - \overset{\circ}{p})I_{xz}
\end{aligned} \tag{1}$$

$$\begin{aligned}
X_g &= -mg * \sin\theta\\
Y_g &= mg * \sin\varphi\cos\theta\\
Z_g &= mg * \cos\varphi\cos\theta
\end{aligned}$$

$$\begin{aligned}
X_a &= \left(C_{X_0} + C_{X_\alpha}\alpha + C_{X_{\overset{\circ}{\alpha}}}\frac{\overset{\circ}{\alpha}\bar{c}}{V} + C_{X_q}\frac{q\bar{c}}{V} + C_{X_{\delta e}}\delta_e\right)1/2\,\rho\,V^2S\\
Z_a &= \left(C_{Z_0} + C_{Z_\alpha}\alpha + C_{Z_{\overset{\circ}{\alpha}}}\frac{\overset{\circ}{\alpha}\bar{c}}{V} + C_{Z_q}\frac{q\bar{c}}{V} + C_{Z_{\delta e}}\delta_e\right)1/2\,\rho\,V^2S\\
M &= (C_{m_0} + C_{m_\alpha}\alpha + C_{m_{\overset{\circ}{\alpha}}} + \frac{\overset{\circ}{\alpha}\bar{c}}{V} + C_{m_q}\frac{q\bar{c}}{V} + C_{m_{\delta e}}\delta_e)1/2\,\rho\,V^2S
\end{aligned} \tag{2}$$

$$\begin{aligned}
Y_a &= \left(C_{Y_0} + C_{Y_\beta}\beta + C_{Y_{\overset{\circ}{\beta}}}\frac{\overset{\circ}{\beta}b}{2V} + C_{Y_p}\frac{pb}{2V} + C_{Y_r}\frac{rb}{2V} + C_{Y_{\delta a}}\delta a + C_{Y_{\delta r}}\delta_r\right)1/2\,\rho\,V^2S\\
L &= \left(C_{l_0} + C_{l_\beta}\beta + C_{l_{\overset{\circ}{\beta}}}\frac{\overset{\circ}{\beta}b}{2V} + C_{l_p}\frac{pb}{2V} + C_{l_r}\frac{rb}{2V} + C_{l_{\delta a}}\delta a + C_{l_{\delta r}}\delta_r\right)1/2\,\rho\,V^2Sb\\
N &= (C_{n_0} + C_{n_\beta}\beta + C_{n_{\overset{\circ}{\beta}}}\frac{\overset{\circ}{\beta}b}{2V} + C_{n_p}\frac{pb}{2V} + C_{n_r}\frac{rb}{2V} + C_{n_{\delta a}}\delta a + C_{n_{\delta r}}\delta_r)1/2\,\rho\,V^2Sb
\end{aligned} \tag{3}$$

Parameters of the particular air-plane can be find differently. Some of them can be find by simple or indirect measurement like the geometric ones (S, $\bar{c}$, b), mass (m), location of the centre of mass, inertial moments (I_{xx}, I_{yy},…). The more difficult but more interesting for researcher is investigation of the aerodynamic coefficients which in the Eq. (3) as C_x. These coefficients are the functions of such variables [1–5]:

- angle-of-attack α and angle of side-slip β, which are general parameters of the airflow around the body,
- control surfaces deflections (δ_e—elevator, δ_a—aileron, δ_f—flaps, δ_r—rudder…)
- propulsion dependently on its kind and its location,
- and also external conditions like:
- air velocity around the body (important for the air-planes flying in wide range of velocities), and lightly the air composition for example humidity,
- nearest outer objects and especially surface of the Earth if it is close enough.

The angle-of-attack and angle of side-slip and intentional deformations of the geometry of the body by the control surfaces deflections in the given flight conditions create functions $C_m(\alpha, \delta_e, \delta_f)$, $C_z(\alpha, \delta_e, \delta_f)$, $C_x(\alpha, \delta_e, \delta_f)$, $C_l(\alpha, \beta, \delta_a, \delta_r)$, $C_n(\alpha, \beta, \delta_a, \delta_r)$, $C_y(\alpha, \beta, \delta_a, \delta_r)$, hence the aerodynamic properties of the air-plane.

Another variables of the state like angle rates neglected above but shown in Eq. (3) are the measure of the unsteadiness of the airflow. Such unsteadiness is difficult to describe by equations and it is lightly predictable by theory but has also influence on the value of the coefficients. The non-linearities in the functions $C(\ldots)$ makes the relations more complicated comparing to motion the simply-looking Eqs. (1)–(3).

In many practical design and also in research activities the models are simplified to the linear form of matrix state and output equations [1, 3–5]. Such form is used during the controller design. In such a case the matrix coefficients are the target of the in-flight investigation of the aircraft model. The limitations of the linear modelling, really it is only model of the local dynamics near set-point, makes it unsuitable for the automatic control system design which should safely control the flight of the air-plane in the wide range of the state.

From this reason the tests are taken up to find out the most accurate as possible aero-dynamical characteristics of the motion of flying vehicles. The knowledge about real values of the parameters which describe the object is also required during both the advanced control method usage and current analysis of the state of the object.

3 Review of Estimation Methods

Depending on the modelling purpose and accepted structure of the model for estimation of the model coefficients the off-line estimation is possible to be used. The linear estimators are described with details and they are relatively the simplest

for usage [3, 4, 6–8], but because of limited range of its representation of the real motion of flying vehicle the non-linear estimators are much interesting. The example of the linear estimation use for determination of the non-linear characteristics of aero-dynamical coefficients of the airplane is shown in [9, 10].

The second group of the estimators are that operating on-line. These estimators give possibility to achieve estimation of the model coefficients during the test. In a case of the experiments conducted in-flight this fact is very favorable. The correctness of the results of these experiments depend on both proper plan of the experiment and environment mostly atmospheric conditions during the test. For that reason it is worth to know the results of experiment to correct the plan. The activity described above is not possible in a case of off-line estimators. In a case of linear model in the literature there are described some methods of estimation of the coefficients in the time of experiment (on-line methods) which are possible for usage, like recursive least squares, recursive weighted least squares, recursive prediction error method, Fourier transform regression, filtering recursive methods [1, 4, 7, 11–14].

In the paper the method of estimation of the coefficients is presented, which describes the characteristics of the aero-dynamical coefficients with the recursive weighted least squares method usage. The characteristics of the aero-dynamical coefficients are described by the values of the coefficients at points lying in the intersections of the grid which is given by ordered sets of values of the appropriate vehicle state variables and control signals. In the Fig. 1 there presented the grids described by the input values $x_1(k_1)$ and $x_2(k_2)$ and the exemplary values of output y at the nodes. For the purpose of both modeling of the airplane motion and estimation of using node values $y(k_1, k_2)$ the coefficients the interpolation of the value of the particular coefficient y is computed. During the activity of the estimator

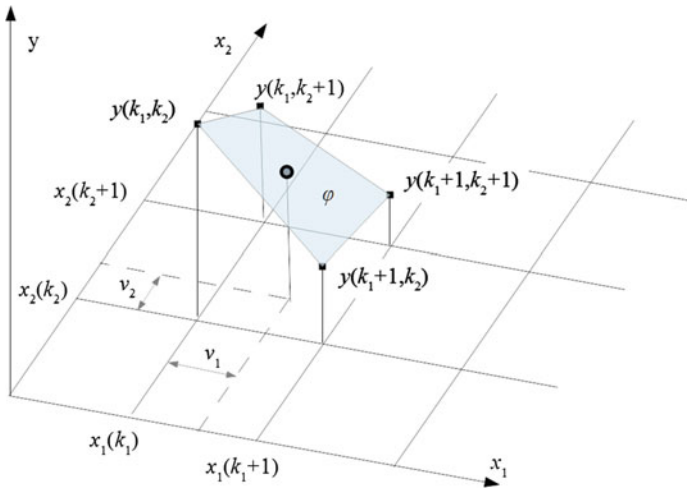


Fig. 1 Example of multivariable aerodynamic characteristic using grid representations

in every moment the linear estimation of the coefficients is conducted. These coefficients describes the shape of the section and the general shape of estimate function can be a fuzzy system.

3.1 Identification of Non-linear Functions of Aerodynamic Coefficients

The information set necessary for aerodynamic characteristics reconstruction can be obtained by using various method, like analytical and computer modelling of the airflow and computation of the coefficients and also experimental research. The second can be realized by reduced scale models and wind tunnel experiments. Next the results are transformed to the full scale conditions but the errors appears due the non-perfect airflow similarity and also rough shape reconstruction of the scaled models. The full scale models or real air-plane investigation in wind tunnels may reduce the errors but it is technically more difficult or simply it is impossible because of the size of a body.

In-flight experiments with real air-plane looks as the best method but the constraints, like the limitations of accessible values of variable and generally more difficult measure conditions, and possibly existing disturbances are important factors which should be considered.

The identification is process consist of:

- design of the general method, planning of the experiment, selecting data representation and choosing appropriate computation algorithm,
- conduct experiment and obtaining data,
- estimation of parameters
  - on-line if computation is executed during the experiment,
  - off-line if the recorded set of data is used,
- verification of the results.

In the following analysis it is assumed that, the control signals for the test of the air-plane are developed according to the identification experiment. It means that the steady flight is disturbed only by the control surfaces which influence the motion of the investigated coefficient. As the result the information necessary for estimation is in the variables:

- a_x, a_z—for C_x and C_z
- a_y—for C_y
- $\overset{\circ}{q}$—for C_m
- $\overset{\circ}{p}$—for C_l
- $\overset{\circ}{r}$—for C_n

The general form of the aerodynamic coefficients is known as some functions, which are using linear as stated in Eq. (3) but it should be noticed these equations are only rough approximation of reality when the non-linearity, very important when the angle-of-attack is high, and the influence of the airflow condition on the controls effectiveness are neglected.

The general form of function description, which is very useful in computer simulations, is a table. In the case of aerodynamic coefficient investigation the node values can be computed analytically and directly put into the table as initial values. Some values can be identified using the method presented in [9] and put into the table or correct the analytical method. The table form is also useful in an on-line identification and has a property to transform non-linear approximation into the linear in the parameters problem as will be stated further.

3.2 Function Representation in Table

A function $y = f(x^1, x^2, \ldots x^p)$ in the form of the table consists of the:

- one-dimensional tables of the values of input variables, $X^K = \{x_1^k, x_2^k, \ldots x_{nk}^k\}$, number of tables equals p,
- the p-dimensional table of the output values.

The values of output in the node it is for $x^k = x_i^k$ for every $k = 1 \ldots p$, are given in the table.

The general output from the table is computed using Eq. (4)

$$y = y(k_1, k_2, \ldots k_p) + \varphi(v_1, v_2, \ldots v_p) \tag{4}$$

where

$k_i = \min_{j \in \{1,2,\ldots p\}} (k_j)$ such that, fulfil condition $x^i - x_{k_j}^i > 0$,

$$v_i = 1 - \frac{x^i - x_{k_i}^i}{x_{k_i+1}^i - x_{k_i}^i} \tag{5}$$

and $\varphi(v_1, v_2, \ldots v_p)$ is an approximation in the p-dimensional cuboid determined by the pairs of input signal values $x_{k_i}^i, x_{k_i+1}^i$, (in the Fig. 1 this cuboid reduces to the rectangle) which should fulfil condition

$$\varphi(v_1, \ldots v_p) = y(m_1, m_2, \ldots m_p), n_i = \begin{cases} k_i, \text{ if } v_i = 0 \\ k_i + 1, \text{ if } v_i = 1 \end{cases} \tag{6}$$

The table can be designed by using the values of the output obtained by any method. This make possible to use the results of theoretical research as the initial values. The results from identification experiment correct them. Such possibility of

initialize by near real values can be important because of the limited range of achievable values of control signals and also safety limited range of the state variables in read flight. The resulting set of input data points achievable during flight is only subset of the set $\langle x_1^1, x_{n_1}^1 \rangle \times \langle x_1^2, x_{n_2}^2 \rangle \times \cdots \times \langle x_1^p, x_{n_p}^p \rangle$ as it is presented in part 4.

The minimal interpolation in the p-dimensional cuboid is denoted by Eq. (7) and is used as approximation of the data inside them.

$$\varphi(v_1, v_2, \ldots, v_p) = a_{0(k_1,k_2)} + \sum_{k=1}^{p} a_{k(k_1,k_2)} \cdot v_k + \sum_{k=1}^{p} \sum_{l \neq k} a_{kl(k_1,k_2)} \cdot v_k \cdot v_l + \sum_{k=1}^{p} \sum_{l \neq k} \sum_{m \neq k, m \neq l} a_{klm(k_1,k_2)} \cdot v_k \cdot v_l \cdot v_m + \cdots \tag{7}$$

3.3 Identification of Aerodynamic Coefficient

The aerodynamic coefficients are computed using the Eq. (8) obtained from (1) and (3).

$$\begin{aligned} c_\zeta(\alpha, \delta \ldots) &= \frac{1}{p_d \cdot S \cdot l} \cdot I_\zeta \cdot \dot{\Omega}_\zeta \\ c_\xi(\alpha, \delta \ldots) &= \frac{1}{p_d \cdot S} \cdot m \cdot a_\xi \end{aligned} \tag{8}$$

$$\text{where} \quad \begin{aligned} c_\zeta &\in \{c_l, c_m, c_n\} \\ I_\zeta &\in \{I_{xx}, I_{yy}, I_{zz}\} \\ \dot{\Omega}_\zeta &\in \{\dot{p}, \dot{q}, \dot{r}\} \\ c_\xi &\in \{c_D, c_y, c_L\} \\ a_\xi &\in \{a_x, a_y, a_z\} \end{aligned}$$

It is important that the components of linear acceleration and of the frame rate are presented in the wind frame, and identified coefficients are drag C_D and lift C_L.

For the purpose of presentation more detailed analysis of the coefficient C_m is presented. This coefficient depends mainly on the angle of attack and deflections of the elevator and also deflection of flaps. Considering that the flaps are used as three or four position control surface such number of tables is the assumed result of full process of identification. Then every table depends only on two variables. The value of actual C_m is computed using (9) and approximation of angular acceleration (10).

$$C_m(\alpha, \delta_H) = \frac{1}{p_d \cdot S \cdot l} \cdot I_{yy} \cdot \dot{q} \tag{9}$$

$$(\dot{q})_k = \frac{q_{k+1} - q_k}{\tau} \tag{10}$$

The approximation is used instead of non-existing direct measurement. The usage of the (10) is a cause of increase of the noise level in the angular acceleration by the quantization noise and also because of amplification of the gyro noise.

The data measured and computed during the experimental flight are the teaching set (11).

$$S = \{(\alpha, \delta_H, c_m)_k\}, k = 1, \ldots N \tag{11}$$

The representation of the function $C_m(\alpha, \delta_e)$ between the nodes is obtained from the general form (7) is given by (12).

$$\varphi(v_1, v_2) = a_{0(k_1,k_2)} + a_{1(k_1,k_2)} \cdot v_1 + a_{2(k_1,k_2)} \cdot v_2 + a_{12(k_1,k_2)} \cdot v_1 \cdot v_2 \tag{12}$$

The coefficients in Eq. (12) and values in the data table depends according to the Eq. (13)

$$\begin{aligned} a_{0(k_1,k_2)} &= y(k_1, k_2) \\ a_{1(k_1,k_2)} &= y(k_1 + 1, k_2) - y(k_1, k_2) \\ a_{2(k_1,k_2)} &= y(k_1, k_2 + 1) - y(k_1, k_2) \\ a_{12(k_1,k_2)} &= y(k_1, k_2) + y(k_1 + 1, k_2 + 1) - y(k_1, k_2 + 1) - y(k_1, k_2 + 1) \end{aligned} \tag{13}$$

The dependence (12) can be stated in the form (14) as the basis for linear least-square problem.

$$\begin{aligned} y &= X^T \cdot \Theta \\ X &= [1\ v_1\ v_2\ v_1 \cdot v_2]^T \\ \Theta &= [a_{0(k_1,k_2)} a_{1(k_1,k_2)} a_{2(k_1,k_2)} a_{12(k_1,k_2)}]^T \end{aligned} \tag{14}$$

3.4 Recursive Solution—On-Line Identification

Let assume there exists the proper parameters $\Theta(k_1,k_2)$ in every rectangle of input variables. The estimation error is given by (15). Taking the values of X as the weights of error the recursive algorithm is given by dependence (16).

$$e = y - \hat{y} = X^T_{(k_1,k_2)} \cdot \left(\Theta_{(k_1,k_2)} - \hat{\Theta}_{(k_1,k_2)}\right) \tag{15}$$

$$\hat{\Theta}_{(k_1,k_2)}(k+1) = \hat{\Theta}_{(k_1,k_2)}(k) + \gamma \cdot X_{(k_1,k_2)}(k) \cdot e_k = \hat{\Theta}_{(k_1,k_2)}(k) + \Delta\Theta(k) \tag{16}$$

The change of parameters of Eq. (12) influence the node values in the output table as stated in the Eq. (17).

$$\begin{bmatrix} y_{k_1,k_2}(k+1) \\ y_{k_1+1,k_2}(k+1) \\ y_{k_1,k_2+1}(k+1) \\ y_{k_1+1,k_2+1}(k+1) \end{bmatrix} = \begin{bmatrix} y_{k_1,k_2}(k) \\ y_{k_1+1,k_2}(k) \\ y_{k_1,k_2+1}(k) \\ y_{k_1+1,k_2+1}(k) \end{bmatrix} + \begin{bmatrix} \Delta a_0 \\ \Delta a_0 + \Delta a_1 \\ \Delta a_0 + \Delta a_2 \\ \Delta a_0 + \Delta a_{12} - \Delta a_1 - \Delta a_2 \end{bmatrix} \tag{17}$$

The values computed by (17) if e is given by (15) converges to the proper parameters so the value of the learning factor γ can be set near 1. But the real values of the C_m are disturbed by the cumulative measurement noise and also by the disturbance which is the effect of numerical differentiation (10). Cumulative disturbance can be treated as the noise with expected value 0 if the experiment is long enough and the stimulation trajectories are dense in the input area.

The remaining noise in the response of (16) and (17) decrease with reduction of the learning coefficient. Such property can be used in such a way, that the value of γ decrease with time of experiment, because it is expected that, the estimated values are closer to the proper values.

4 Experiment Plan and Stimulation

Contrary to the wind tunnel research during the in-flight experiment it is impossible to independently change the values of the state variables and control variables. The proper stimulation is achieved by using control signals which contains two components. The first is the low frequency one, which change the set-point. The second is high frequency but more important property of them is activation in the widest region of the inputs.

As the stimulus the function given by Eq. (18).

$$\delta_e = LPF(A_1 \cdot 1(t + k_1 \cdot \tau_1) + A_2 \cdot 1(t + k_2 \cdot \tau_2) + s(t) \cdot [1(t + k_3 \cdot \tau_3) - 1(t + (k_3 + 1) \cdot \tau_3)]) \tag{18}$$

The function (18) contains two pseudo-random components (random amplitudes A_1 and A_2 and fixed durations τ_1 and τ_2) and signal s of the shape of two short opposite pulses repeating with period τ_3. The LPF (low-pass filter) operator is used because it is impossible in the reality to create the step responses of control surface deflections and its output is more realistic. The resulting control variable values with time are shown (short part of experiment time) in the Fig. 2.

In the Fig. 3 it is presented the function of C_m, which is used in the simulation as real characteristic of the modelled air-plane motion. The points which are achieved during the experiments are shown on this surface. The two sets distinguished by

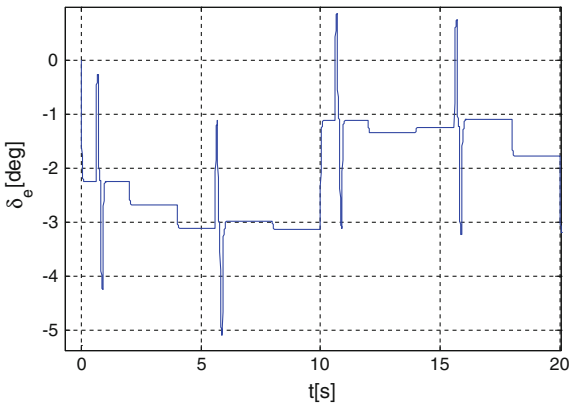


Fig. 2 Function of control variable d_e

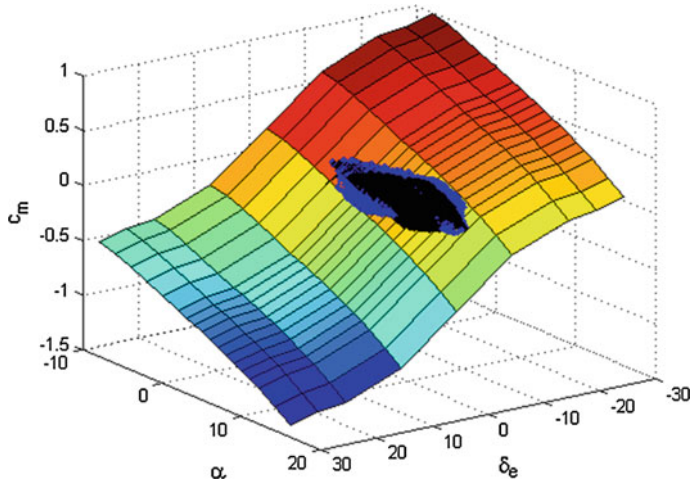


Fig. 3 Input variables covering on the c_m characteristic, *black*—without *s*, *blue*—with *s*

colours gives a view of the increase of the stimulated area when component *s* is used as given in Eq. (18)—blue points, comparing to the case of lack of them—black points.

5 Experiment Results

The simulation experiment has been developed for the purpose of evaluation the possibility of on-line identification of the aerodynamic coefficients. It is assumed the initial node values are computed theoretically but this function differ from the real air-plane characteristic. The experiment time was set constant as 600 s.

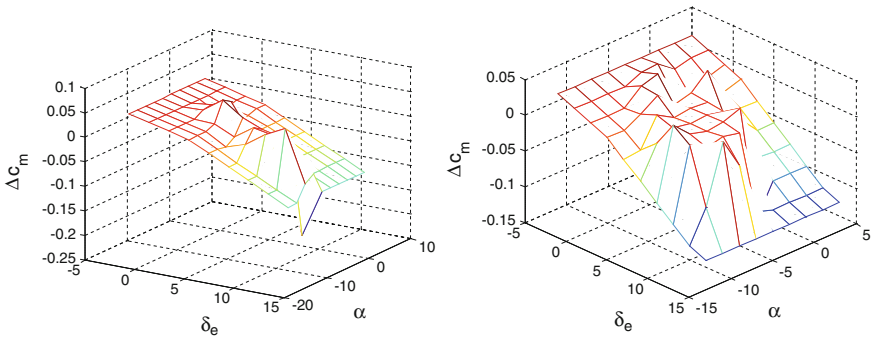


Fig. 4 The remaining error of the c_m estimation

The control sequence given by (18) has property that the air-plane does not void the airspeed limits. During simulation the improvement due to using the component s is investigated. Two test was conducted, one without the component *s* and second one with it. The remaining errors are presented in the Fig. 4. The very narrow area where the initial error is reduced in the Fig. 4 (left) comparing to the more better estimation when the pulses increase excitation (Fig. 4 right) confirms necessity of using them.

6 Conclusion

The novel method the on-line identification of non-liner function as the part of dynamical model of motion is presented in the paper. The performed investigation shows the possibility to identify the non-linear function in the form of tabular data but also demonstrate the limitations of the real-object and real-conditions experiments. The dynamics of an object and the necessity of the safety limiting of the state variables reduces the area of excitation in the normal flight. The special kind of the excitation can be used but unfortunately it should be noticed that the added dynamical component increase the rather unpleasant responses of an air-plane.

The table representation of a function makes possible to combine the results of different methods as the theoretical analysis and the identification. The theoretical results can be used as the initial values and also as final values outside area of excitation. The inaccuracies on the border of excitation should be reduced in the further work, and at this moment assuming that, the coefficient characteristic should be rather smooth, the development of the identification algorithm by using some smoother is possible.

The simulation testing and possibly reshaping of the stimulation signals looks as important part of identification process.

References

1. Laban, M.: On-line Aircraft Aerodynamic Model Identification. Proefschrift Technische Universiteit, Delft (1994)
2. Ignatyev, D.I., Khrabrov, A.N.: Neural network modeling of unsteady aerodynamic characteristics at high angles of attack. Aerosp. Sci. Technol. **41**, 106–115 (2015)
3. Klein, V., Morelli, E.A.: Aircraft System Identification. AIAA Inc., Reston Virginia (2006)
4. Jategaonkar, R.V.: Flight Vehicle System Identification: A Time Domain Methodology. AIAA Inc., Reston Virginia (2006)
5. Etkin, B.: Dynamics of Atmospheric Flight. Wiley, London (1972)
6. Zhu, Y.: Multivariable System Identification for Process Control. Elsevier Science, Oxford (2001)
7. Soderstrom, T., Stoica, P.: System Identification. Prentice Hall International Ltd, UK (1994)
8. Ljung, L.: Perspectives on system identification. Ann. Rev. Control **34**, 1–12 (2010)
9. Cieciński, P., Pieniążek, J.: Identification of a UAV model for control. Lect. Notes Control Inf. Sci. **396**, 13–22 (2009)
10. Parameter Estimation Techniques and Applications in Aircraft Flight Testing. NASA Technical Note, Flight Research Center Edwards, Calif (1973)
11. Chowdhary, G., Jategaonkar, R.: Aerodynamic parameter estimation from flight data applying extended and unscented kalman filter. Aerosp. Sci. Technol. **14**, 106–117 (2010)
12. Leou, Maw-Lin, Liaw, Yi-Ching, Chien-Min, Wu: A simple recursive least square algorithm of space–temporal joint equalizer. Dig. Signal Proc. **22**, 1145–1153 (2012)
13. Hall, W.E., Gupta, N.K., Smith, R.G.: Identification of aircraft stability and control coefficients for the high angle-of-attack regime. Report documentation page, Palo Alto (1974)
14. Young, P.C.: Recursive Estimation and Time-Series Analysis. Springer, London, New York (2011)

Actor Model Approach to Control Stability of an Articulated Vehicle

Kornel Warwas and Krzysztof Augustynek

Abstract The paper presents an application of actor model to control braking torques on wheels of an articulated vehicle in an untripped rollover manoeuvre. The numerical model of the articulated vehicle and dynamic optimisation have been used to calculate appropriate braking torques for each wheel in order to restore stability. The optimisation problem requires the equations of motion to be integrated at each optimisation step and its time-consuming. Therefore, parallel computing with using actor model system has been proposed. Actor model system has been implemented in genetic algorithm. In the paper, formulation of genetic algorithm with actor system and results obtained from dynamic optimisation have been presented and compared.

Keywords Actor model · Genetic algorithm · Parallel computing · Dynamic optimisation · Articulated vehicle

1 Introduction

The dynamic optimisation of sophisticated physical systems like multibody system, articulated vehicle is time-consuming task. Improvement of the optimisation calculation time is a subject of many papers [1–3]. Parallel and distributed systems are often used in order to improve the efficiency of optimisation calculations [1, 2, 4]. Some of the algorithms allow to split computational effort on separate threads, processes or cluster's nodes in a natural way. Complexity of the mathematical model of the system and implementation of the optimisation methods have big influence on the effectiveness of the optimization process. Currently, the

K. Warwas (✉) · K. Augustynek
Department of Computer Science and Automatics, University of Bielsko-Biala, Bielsko-Biala, Poland
e-mail: kwarwas@ath.bielsko.pl

K. Augustynek
e-mail: kaugustynek@ath.bielsko.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_4

development of computer hardware and software allow to release from the monolithic architecture to micro-services. Then the business logic can be separately processed and particular parts of the system can communicate with each other [5]. Such approach allows to reduce significantly the time of calculation by parallel processing and provide high scalability of the system. One of the ways such implementation of the system and calculations is actor model approach [4, 6, 7]. The actor model for concurrent and parallel programming is gaining popular due to its high level of abstraction and its ability to efficient use of multicore and multi-processor machines. Implementing applications on top of those primitives has proven challenging and error-prone. Additionally, mutex-based implementations can cause queueing and unmindful access to (even distinct) data from separate threads in parallel can lead to false sharing: both decreasing performance significantly, up to the point that an application actually runs slower when adding more cores. The actor model describes calculation process as a result of interaction between active objects (actors). Actors interact with each other by asynchronous messages passing. Each actor can process the received message according to implemented behavior, forward a message to other actor or wait for a new message. It exists many actor model implementations like Akka, Akka.NET, CAF, Theron [8, 9]. Each one has embedded queue system and allows to implement well known architectonic (e.g. CQRS) and design patterns (e.g. chain of responsibility). Some of actor model systems can be deployed in cloud systems like Amazon Web Services or Microsoft Azure.

In this paper genetic algorithm with actor model approach has been presented. During optimisation cornering manoeuvre of the articulated vehicle has been considered. The aim of the optimisation calculations is maintaining stability of vehicle and preventing before its rollover during manoeuvre. Rollover accidents of articulated vehicles are especially violent and cause greater damage and injury than other accidents. Many systems help in preventing vehicle rollovers, as they can automatically adjust the braking pattern for each wheel, possibly giving the driver greater control [3, 10–12]. Presented method is based on the control of braking torques in the case of losing the stability. Braking torques patterns, which have to be applied to each wheel of the vehicle, are obtained by solving an optimisation task.

2 Mathematical Model

The model of the articulated vehicle has been formulated as a system of rigid bodies consisting of a tractor, a fifth wheel and a semi-trailer. It is assumed that the tractor is a rigid body, which motion has been described by means of six generalized coordinates (Fig. 1) which can be written in the following form:

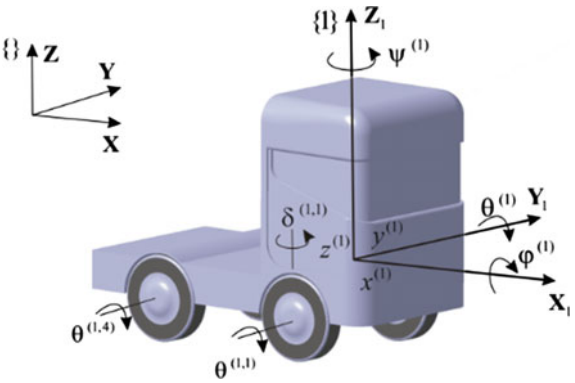


Fig. 1 The model of the articulated vehicle tractor

$$\tilde{\mathbf{q}}_T^{(1)} = \begin{bmatrix} x^{(1)} & y^{(1)} & z^{(1)} & \psi^{(1)} & \theta^{(1)} & \varphi^{(1)} \end{bmatrix}^T \tag{1}$$

where: $x^{(1)}, y^{(1)}, z^{(1)}$—translational components of the tractor displacement, $\psi^{(1)}, \theta^{(1)}, \varphi^{(1)}$—rotational components of the tractor displacement.

Wheels are connected with the tractor and each wheel has one degree of freedom $\theta^{(1,i)}, i = 1, \ldots, 4$ and the generalized coordinates vector has the form:

$$\tilde{\mathbf{q}}_{TW}^{(1)} = \begin{bmatrix} \theta^{(1,1)} & \theta^{(1,2)} & \theta^{(1,3)} & \theta^{(1,4)} \end{bmatrix}^T \tag{2}$$

It has been assumed that suspension stiffness has been reduced to contact point of a tire with a road. In paper [3] it has been shown that the relative error resulting from such simplification is less than 5 % in relation to results obtained from the road tests. Therefore, vector of the generalized coordinates contains only front wheels steering angles $\delta^{(1,1)}$ and $\delta^{(1,2)}$:

$$\tilde{\mathbf{q}}_{TS}^{(1)} = \begin{bmatrix} \delta^{(1,1)} & \delta^{(1,2)} \end{bmatrix}^T \tag{3}$$

Motion of the fifth wheel in relation to the tractor (Fig. 2) has been described by one degree of freedom (pitch angle $\theta^{(2)}$).

Generalized coordinate of the fifth wheel is the component of the following vector:

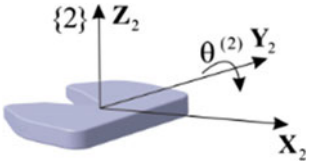


Fig. 2 The model of the articulated vehicle fifth wheel

$$\tilde{\mathbf{q}}_F^{(2)} = \left[\theta^{(2)}\right] \tag{4}$$

Semi-trailer (Fig. 3) has one degree of freedom, inclination angle $\psi^{(3)}$, with respect to the fifth wheel and its generalized coordinates vector are given by:

$$\tilde{\mathbf{q}}_S^{(3)} = \left[\psi^{(3)}\right] \tag{5}$$

The semi-trailer has six wheels (Fig. 3) and each has one degree of freedom, $\theta^{(3,i)}, i = 1, \ldots, 6$, which is component of the following generalized coordinates vector:

$$\tilde{\mathbf{q}}_{SW}^{(3)} = \left[\, \theta^{(3,1)} \quad \theta^{(3,2)} \quad \theta^{(3,3)} \quad \theta^{(3,4)} \quad \theta^{(3,5)} \quad \theta^{(3,6)} \,\right]^T \tag{6}$$

Equations of vehicle motion have been formulated using Lagrange equations of second kind and homogenous transformations [7]. It can be written in the following form [6, 13]:

$$\begin{aligned} \mathbf{A}\,\ddot{\mathbf{q}} + \boldsymbol{\Phi}_q \mathbf{r} &= \mathbf{f} \\ \boldsymbol{\Phi}_q^T\,\ddot{\mathbf{q}} &= \mathbf{w} \end{aligned} \tag{7}$$

where: $\mathbf{A} = \mathbf{A}(t, \mathbf{q})$—mass matrix,
$\mathbf{f} = \mathbf{f}\left(t, \mathbf{q}, \dot{\mathbf{q}}, \mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(i)} \ldots, \mathbf{M}^{(n_w)}\right)$—vector of external, Coriolis and centrifugal forces,
$\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}$—displacement, velocity and acceleration vectors,
$\mathbf{M}^{(i)}$—vector of discrete values of braking torques acting on the ith wheel,
n_w—number of wheels,
$\boldsymbol{\Phi}_q$—constraints matrix,

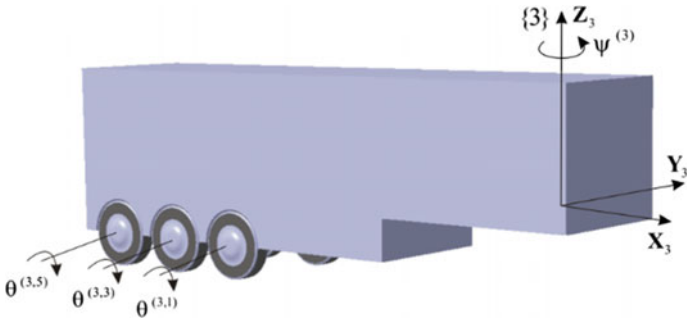


Fig. 3 The model of the articulated vehicle semi-trailer

r—vector of constraint reactions,
w—vector of the right sides of the constraint equations.

The details of the procedure which leads to formation of Eq. (7) with a description of elements in the matrix **A** and the vector **f** is presented in [6].

3 Formulation of the Articulated Vehicle Optimisation Problem

Articulated vehicles rollover is one of the most dangerous road manoeuvres. This situation happens mostly during the unforeseen lane-change manoeuvre [5] whilst cornering of the vehicle. Such manoeuvre has been performed when the preplanned vehicle trajectory would collide with an obstacle. When the obstacle is detected, the trajectory is translated to other traffic lane, as shown in Fig. 4, in order to avoid collisions.

Stability of the articulated vehicle can be restored by an appropriate control of braking torques applied to each wheel of the vehicle. Let us consider vector $\mathbf{M}^{(i)}$ containing discrete values of braking torque applied on the ith wheel. A continuous function $M^{(i)}(t)$ can be obtained using spline functions of the 3rd order. The vector of the decisive variables, which contains discrete values of the braking torques of all wheels, can be written in the form:

$$\mathbf{M} = \left[\mathbf{M}^{(1)} \quad \ldots \quad \mathbf{M}^{(i)} \quad \ldots \quad \mathbf{M}^{(n_w)}\right]^T = \left(M_j\right)_{j=1,\ldots,m} \tag{8}$$

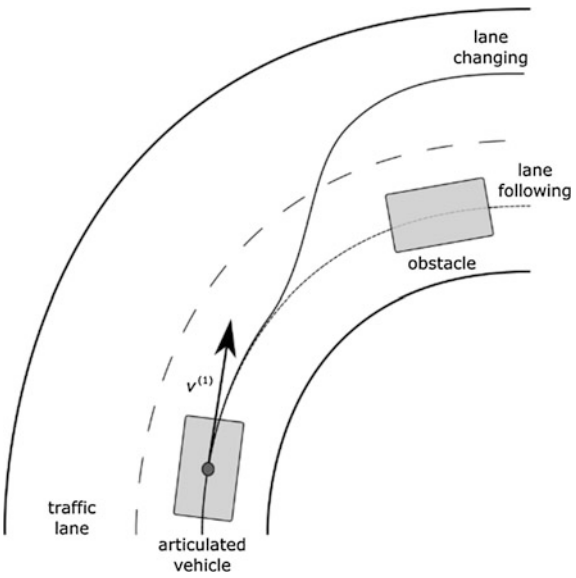


Fig. 4 Lane changing and lane following manoeuvre during cornering

where m—number of decisive variables.
$\mathbf{M}^{(i)} = \left(M_k^{(i)}\right)_{k=1,\ldots,n}$,
n—number of discrete values of the braking torque.

In the presented problem, braking torques calculated for fixed initial vehicle velocity and the front wheels' steering angle have to fulfil the following conditions:

- the articulated vehicle cannot lose stability during the manoeuvre,
- longitudinal velocity loss has to be as small as possible,
- lateral displacement of the vehicle is limited by the standard road width.

Above assumptions are taken into account in the objective function and also in optimisation constraints. The stability conditions can be assured by minimizing the functional [8]:

$$\Omega(\mathbf{M}, \mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{t_e}\left(C_1 \int_0^{t_e} \left(\varphi^{(1)}\right)^2 dt + C_2 \int_0^{t_e} (v_0 - v_e)^2 dt\right) \rightarrow \min \tag{9}$$

where C_1, C_2—empirical coefficients,
t_e—time of simulation,
v_0—initial velocity.
$v_e = \sqrt{\left(\dot{x}^{(1)}(t_e)\right)^2 + \left(\dot{y}^{(1)}(t_e)\right)^2}$—total final velocity of the tractor.

In the considered optimisation problem, inequality constraints can be written as follows:

$$\left[M_{min} - M_1 \quad \ldots \quad M_{min} - M_i \quad \ldots \quad M_{min} - M_m\right]^T \leq \mathbf{0} \tag{10}$$

$$\left[M_1 - M_{max} \quad \ldots \quad M_i - M_{max} \quad \ldots \quad M_m - M_{max}\right]^T \leq \mathbf{0} \tag{11}$$

where: M_{min}, M_{max}—acceptable minimal and maximal braking torque values.

The classical genetic algorithm and genetic algorithm with actor model approach have been applied in order to solve formulated minimization problem [8, 9]. In the next paragraph genetic algorithm with actor model approach method will be presented in details.

4 Actor Model Approach in Genetic Algorithm

Actors are objects which encapsulate state, behaviour and they can communicate exclusively by exchanging messages which are placed into the recipient's mailbox (Fig. 5).

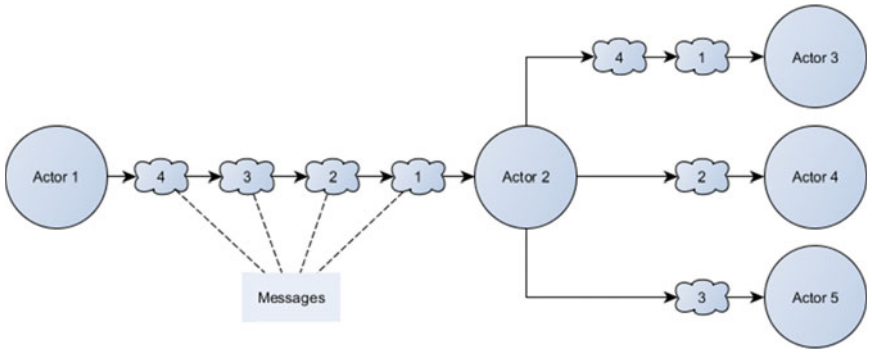


Fig. 5 Messages exchanging in actor model system

One actor can split up its tasks and delegate until they become small enough to be handled in one piece in order to build fault-tolerant, hierarchical system. Each actor can be treated as separate thread on one computer or it can be a process on machine in cluster of distributed systems. Such approach is therefore suitable for designing and development of a parallel and distributed information systems. Actor model basis on event-driven system in which actors process events and generate responses or more requests in asynchronous way. Message passing in actor model is network transparent.

In this paper CAF framework has been applied in order to implement genetic algorithm with actor model approach. CAF allows to transparently connect actors running on different machines and operating systems via the network. It integrates multiple computing devices such as multi-core CPUs, GPGPUs, and even embedded hardware. In presented approach first actor (system) creates next actor (generation) which is responsible for managing genetic algorithm generations (Fig. 6a). This actor produces many coworkers (crossover operators) which perform next operation in non-blocking and asynchronous way. Number of those child actors depends on number of individuals in population. When crossover operation is finished each actor produces new actor which is responsible for mutation operation. Individuals obtained from mutation are forwarded to actors which evaluate objective function value. This operation is time consuming because it is necessary to integrate dynamic equation of motion of the system. The main advantage of such approach is that the calculation of the objective function can be performed in parallel and non-blocking way. The next actor waits until calculation of the objective function will be finished by all actors. It collects results and creates new parent's population according to natural selection rule. These steps are repeated until stop condition is not satisfied. Messages which are sent during one step of presented approach have been shown in Fig. 6b.

It is assumed real-number representation of genes in chromosomes and the following genetic operators have been used [14]: hybrid selection consists of natural selection combined with elitist selection, arithmetical one-point crossover, in

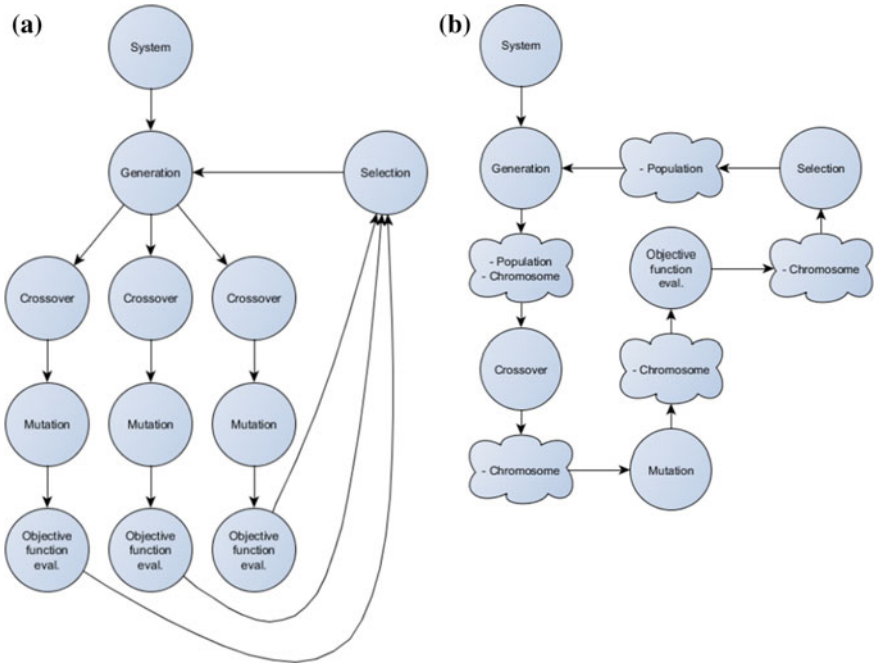


Fig. 6 Actors (**a**) and message flow between actors (**b**) in genetic algorithm

which a new chromosome is a linear combination of two vectors and non-uniform mutation.

It can be concluded that application of actor model frameworks offers many facilities which provide abstraction layer for low-level operations such as: queuing messages, multithreading calculations, synchronization mechanisms and location transparency.

5 Numerical Simulations

During simulations optimal braking torques have been calculated using classical genetic algorithm and its modification using actor model approach. It has been analysed lane change manoeuvre whilst cornering of the articulated vehicle. When the appropriate braking torques are not applied the rollover of the vehicle occurs. Additionally, when only the articulated vehicle cornering manoeuvre is considered the truck is stable. Lane change manoeuvre start at $t = 3$ s of simulation and in the same time additional braking torques have been applied. This torques are acted till the end of the simulation ($t_e = 6$ s). It has been assumed that vehicle initial velocity $v_0 = 45$ km/h. Physical parameters of the articulated vehicle have been taken from

[3]. Interpolation of the braking torque has been performed for $m = 7$ interpolation nodes which are decisive variables in considered problem. Bulrish-Stoer-Deuflhard [15] method with adaptive step size has been used for integration equations of motion. Computing has been performed on Intel Core i5 2, 6 GHz and 8 GB LPDDR3 1600 MHz RAM computer with OS X El Capitan.

Objective and fitness function values obtained for various number of individuals in the population (actors) using genetic algorithm with actor model approach have been presented in Fig. 7a. Differences between time of optimisation calculations obtained for classical genetic algorithm and its modification with actor model can be seen in Fig. 7b.

As shown in Fig. 7 differences between objective function or fitness function for various number of individuals are small. For further calculations it has been assumed thirty individuals in the population. Such number of individuals seems to be a good compromise between time and accuracy of calculations. When comparing the times of the calculation obtained for both analysed methods it can be noticed that optimisation time is significantly shorter in the case of a model actor approach regardless of the number of individuals in the population. Figure 8a shows front wheels steering angle course of the articulated vehicle applied during simulations. Results obtained from optimisation using analysed in the paper modification of genetic algorithm are presented in Fig. 3b, c which show courses of tractor roll angle and its trajectory. Figures present also results of simulations for lane following (cornering) manoeuvre (1) and those with the lane change manoeuvre before optimisation (2) and after optimisation (3). Courses of the optimal braking torques acting on the wheels $w_1, \ldots, w_6$ have been shown in Fig. 8d.

It can be seen that dynamic optimisation using actor model approach gives solution which allows articulated vehicle to maintain the stability during manoeuvre.

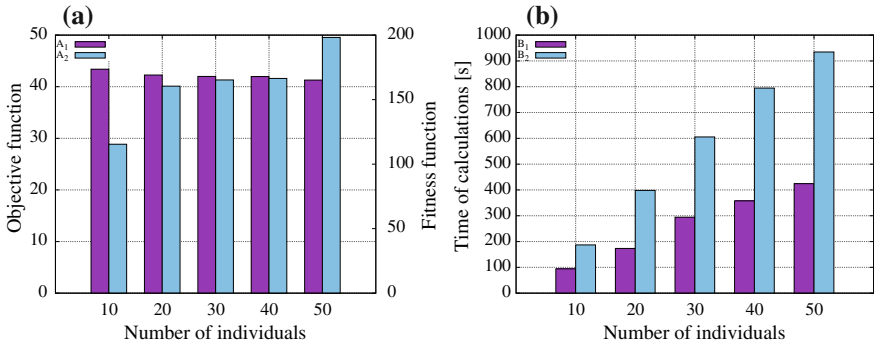


Fig. 7 Values of the objective (A_1) and fitness (A_2) function obtained for various number of individuals (**a**), values of the optimisation calculations time (**b**) obtained for various number of individuals with actor model system (B_1) and with classical genetic algorithm approach (B_2)

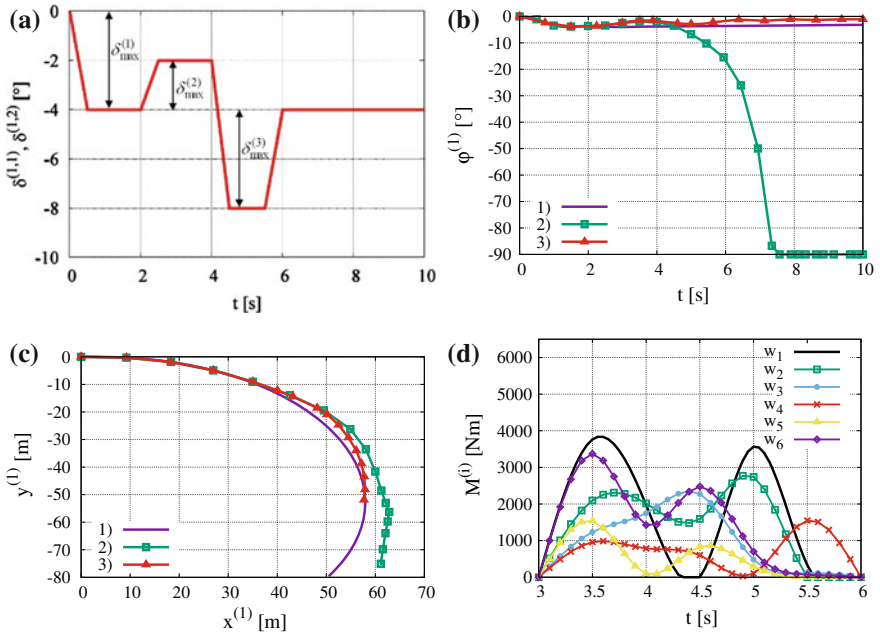


Fig. 8 Courses of: steering angle (**a**), the tractor roll angle (**b**), the tractor trajectory (**c**), optimal braking torques (**d**)

6 Conclusions

The articulated vehicle rollover is strongly associated with severe injury and fatalities in highway accidents. Stability can be achieved by an appropriate control of braking torques. In the paper, the problem of controlling brakes has been formulated as dynamic optimisation task. This task has been solved using classical genetic algorithm and its modification with using actor model approach. The long duration of the optimisation process results from the necessity of integrating equations of motion in each step. Actor model approach allows to reduce time of calculations by dividing computational effort into smaller tasks which are performed by single actor in asynchronous way. Results show that time of calculations obtained for genetic algorithm with various number of individuals using actor model approach is averagely 50 % shorter than the time obtained without this modification. It should be noted that the results have been obtained on a personal computer with a 4th core processor. According to the authors better results can be obtained on servers with larger number of cores or on computing cluster. Although the actor model approach is known since the 70 s of the last century, but now it can be noticed that interest in this approach in application to modern business systems have been increased. This approach can be easily applied to scientific/numerical

applications. In addition to the benefits mentioned previously, it can be obtained clear source code which is logically split into small atomic parts, well-designed object-oriented architectures and easy to maintain and extend.

References

1. Eberhard, P., Dignath, F., Kübler, L.: Parallel evolutionary optimization of multibody systems with application to railway dynamics. Multibody Syst. Dyn. **9**(2), 143–164 (2003)
2. Augustynek, K., Warwas, K., Polański, A.: Application of the genetic algorithms and distributed computing in task of the reduction of vibrations of a satellite. In: 7th Conference Computer Methods and Systems, pp. 237–242 (2009)
3. Warwas, K.: Analysis and control of motion of articulated vehicles with flexible elements. PhD Thesis, University of Bielsko-Biała, Bielsko-Biała (2008)
4. Plotnikova, N.P., Fedosin, S.A., Teslya, V.V.: Gravitation Search Training Algorithm for Asynchronous Distributed Multilayer Perceptron Model, New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, vol. 312, Lecture Notes in Electrical Engineering, pp. 417–423 (2015)
5. Amaral, M., Polo, J., Carrera D., Mohomed, I., Unuvar, M., Steinder, M.: Performance evaluation of microservices architectures using containers. In: Network Computing and Applications (NCA), 2015 IEEE 14th International Symposium, pp. 27–34 (2015)
6. Hewitt, C.: Actor model of computation for scalable robust information systems. Inconsistency Robustness (2015)
7. Lim, Y.H., Tana, J., Abramson, D.: Solving optimization problems in Nimrod/OK using a genetic algorithm. Proc. Comput. Sci. **9**, 1647–1656 (2012)
8. Vaughn, V.: Reactive Messaging Patterns with the Actor Model Applications and Integration in Scala and Akka. Addison-Wesley (2016)
9. Charousset, D., Schmidt, T.C., Hiesgen, R.: CAF—The C++ actor framework for scalable and resource-efficient applications. IN: Proceedings of the 5th ACM SIGPLAN Conference on Systems Programming and Applications (SPLASH '14) Workshop AGERE!, New York (2014)
10. Huang, H.H.: Controller design for stability and rollover prevention of multi-body ground vehicles with uncertain dynamics and faults. PhD Thesis, Graduate School of The Ohio State University, Ohio (2009)
11. Yedavalli R.K.: Robust stability and control of multi-body ground vehicles with uncertain dynamics and failures. Technical Report, Ohio State University Research Foundation, Ohio (2010)
12. Yao, Z. et al.: Dynamic simulation for the rollover stability performances of articulated vehicles. J. Automobile Eng. 771–783 (2014)
13. Bauchau, O.A.: Flexible Multibody Dynamics, Solid Mechanics and Its Applications. Springer, Netherlands (2011)
14. Affenzeller, M., Wagner, S., Winkler, S., Beham, A.: Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications. CRC Press (2009)
15. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical recipes, 3rd edn. In: The Art of Scientific Computing. Cambridge University Press, Cambridge (2007)

Part II
Image Processing

Automatic Detection of Nerves in Confocal Corneal Images with Orientation-Based Edge Merging

Adam Brzeski

Abstract The paper presents an algorithm for improving results of automatic nerve detections in confocal microscopy images of human corneal. The method is designed as a postprocessing step of regular detection. After the nerves are initially detected, the algorithms attempts to improve the results by filling undesired gaps between single nerves detections in order to correctly mark the entire nerve instead of only parts of it. This approach enables for reliable detection of long nerves, which can be used for more accurate elimination of short detections and therefore eliminating noise. The method evaluates candidate gaps by analysing the orientation of segments to be merged in the area near the gap. Segments with sufficiently high orientation compliance are merged to form a single nerve detection. Despite using only a simple technique for initial detection of nerves, the method enabled achieving a fairly low ratio of wrong nerve detections with a balanced level of overall accuracy.

Keywords Confocal microscopy · Automatic corneal nerve detection · Nerve segmentation

1 Introduction

Confocal microscopy is a novel, non-invasive technique in diagnosing pathologies in human corneal. The examination provides a crucial insight into cell and microstructure level of the corneal, also enabling imaging of almost transparent structures. Also, it enables observation of all of the layers of the corneal, which opens new fields of diagnostics. One of the interesting applications of corneal confocal microscopy (CCM) is the examination of nerves. Since the nerve fibres in corneal are dense and at the same time can be well detected with CCM, changes in

A. Brzeski (✉)
Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, Gdansk, Poland
e-mail: brzeski@eti.pg.gda.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_5

nerves structure can be observed in early stages. For this reason, CCM gains importance as a technique for early diagnosis of diabetes [1], which causes twisting and thinning of the nerves. This approach requires however proper assessment of the scale of the tortuosity of the nerves for making correct decisions. Objective and precise measurement of features of the nerves structure is also difficult for humans, which results in a need for computer-based measurements with automatic detection of nerves. Computer-based detection, in turn, faces the problem of precise identification of the nerve fibres, especially when they are severely affected by the disease. Nerves tend to have small gaps, which should be identified in order to correctly mark the entire nerve instead of splitting it into several fibres. Confocal images also include a significant amount of noise, and distinguishing the nerves from other structures visible in the pictures is not always clear. Precise detection and tracking of the nerves is therefore both a important and a challenging task for computer vision in medicine.

2 Existing Work

The problem of automatic detection, tracking and measurement of corneal nerves in confocal images has already been investigated for at least a decade. In 2006 Ruggeri et al. [2] presented nerve recognition approach designed as a modification of an algorithm for tracking vessels in retinal images. The method firstly involved a preprocessing step to normalize luminosity and contrast, followed by an averaging resulting in a clearer appearance of the nerves. Then seed points were identified from which nerve detections were grown by classifying pixels with a fuzzy c-mean clustering. Further tracking of the nerves was carried out on a high-pass filtered image. The authors also propose a technique for connecting nerves between segments. Pairs of end-points are evaluated by drawing five different arcs simulating possible connections. Arcs covering most bright pixels are chosen and then accepted as connections if they are sufficiently brighter than neighbouring arcs. Finally, the method included also techniques for eliminating false recognitions. In 2008, the authors reported improved algorithm [3], including application of Gabor filter in the additional nerve tracking step as well as measures for total lengths of nerves and nerve densities. In 2011 also nerve tortuosity measures were added [4].

In 2010 Dabbah et al. [5] presented a dual-model nerve detection algorithm. The two separate models were designed for describing the nerves and the background, utilizing 2D Gabor wavelet and a Gaussian envelope. The detected nerves are then measured and the distributions are passed as input to SVM classifier for a task of binary classification detecting the presence of nerves. In their next work [6], the authors also utilize a method based on least mean square for assessing nerve orientation consistence, enabling reducing errors in nerve detection. The proposed algorithm was published and distributed in a notable software package named ACCMetrics, enabling automatic detection of nerves and measurement of nerve lengths, areas, fractal dimensions and other features.

In 2012 Ferreira et al. [7] proposed another approach based on Gabor filters. The image is firstly equalised with contrast-limited adaptive histogram equalization method. The nerves are detected by measuring phase symmetry of the structures in the image using a quadrature of Log Gabor filters evaluated for six orientations. Then nerve reconstruction process is carried out. Firstly seed points are selected by choosing nerve candidates that intersect with image diagonals. A few conditions reflecting expected features of nerves were then applied to eliminate false seeds. From the chosen seeds the nerves are reconstructed using a sequence of simple morphological operations. False nerves are again eliminated by evaluating skeleton of the detection and discarding small nerve branches. Finally a set of measures are evaluated on the detected nerves: toruosity, nerve length and nerve density.

3 The Proposed Algorithm

The presented method is designed as postprocessing algorithm for the results of regular nerve detection procedures enabling merging disjoint edges resulting from noise or depth variation of the nerves, which makes them fade out. The algorithm enables achieving higher accuracy in nerve tracking and identification of separate nerves, which is important for evaluating features of the nerves. The accuracy is increased by limiting the number false detections. Since the method enables more precise identification of long nerve fibres, the short detections can be identified as

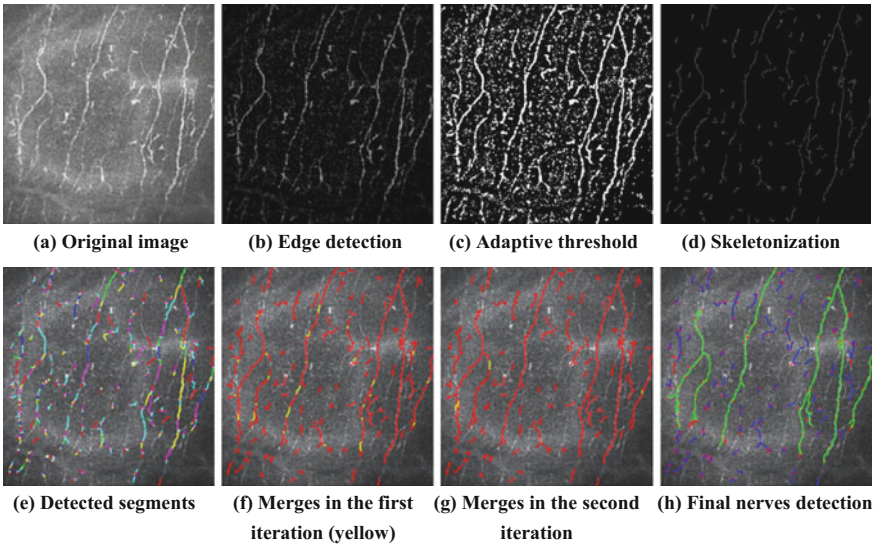


Fig. 1 Intermediate and final results of the algorithm acquired on a sample corneal image

noise and discarded with greater accuracy. The merging is achieved by deducting the gaps to be filled on the basis of the orientation of the neighbouring nerves. Consideration of the orientation and its consistence enables distinguishing the nerves from contours originating from noise.

The algorithm has several processing steps, which are described in the following subsections. Intermediate results of the algorithm for sample confocal image were presented in Fig. 1. The outline of the algorithm was presented in the Fig. 2.

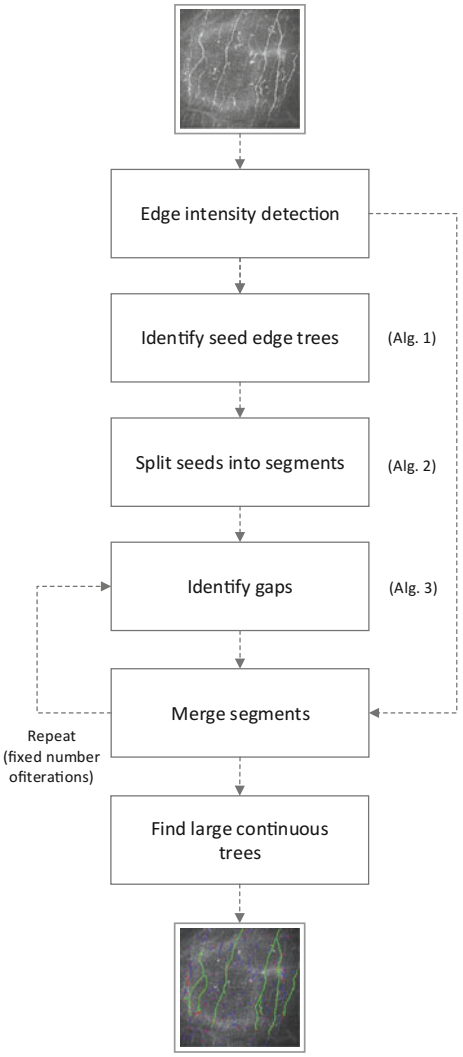


Fig. 2 The outline of the algorithm

3.1 Edge Intensity Detection

Since nerves appear in the confocal images in white color on dark background, in this step we look for white edges. The nerves however vary in intensity and this information will be used later in the processing. The images also contain large, bright areas in the background that make nerves less clear. In order to eliminate the bright areas, median blur with large kernel is applied to the image and the result is subtracted from the original image. The operation results in clearer appearance of the nerves, at the same time preserving the intensity their intensity values.

3.2 Seed Trees Identification

In this step seed edge trees for further merging are detected. The edge tree is defined as separate nets of segments composed of edges. The seeds need to have high probability of belonging to actual nerve. Therefore, only long, continuous edge trees with high intensity are extracted. The procedure is outlined in Algorithm 1.

Algorithm 1 Seed edge trees identification

```
1: function GET_SEEDS(img)
2:     img ← adaptive_threshold(img)
3:     img ← thinning(img)
4:     edge_trees ← get_separate_trees(img)
5:     seeds ← ∅
6:     for all tree ∈ edge_trees do
7:         if tree.size() > thresh then
8:             seeds ← seeds + segment
9:     return seeds
```

3.3 Splitting Seeds into Segments

The purpose of this step is to split the seed trees into elementary segments. Segments can be curved. The trees are split into a minimum possible number of segments. The procedure is described in Algorithm 2.

3.4 Finding and Filling Gaps Between Segments

The main part of the algorithm. The purpose is to identify gaps that have sufficiently good properties for being filled in order to merge the two segments. For this purpose each gap is evaluated in terms of orientation compliance. The measure is computed

for subsegments of the candidate segments, which start at the ends to be connected and with length equal to the distance between the endpoints multiplied by two. For each pixel of the subsegments an angle is computed defined by the pixel and the two endpoints. An average angle is evaluated for each pixel of both subsegments, which denotes the cost of connecting the segments. If a any of the segments is not long enough to evaluated the measure, penalty is applied. Finally, harmonic mean of the costs for the two subsegments are evaluated, and if the result is below a fixed threshold, the segments are chosen to be merged. The chosen gaps are then filled by finding the shortest path connecting ends of the segments to be merged with A-star algorithm with costs evaluated over the edge intensity map acquired from the first step of the algorithm. The procedure is outlined in Algorithm 3.

Algorithm 2 Splitting seeds into segments

```
1: function GET_SEGMENTS(tree)
2:     segments ← ∅
3:     for all tree ∈ edge_trees do
4:         junctions ← get_junctions(tree)
5:         id ← 0
6:         if junctions = ∅ then
7:             segments ← segments + tree
8:             continue
9:         for all j ∈ junctions do
10:            adjacent_points ← get_adjacent_points(j)
11:            for all point ∈ adjacent_points do
12:                if point.id = null then
13:                    point.set_id(id)
14:                    id ← id + 1
15:        for all j ∈ junctions do
16:            adjacent_points ← get_adjacent_points(j)
17:            for all point ∈ adjacent_points do
18:                segment ← get_segment_from(j, point)
19:                segments ← segments + segment
20:    return segments
21:
22: function GET_JUNCTIONS(tree)
23:    junctions ← ∅
24:    for all point ∈ tree do
25:        if degree(node(point)) > 2 then
26:            junctions ← junctions + point
27:    return junctions
```

3.5 *Finding Final Segments*

Filling gaps between segments results in merging of some of the initial edge trees and leads to a new set of edge trees. For choosing the final trees for the result of nerve detection, again we assume extracting only the trees what are large enough. Therefore all trees smaller than a fixed threshold are discarded, while passing trees are returned as the final nerve detection result.

Algorithm 3 Find segments to be merged

```
1: function FINDS_SEGMENTS_FOR_MERGING(segments)
2:     end_points ← get_end_points(segments)
3:     proposal_map ← ∅
4:     for all p ∈ end_points do
5:         proposals ← get_close_points(p, end_points)
6:         results ← null
7:         for all proposal ∈ proposals do
8:             result ← evaluate(p, proposal)
9:             results ← results + result
10:        best_proposal ← get_best(proposals, results)
11:        proposal_map[p] ← best_proposal
12:    for all p ∈ end_points do
13:        proposal ← proposal_map[p]
14:        if proposal_map[proposal] = p then
15:            merge(p, proposal)
16:
17: function GET_END_POINTS(segments)
18:    end_points ← ∅
19:    for all segment ∈ segments do
20:        for all pointinsegment do
21:            if degree(node(point)) = 1 then
22:                end_points ← end_points + point
23:    return end_points
```

4 Results

The algorithm was evaluated on the publicly available Corneal Nerve Tortuosity Data Set [4], containing 30 images 384×384 resolution. The detections of the algorithms were compared to reference masks, acquired by manually marking the nerves in the dataset images. Positive detection was considered correct if there was a positive value in the mask within the distance of 3 % of the input image width from the detection.

For each image, following measures are evaluated as ratios in respect to the total number of nerve pixels in reference masks: ratio of correctly detected nerve pixels (detected), the ratio of undetected nerve pixel (missed) and the ratio of wrongly

identified nerve pixels. The results were averaged over the entire dataset. The correct detection rate was 0.82, missed 0.39, and wrong detection rate equaled to 0.15.

5 Conclusions

The acquired results showed a balanced level correctly and wrongly identified nerve pixels. The both ratios are fairly good, considering the fact the algorithm uses a very simple technique for initial detection of nerves, comprising only subtracting mean image and applying adaptive threshold. The task was also challenging, since the dataset contains many unclear nerves and the reference mask were prepared with an intention of choosing only the significant nerves. This behavior was therefore also expected from the algorithm, which introduces higher risk of returning wrong detection. Thus, the acquired wrong detection ratio is considered promising. Unfortunately, the missed nerve ratio is slightly too high. This result is even more affected by the imperfection of initial detection of nerves. However the nature of the reference detections that focus on the main nerves in this case, in turn, allows getting better results. The problem will be considered in the future improvement of the algorithm, which will include consideration of the length of nerves in the step of filling detection gaps, while currently the orientation is the only factor. The future work will also include application of more advanced initial nerve detection employing Gabor filters, which should significantly improve initial seeds selection.

Acknowledgment The research was funded by grants from National Centre for Research and Development (PBS2/A3/17/2013, Internet platform for data integration and collaboration of medical research teams for the stroke treatment centers).

References

1. Rosenberg, M.E., Tervo, T.M., Immonen, I.J., Muller, L.J., Gronhagen-Riska, C., Vesaluoma, M.H.: Corneal structure and sensitivity in type 1 diabetes mellitus. Invest. Ophthalmol. Vis. Sci. **41**(10), 2915–2921 (2000)
2. Ruggeri, A., Scarpa, F., Grisan, E.: Analysis of corneal images for the recognition of nerve structures. In: Engineering in Medicine and Biology Society EMBS'06, 28th Annual International Conference of the IEEE, 2006, pp. 4739–4742
3. Scarpa, F., Grisan, E., Ruggeri, A.: Automatic recognition of corneal nerve structures in images from confocal microscopy. Invest. Ophthalmol. Vis. Sci. **49**(11), 4801–4807 (2008)
4. Scarpa, F., Zheng, X., Ohashi, Y., Ruggeri, A.: Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. Invest. Ophthalmol. Vis. Sci. **52**(9), 6404–6408 (2011)
5. Dabbah, M.A., Graham, J., Malik, R.A.: Corneal confocal microscopy image quality analysis and validity assessment (2010)

6. Dabbah, M.A., Graham, J., Petropoulos, I., Tavakoli, M., Malik, R.A.: Dual-model automatic detection of nerve-fibres in corneal confocal microscopy images. Med Image Comput Comput-Assist Interv MICCAI **2010**, 300–307 (2010)
7. Ferreira, A., Morgado, A.M., Silva, J.S.: A method for corneal nerves automatic segmentation and morphometric analysis. Comput. Methods Progr. Biomed. **107**(1), 53–60 (2012)

Methods of Digital Hilbert Optics in the Analysis and Objects' Recognition

Adam Sudol

Abstract This paper describes methods on how to increase the effectiveness of objects' pictures identification based on correlation methods. The main concept of increasing the discriminant effectiveness is based on highlighting of characteristic points of recognized objects by applying Hilbert transformations. Study of the effectiveness of Digital Hilber Optics (DHO) have been performed on a set of aircrafts, whose models rendered first as binary images, and then as grayscale. It has been performed a very detailed analysis of requirements on resources of information system's which would in a real world support the discriminatory decision of objects' class for which the sample database has been created.

Keywords Object identification · Texture identification · Digital Hilbert transformations

1 Introduction

There are many ways to recognize and support object recognition, and each of them has a different efficiency, the time needed for identification, the degree of probability of correct identification and the demand for ICT system resources. These methods according to the general classification can be divided into:

- methods based on artificial intelligence,
- methods based on a comparison with the database samples.

This paper focuses on the latter methods. Not because these methods have not been well researched, but because they have not been sufficiently explored all methods of defining a unique and concise descriptions of the objects, which would permit efficient database search.

A. Sudol (✉)
Department of Technology, Opole University, Opole, Poland
e-mail: dasiek@uni.opole.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_6

Identification of the objects and the textures is very wide problem, primarily due to the countless number of domains in which it is now used [1, 3]. The final, identification phase, consists of the number of earlier stages, one of which is the process of extracting [2], or signal acquisition, as a basis for further identification procedures. During the acquisition, images can be acquired with the variant information capacity. Under natural conditions, it can be difficult to obtain images of sufficient quality. Some may be low resolution, while others suffer from noise. In such situations, it has become necessary to carry out image analysis in one of manner:

- artificially increasing the informational potential of the images by preprocessing, aimed at reducing noise, histogram filtering or increasing local contrast, for example by applying the Unsharp Mask filter or by merging several images into one [5].
- application of methods manifesting high efficiency, enabling the identification of an object with sufficient probability from the image of a naturally reduced informational potential (resolution, noise).

The latter methodology also include Digital Hilbert optics, which consists of a large collection of methods of the image post processing.

Digital Hilbert optics is a collection of methods and tools designed to improve the informational potential of digital images of identified objects, and thus the effectiveness of identification. Most correlation methods, operation of which is based on more or less complex comparative analysis have a very big problems with the identification of objects in digital images burdened with all sorts of noise. It is not just about classical additive noise but also the angular noise at which the classic comparative methods are useless. Digital Hilbert optics provides many methods to eliminating or even bypass certain types of interference that are a natural part of real images of 3D objects. Methods of identification based on a digital Hilbert optics significantly reduce computational complexity, which has a substantial impact on the time it takes to obtain a result with a certain probability of correct identification.

The purpose of this paper is to propose and to evaluate a set of methods based on Hilbert optics, improving the effectiveness of the identification of objects on digital images, which are a special form of three-dimensional signals.

For object identification it has been used simple and hybrid Hilbert transformations and post-processing aimed at significant reduction in the complexity of the description of objects so-called designators, (a unique signatures). For the latter, it became necessary to develop a database which is the main source of comparative material. Descriptions, signatures, besides minimizing their volume (creating a signature) should facilitate the development of methods for identification of an object, independent of:

- angular noise—the object's rotation in three-dimensional space,
- translational—movement of the object in three-dimensional space,
- scalar—closer or farther away from the observer.

2 Materials and Methods

Problems of identifying of object-oriented scenes belong to a class of problems poorly mathematically formalized, and the proper solutions require the use of different algorithmic and heuristic methods, advanced equipment and software. One of the main problems of this kind of information technology (IT), synthesis, and implementation of identification systems is to reduce the description volume of the identified complex shape object (CSO). It also significantly accelerates the whole process of identification with one hand, but also leads to an increase in the number of false identification.

2.1 Theoretical Basics of Digital Hilbert Transformations

Identification of objects as structural elements of the scene is a major methodological challenge in the field of artificial intelligence, and it is based on different types of filtering methods and evaluation parameters such as their shape changes and direction vector. Increasing the sensitivity of video and information systems is intended to increase in discriminatory ability in the identification process.

Digital optical methods are developed over the past half century, however, the most studied and used methods based on Fourier transformations. The use of digital Hilbert optics methods (DHO), a digital Hilbert signal processing for processing and analyzing multi-dimensional signals (images) is not yet very widespread and is in the experimental and modeling phase [4].

Hilbert Transformation (analytical signal) is defined as complex and has the form:

$$z(t) = x(t) + j\hat{x}(t) \tag{1}$$

$$\Psi(s,t) = \frac{1}{\pi(s-t)},\ \varphi(t,s) = \frac{1}{\pi(t-s)},\ X(s) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(t)}{s-t}dt,\ x(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{X(s)}{t-s}ds \tag{2}$$

where:

s time variable
$X(s)$ signal Hilbert transform—time function.

For this reason, these definitions as formulas are often expressed as:

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau = \frac{1}{\pi t} \otimes x(t), \qquad x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{x}(\tau)}{t-\tau} d\tau = -\frac{1}{\pi t} \otimes \hat{x}(t) \quad (3)$$

Due to Hilbert transformation can be practically hybridize in any kind—combined with other transformations, and with itself—it has been selected four transformations, that in previous author's studies [6, 7] exhibited the highest ability to discriminate.

Digital images are a special form of discrete signals. Their matrix character in grayed (gray-scaled) images forces periodic, local, Hilbert transform operation. By its basic nature, it is an infinite in its form, as well as Fast Fourier Transform, from which the above mentioned uses.

In order to allow its use on digital images of scenes containing elements, they must be such situated that the test subjects did not touch the border of the image. It is experimentally determined that the transition to Hilbert did not violate the cyclical nature of the image, the image must be expanded on each side by half its dimension. This approach is illustrated on Fig. 1, wherein the test object is a white rectangle 128 × 128 pixel, and black border (background) has then a width of 64 pixels on each side.

Figure 1 illustrates the result of Hilbert transformation on the binary image. It is the basis for all examined hybrids.

The input image shown in the Fig. 1 has the dimensions of 256 × 256 pixels. Blacks area a value of 0 and whites—the value of 1.

Figure 2 shows the spectral form of the Hilbert transform of the square. The paper simultaneously uses both two approaches as shown in Fig. 2a, b. Figure on the left (Fig. 2a) represents the natural form of the Hilbert transform, and on the right (Fig. 2b) there is its module. It is the basis for one of the many methods of object's description minimization, so-called the characteristic points method (CPM).

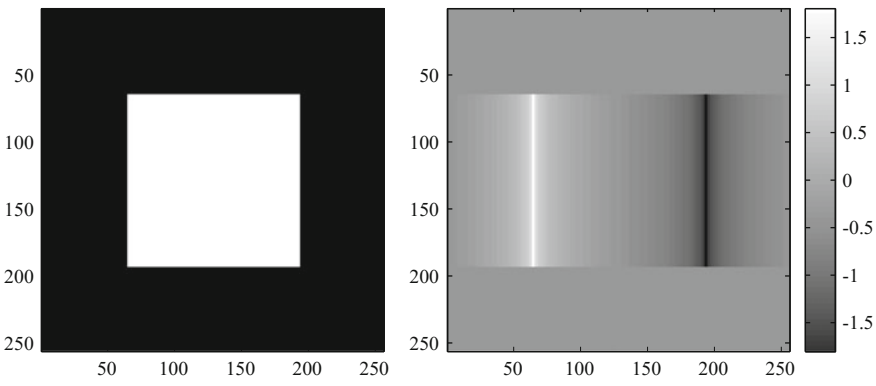


Fig. 1 The primary image and its Hilbert transform

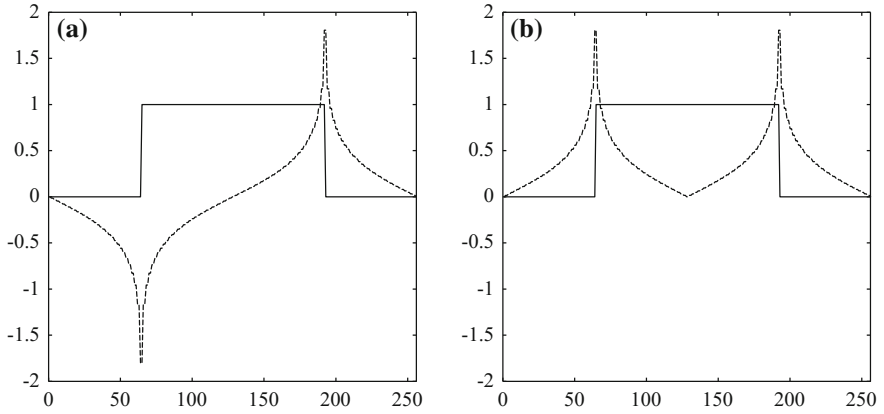


Fig. 2 Spectral form of Hilbert transform. **a** Hilbert transform (the *dotted line*), **b** Hilbert transform module (the *dotted line*)

2.2 Investigated Objects

Based on a preliminary correlation analysis of the flying objects' images, it has been selected six classes of objects (airplanes), which were further investigated. The study involved aircraft images in black and white and shades of gray. They have been rendered in two main views. One view was from the side, imitating observation the aircraft from the ground, and the second view was from above, imitating observation from the satellite.

Whenever paper mentions a particular object, then in order to uniquely identify this object, the number described in Fig. 3 will be used.

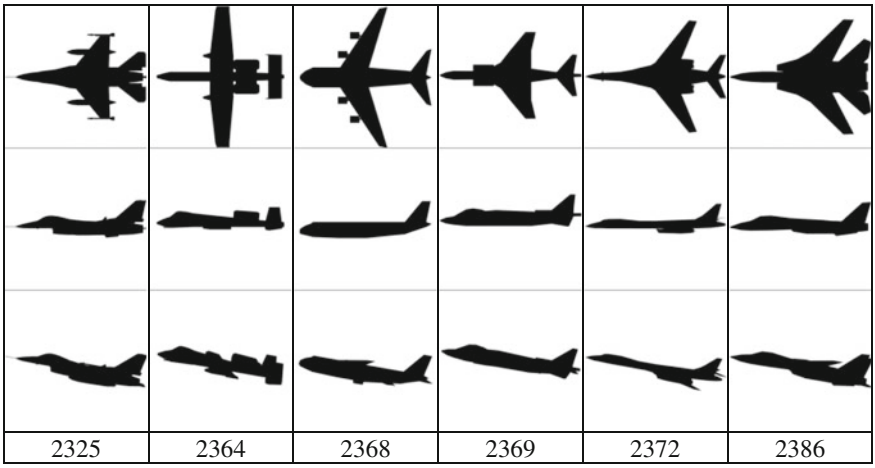


Fig. 3 Set of aircrafts investigated in the experiment

All tested objects were preprocessed by artificial increasing in the size of the image of the object under test. Originally, all object images have had a size of 128 × 128 pixels. Preprocessing consisted of adding some white space (rims with a thickness of 64 pixels) around the object symmetrically on each side, thereby generating bitmap with a size of 256 × 256 pixels.

2.3 *Investigated Transformations*

It has been investigated in detail the following forms of digital images:

- primary images—not processed any transformation (Fig. 3).
- images processed with fast Fourier transformation,
- images processed with isotropic Hilbert transformation,
- images processed with anisotropic Hilbert transformation,
- images processed with isotropic Foucault transformation,
- images processed with anisotropic Foucault transformation.

Fast Fourier Transformation (FFT)
Already at the stage of preprocessing it has been applied certain mechanisms increasing ability to discriminate of investigated transformations.

E.g. for fast Fourier transform it has been used its variety FFTShift with exchanged positions of its quarters and removed a "0 Hz" region (center of matrix) (Fig. 4).

Below are the results of calculations interclass correlation (different objects) for deleted and undeleted zero elements of the Fourier transform.

The above data clearly shows that the zero frequency components, which are currently located in the center of transform increase interclass correlation coefficient, which at a later stage of calculations would have a significant impact on the expansion of a set of similar objects. The relative reduction in the correlation coefficients shown in Table 1 compared to identical coefficients in Table 2 from 6.7 to 10.3 % allows quite reduced the set of the objects "suspected" to similarity to the investigated object.

Isotropic Hilbert Transformation
Isotropic Hilbert transformation (HTI) is the first transformation commonly used in this study. It sums up the two transformed images: one converted in accordance with the nature of the Hilbert transformation (on x-axis) and the second converted as transposition of the original image (on x′). Hilbert transforming itself, is effective only on x-axis, however in order to obtain an increase in the local contrast and thereby detect the edge, regardless of their projection to the axis of the transformation, it is necessary to summing two transformations.

Figure 5 shows the Hilbert transform applied to the image on Fig. 3. The transformation is extremely useful in detecting the edges of the object, where HTI (Fig. 5a) itself detects the edges of the object, indicating the trends in brightness on

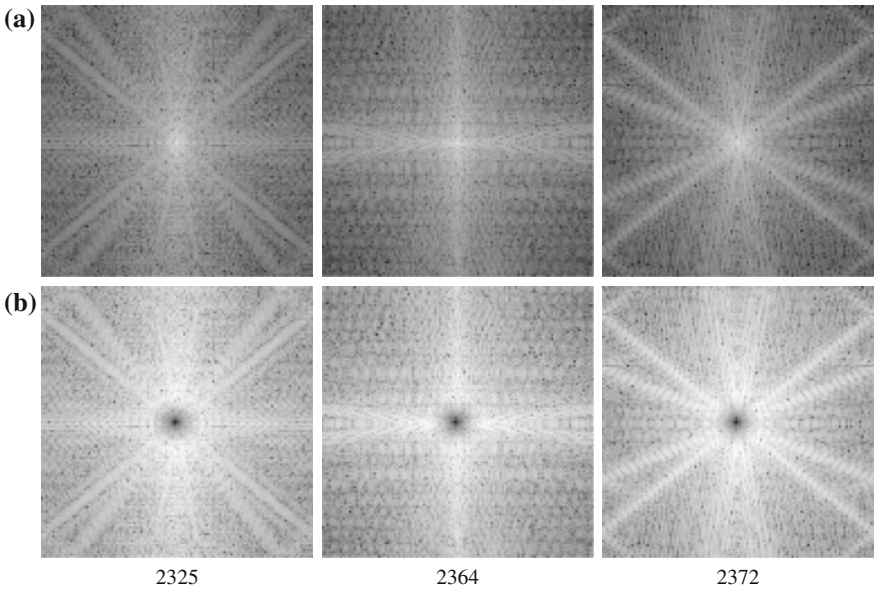


Fig. 4 The Fourier transform with the exchanged quarters (**a**) the Fourier transform of (**a**) with a solid component removed from transform (**b**)

Table 1 The correlation coefficients for objects with undeleted center of the transform

	2325	2364	2372
2325	1.0000	0.6811	0.6458
2364	0.6811	1.0000	0.5958
2372	0.6458	0.5958	1.0000

Table 2 The correlation coefficients for objects with deleted center of the transform

	2325	2364	2372
2325	1.0000	0.6352	0.5888
2364	0.6352	1.0000	0.5343
2372	0.5888	0.5343	1.0000

grayed images. Figure 5b shows the sum of modules of Hilbert transform applied separately on both axes. It is a very useful form of transformation of HTI, through which it is very easily to determine the boundaries of the object.

Anisotropic Hilbert Transformation

The second and also the most common in this paper (and studies) transformation is anisotropic Hilbert transformation (HTA) (the script below this page). Unlike HTI detecting the boundaries of an object, a HTA emphasis strong change in direction of the edge of the object. Arise then so-called characteristic points, that for each of the objects are different.

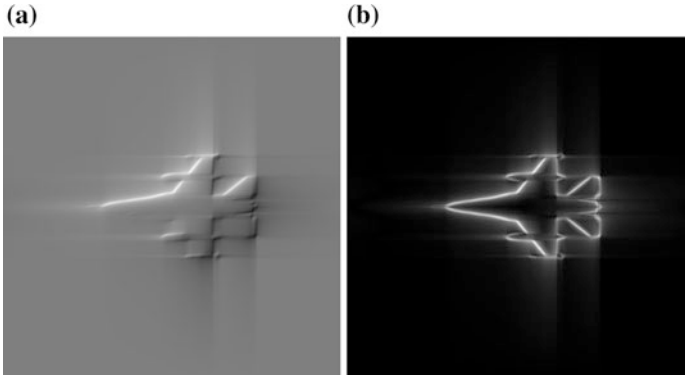


Fig. 5 Isotropic Hilbert transformation—HTI (**a**) and its module—|HTI| (**b**)

```
function [trans] = HTA(image)
original_im = im2double(image);
ox = imag(hilbert(original_im));
trans = imag(hilbert(ox'))';
The script executing the anisotropic Hilbert transform
```

As can be seen from the above script the HTA in the second step forms a Hilbert transformation is not from the primary image, as is the case of HTI, but the transposed form of a first Hilbert transform. Figure 6 illustrates a variant of action HTA.

HTA, like HTI, exposes despite to the characteristic points, their nature and the nature of the change of shades of gray in the original image. |HTA|—the module of HTA exposes locations of abrupt changes of the object, regardless of the direction of changes of gray levels on the original image.

The |HTA| transform is an excellent material for testing by characteristic points method (CPM), in which it is assumed that information about the characteristic

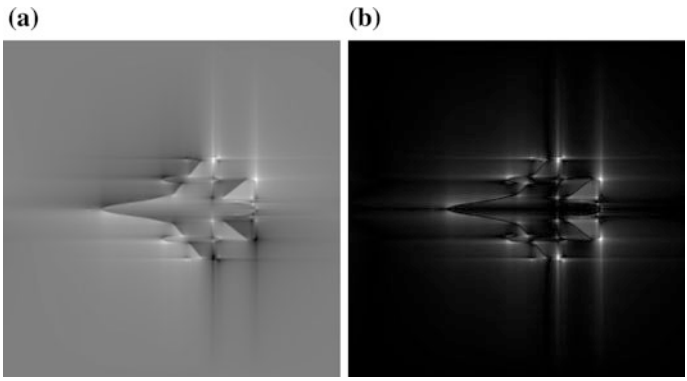


Fig. 6 Anisotropic Hilbert transformation—HTA (**a**) and its module—|HTA| (**b**)

points can be saved as a list of coordinates of centroids of individual points. The choice of the coordinate system used when creating this list also gives the opportunity to develop vectors with features of a very high capacity discriminatory. In addition, the use of relative measures in this feature vector eliminates the need to adjust the scale of the object to the feature vector.

Isotropic and Anisotropic Foucault Transformation

Another hybrid transformation from a family of Hilbert transformations is the isotropic Foucault transformation. This transformation, although it is called isotropic does not use the results of HTI or |HTI| but calculates the squares of primary Hilbert transformations on both axes. The main difference between isotropic Hilbert transformation and Foucault-a is that the HTI in addition to the image, resulting from the transformation also includes the original image.

Hybrid applied here gives only positive values due to the fact that all the operations which the HTI consists of operate on squares of values rather than their original values, the imaginary part separation using a Hilbert transform to the real form of the tested matrix.

As the isotropic Foucault transformation is similar to isotropic Hilbert transformation, the anisotropic Foucault transformation is similar to anisotropic Hilbert transformation.

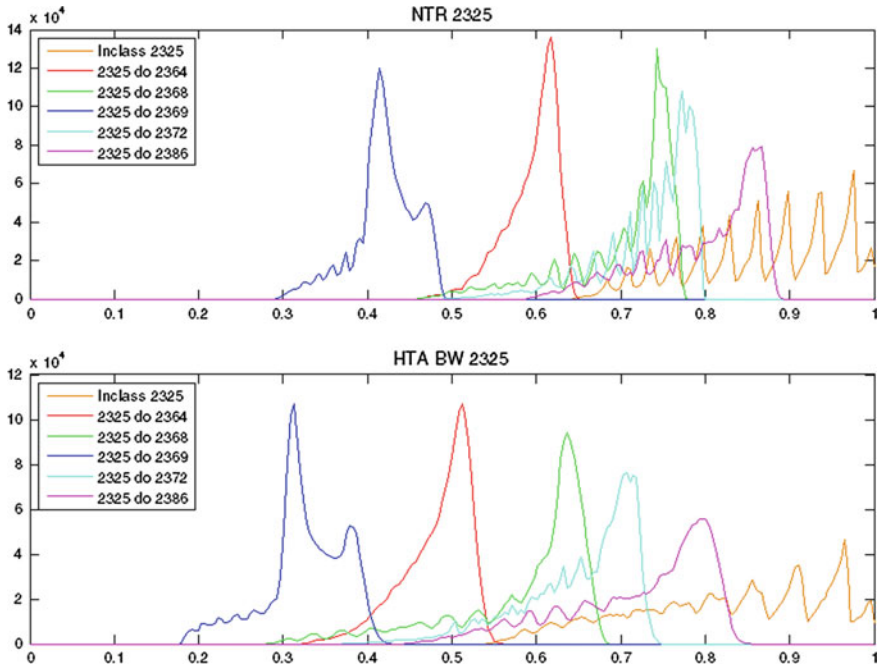


Fig. 7 Histograms of correlation coefficients of binary images of objects untransformed (NTR) and objects transformed by HTA—for objects observed from the side

In the definition of Foucault transformation there is however some difference: in the anisotropic transformation after second iteration from the matrix as a result of first transformation, the result is not squared.

This difference is due to the fact that the result of Hilbert transformation raised to its square according to this expression:

$$img = hilbert\left(hilbert(x)^{2'}\right)' \tag{4}$$

gives a very strong exposure highs in points of sudden changes in the shape of the object relative to the primary image that is contained by the Foucault transformation. This implies that a better solution is to simply convert the HTA, as it gives a result very similar test signals (transform).

3 Results

In the experiment the original (unprocessed) images and transformed images, described in Sect. 2.3 have been investigated as well. Each object has been rotated in three axes in a range of 0°–10° in steps of 1°, which gave for each plane 113

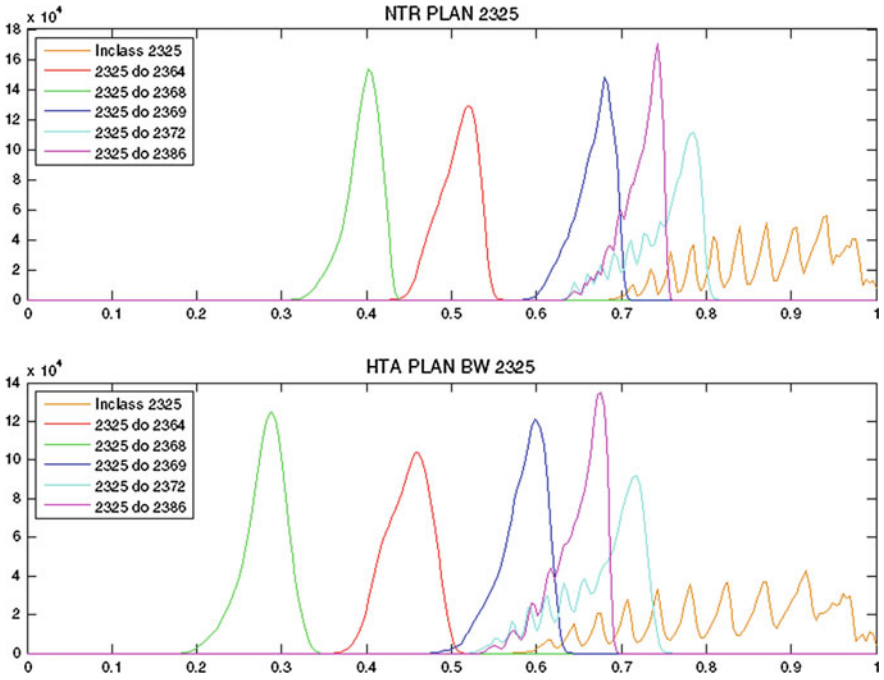


Fig. 8 Histograms of correlation coefficients of binary images of objects untransformed (NTR) and objects transformed by HTA—for objects observed from the top

(1331) of samples. All objects have been tested with correlation methods performed iteratively. As first the correlation properties of in-class (correlation with itself) objects' has been performed, then with the other five objects. The correlation coefficients were collected in the arrays with dimensions 1331 × 1331. Correlation matrices within class (correlation with itself) were the symmetrical matrices along diagonal axis. From those matrices the histograms have been computed, as described on Figs. 7, 8 and 9.

The research clearly shows a very high discrimination ability of HTA transformation, which in Figs. 7, 8 and 9 is presented as increasing the distance of histograms.

The above-described tests have been performed for three different classes of objects' images:

- binary images (BW) of objects as seen from the side (observation from the air),
- binary images (BW) of objects as seen from the top (satellite observation),
- grayscale images (GR) of objects as seen from the side (observation from the air).

For each histogram sets it has been computed the measures of distance between the histograms. These measures are expressed as Minkowski distance, described by the formula (Tables 3, 4 and 5):

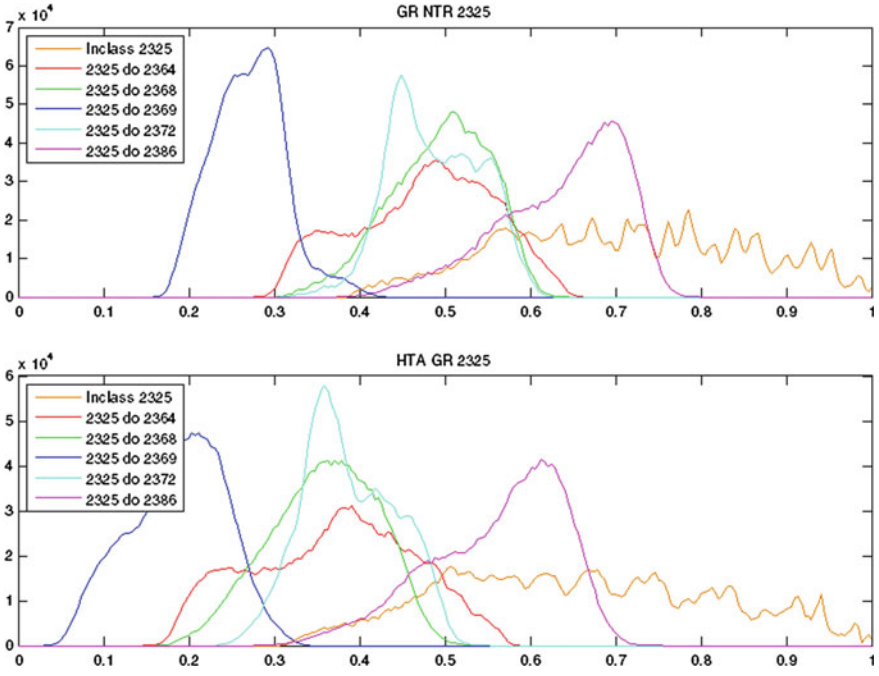


Fig. 9 Histograms of correlation coefficients of grayscale images of objects untransformed (NTR) and objects transformed by HTA—for objects observed from the side

Table 3 Statistics for binary (BW) images of objects as seen from the **side**

Feature	Objects' name				
	2364	2368	2369	2372	2368
Measure for distance for NTR	10.95	4.81	17.54	3.97	2.10
Identification error for NTR*	0.00	0.18	0.00	0.23	0.46
Measure for distance for HTA	10.81	5.30	15.32	4.21	2.39
Identification error for HTA*	0.00	0.13	0.00	0.24	0.49
Changing the distance	−0.14	0.49	−2.22	0.24	0.29
Δ of false identification [%]	0	−28	0	4	7

Table 4 Statistics for binary (BW) images of objects as seen from the **top**

Feature	Objects' name				
	2364	2368	2369	2372	2368
Measure for distance for NTR	15.39	20.94	9.51	4.97	6.74
Identification error for NTR*	0.00	0.00	0.00	0.19	0.07
Measure for distance for HTA	20.73	25.98	10.98	5.3	7.63
Identification error for HTA*	0.00	0.00	0.02	0.21	0.11
Changing the distance	5.34	5.04	1.47	0.33	0.89
Δ of false identification [%]	0	0	–	11	57

Table 5 Statistics for grayscale (GR) images of objects as seen from the **side**

Feature	Objects' name				
	2364	2368	2369	2372	2368
Measure for distance for NTR	8.16	7.76	17.01	8.68	2.10
Identification error for NTR*	0.30	0.25	0.00	0.25	0.60
Measure for distance for HTA	9.89	10.49	17.64	11.28	2.88
Identification error for HTA*	0.25	0.14	0.00	0.18	0.58
Changing the distance	1.73	2.73	0.63	2.60	0.78
Δ of false identification [%]	−17	−44	0	−28	−4

$$d_m = (h_1, h_2) = \frac{\left|\overline{h_1} - \overline{h_2}\right|}{s_1 + s_2} \tag{5}$$

where:

d_m Minkowski distance
h_x first (1) and second (2) histogram
s_x standard deviation of first (1) and second (2) histogram.

4 Conclusions

This paper presents a simple and hybrid Hilbert transformation that was then used to investigate their ability to discriminate when comparing the test object correlation with the pattern. Development of a set of proposed transformations was very laborious process, but their evaluation was necessary because of the virtually unlimited ability of hybridization of Hilbert transformation. Of all the known methods selected it has been selected four of the highest discriminative ability, and for comparison: method for many years used in the analysis of signals—Fast Fourier Transform with swapped quarters of spectrum and also primary, unprocessed images.

A further direction of research is the evaluation of selected methods for the collection of samples prepared for the purposes of an experiment. The study has been performed on images of objects from Fig. 3. In addition, it has been tested several signature transformations, such as Radon transformation that reduces a unique description of the object to a few percent relative to the original image with a slight loss of discriminative ability.

References

1. Al-Atabi, A., Al-Nuaimy, W.: Advanced signal processing techniques in NDT. In: Qahwaji, R., Green, R., Hines, E. (eds.) Applied Signal and Image Processing: Multidisciplinary Advancements, p. 350. IGI Global, Hershey (2011)
2. Bielecki, Z., et al.: Detekcja sygnałów optycznych. WNT, Warsaw (2001)
3. Borden, B.: Problems in airborne radar target recognition. Inverse Prob. **10**(5), 1009–1022 (1994)
4. Hahn, S.L.: Hilbert Transforms in Signal Processing. Artech House, USA (2009)
5. Oriot, H., Michel, A.: Building extraction from stereoscopic aerial images. Appl. Opt. **43**, 218–226 (2004)
6. Sudol, A., Stemplewski, S., Vlasenko, N., Vlasenko, V.: Hybrid transforms applications for objects and textures identification: MDHO based IIT comparative studies. ISAT, Wroclaw (2010)
7. Vlasenko, V., Sudol, A.: Digital Hilbert optics methodology for complex shape objects and textures at dynamic scenes identification: structure design, modeling and verification. Syst. Sci. **35**(3) (2009)

Light-Reflection Analysis Method for 3D Surface Damage Identification

Michał Turek and Dariusz Pałka

Abstract The article introduces a new high resolution 3D mesh comparison method that can be used for 3D object surface analysis. A need of 3D mesh analysis is evident in many procedures in engineering, medical or strictly graphical applications. A typical 3D object scanning process produces a high resolution 3D triangle mesh describing the surface of the object. Surface analysis based on this kind of material is usually complex or inaccurate, because each 3D mesh vertex must be identified, positioned and analyzed. A solution proposed in the paper is focused on using native 3D mesh rendering processes for mesh analysis, especially in the surface damage identification field. 3D graphical acceleration hardware and Pixel/Vertex Shaders technology will be used to prepare sets of 2D images—generated with natively 3D accelerated but specially modified light reflection rendering technique. Images then will be analyzed by comparing their 2D reflections with correct object patterns to find any damage-caused differences. The method has proven very quick to calculate and easy to apply; the test applications were programmed over a standard PC 3D accelerated graphical modules. It can also be flexibly applied, which allows for analyzing only a part of the 3D object surface if needed. Additionally, it can produce very accurate results without any precision lowering mathematical-model assumptions, commonly met and usually necessary to apply in typical 3D triangle mesh analysis.

Keywords Surface damage identification · 3D mesh analysis · Light reflection rendering

M. Turek (✉) · D. Pałka
AGH University of Science and Technology, 30 Mickiewicza Av., 30-059 Krakow, Poland
e-mail: mitu@agh.edu.pl

D. Pałka
e-mail: dpalka@agh.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_7

1 Introduction

Surface analysis is a widely known problem in computer graphics applications. Nowadays highly developed 3D scanning methods can produce advanced multi-point 3D mesh structures, which should be analyzed and compared with patterns. A vast number of triangles and vertices which need to be processed in 3D scanning can cause huge computation complexity problems. Direct 3D triangle mesh calculations are very complex here, so other methods of analysis need to be developed. And if there is a ready comparison pattern also encoded into 3D triangle mesh, much more calculation is needed to perform this process [1, 2]. The concept of work described in the paper assumes the use of a "native graphical 3D object presentation" as a mid-result passed between two phases of surface analysis:

- Mathematically ideal 3D rendering based on native "light reflection like" processing (3D mesh is rendered into 2D image with light-reflections processing that highlights all surface imperfections)
- Domain-dependent 2D image processing (either a cross-section of an image or full 2D image analysis) is used to calculate surface imperfection metric values and judge their fitness.

The approach will enable us to perform a quick object damage analysis with easy precision shifting (zoom-like). It is focused on vast object surface anomaly patterns recognition, not per-vertex calculations. Comparison patterns can be expressed with any method—using 2D pixel-map imaging (direct image comparison) or functional relations (2D image cross-sections as a mid-result will be generated in this particular approach). The following basic assumptions are established:

- Medical applications for the method should be considered a priority, especially CT data analysis (for example, endoprosthesis).
- An entry point for the analysis process will be 3D triangle mesh representing the examined 3D object (created using commonly known methods [3]).
- Mesh resolution or triangle positioning cannot affect the final result as long as it represents a 3D object.

The great advantage of the proposed approach is the use of GPU-accelerated native rendering procedures to blend-in any mesh-specific data (too many triangles, different triangles passing in the mesh describing the same object, etc.)

2 3D Rendering Method

There was a need to develop a 3D rendering engine, which can be used to produce a light-reflective mesh image. The first experiments were performed using a standard OpenGL 3.0 smoothing features (for generating blended per-triangle light-reflection images). However, these attempts gave poor results, and there was no way to

control and customize the rendering process, having every pixel value (of the generated image) calculated correctly. Therefore, a full light-reflection generation process have been developed, based on a 3D Vertex Shader and Pixel Shader collaboration. Vectors called mesh vertex-normals are a base for any calculations here. They are assigned to each vertex in a mesh. To calculate a vertex normal, it is necessary, first, to have a vector perpendicular for each triangle adjacent to that vertex, then, to calculate average values of those perpendicular vectors for their x, y and z components, and finally, to combine them into a new vector, that is a mesh vertex-normal.

Mesh vertex-normals are commonly generated as an addition to a 3D model expressed with a 3D triangle mesh set and will be required by shader transformations. These vectors will be transferred together with vertex position vectors into shader programs as a non-uniform values (via a vertex profile). Inside a shader they will be used to compute the values of so-called vertex light reflections. Vertex light reflection is a vector product of a light direction vector and a vertex normal vector. After the vertex light-reflection computation, it is passed to Pixel Shader via a shader profile. Before it gets to the pixel processing engine, it is interpolated between light-reflection values of two other triangle vertices. Finally, a correct pixel-reflection value is passed to pixel Shader and then further to output 2D image. Any surface (caused by a vertex position shift) will now be exposed. The procedure of such rendering will be extremely fast, because it is natively accelerated by GPU hardware. A test implementation was made in nVidia CG (C for Graphics) shaders (over shader profile 2.0) [4, 5]. The communication between Pixel Shader (in a CG API called—Fragment Shader) and Vertex Shader has been established using extended texture coordinates data passed via a vertex profile [6]. This is a common approach in such cases. Typical implementation of per-pixel shading procedure is based on cooperation between vertex shading and pixel shading programs. Light-reflection vector (calculated in Vertex Shader with material and light features consideration) is passed to Pixel Shader via shader environment profile. During this passing, vector interpolation is calculated each time for particular 2D pixel. Since a "light source" in the proposed method is fixed and single, there is a chance to simplify light-reflection calculations. The goal is to detect surface damage, shown as a normal vector values disturbance spread in the vast number of vertices. A typical per-pixel surface shading technique (with a multi-light reflection color emulation involved) is not needed here. Instead, a vital result needed is just a shaded damage graphical raster. Therefore, a vertex shader procedure can be altered to generate interpolated light reflections based on vertex normals only. Now, a final shaded pixel color for rendered 3D object will express a 3D surface disturbance (possible damage), and not a surface color adjusted by lighting configuration.

Going into details, a vertex normal vector is being normalized first to have all components fixed in [0…1] range. Next, since it is a 3D vector, its x, y and z components are mapped into a new vector inside outbound shader profile (a vertex texture coordinate field was used in this profile). In vertex shader/pixel shader collaboration an additional vector (texture coordinate) will be interpolated for each pixel and passed into a pixel shader. Now, a pixel shader comes to action. It

performs a final pixel colorization using a former vertex normal vector (interpolated) instead of typical pixel shading colorization input values (natural light reflection for a particular surface material color described by diffuse, ambient and specular values). As a result, a precise graphical view of picture deformation is created, with adjustable precision level (zoom) and geometric damage highlights expressed in pixel colors. The solution is also quick, putting lots of calculations in a shader layer and making explicit vertex-set calculation unnecessary.

3 An Example of Application

The usefulness of the method presented above can be shown by the example from the domain of medical images. The example focuses on detecting surface deformations of the mandibular condyle. The analysis of damages of this surface is an important factor in the diagnosis of temporomandibular joint syndrome (in short TMJ syndrome) [7]. This surface can be analyzed on the basis of cross-sections of a 3D bone model obtained from CT pictures. However, the interpretation of these cross-sections, as is the case with the interpretation of other medical images made by humans (even if they are highly-qualified physicians), can be quite difficult, as it was pointed out in [8]. This is a direct result of both vast number of images (cross-sections) to be interpreted and the level of their complexity.

So, this example uses an alternative approach to the problem of detecting damages on the surface of the mandibular condyle. As stated above, the first stage of the presented method is the preparation of a 3D triangle mesh representing a 3D object (in this case a mandibular condyle). It is obtained by the following steps:

- CT Data Acquisition. The data used for the analysis of degenerative changes on the bone surface of the mandibular joint are obtained from a CT scanner with voxel dimension $0.4 \times 0.4 \times 0.4$ mm.
- Surface 3D Construction. Marching cubes algorithm [3] is used to create a polygonal representation of an ISO surface of a 3D scalar field from a CT image set. As a result, a 3D model made up of triangle facets is obtained.

A diagram presenting the process described above is presented in Fig. 1.

Figure 2 shows the surface of a 3D model created in the "Surface 3D Construction" step on the basis of the CT data using marching cubes algorithm. Next, the presented method of damage identification can be used for semi-automatic detection of the surface of the mandibular condyle deformation.

The first way in which the presented method can be used is to semi automatically detect the mandibular condyle deformation, which can indicate TMJ syndrome.

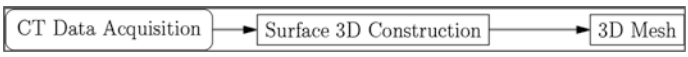


Fig. 1 Key points in the process of the analysis of surface damages

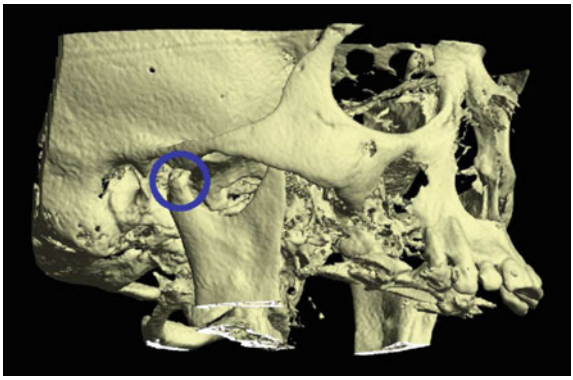


Fig. 2 The 3D model created using the marching cubes algorithm. The *marker* shows the mandibular condyle

In this case, the process consists of the following steps:

- The threshold of the permissible surface deformation level has to be set. This should be done by the domain expert (for example, a physician) by selecting the referential case of the mandibular condyle deformation. Then, every examined case with the surface deformation above this threshold will be treated as a potential cause of TMJ syndrome. After selecting the referential case, the zoom is used to set the referential distance from the surface to the observer for which the vertex-normal vector field is constant (which is represented by the same color of rendered pixels) due to average normal vectors observed from a given distance.
- For each examined case (the surface of the mandibular condyle), the comparison with referential case is performed. The color representation of the vertex-normal vector field is rendered from the referential distance. If the rendered surface color representation is the same as in the referential case (i.e. the vector field seen from the referential distance is constant), the mandibular condyle surface is below the threshold and is not considered the cause of TMJ syndrome. If, however, the rendered color representation of the vertex-normal vector field differs from the referential one, the surface deformations are above the threshold and can be potential causes of TMJ syndrome.

The difference in the color representations of the vertex-normal vector fields for referential and examined cases can be automatically detected using a standard 2D image comparison methods.

Figure 3 presents the mandibular condyle surface rendered with the use of a standard OpenGL procedure and the color representation of the vertex-normal vector field.

Figure 4 presents the color representation of the vertex-normal vector field for three sample distances from the surface to the observer. The distance value unit is the number of pixels from the camera to the nearest surface element.

The second way of using the presented method for detecting deformations of the object surface is the comparison of two CT images of the same object registered in

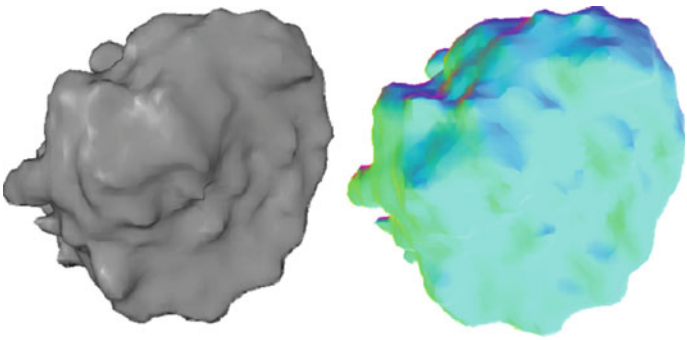


Fig. 3 The mandibular condyle surface. The image on the *left* is rendered with the use of a standard OpenGL procedure, while the image on the *right* is color representation of the vertex-normal vector field

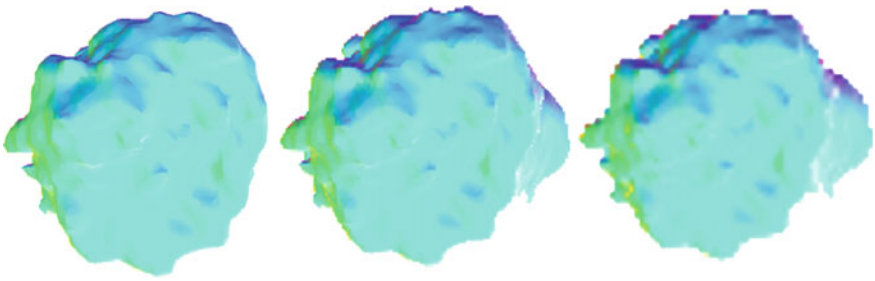


Fig. 4 The color representation of the vertex-normal vector field rendered for three distances—from *left*: distance 50, distance 250, distance 550

two different points in time (for example, the ones registered before and after the treatment). In such case, the comparative analysis of the surface can be provided. The process is as follows:

- The viewpoints for both images should be fitted (relative to the examined surface)
- The color representation of the vertex-normal vector field is rendered using the method presented in this paper
- The difference between the vertex-normal vector fields obtained is calculated. This can be done by a simple subtraction of two images with the color representation of the normal vector fields.

Figure 5 presents the examples of the color representation of the normal vector field of the mandibular condyle surface: original and deformed. The viewpoints in the both images are fitted.

Figure 6 demonstrates the comparison of the original and the deformed mandibular condyle. This figure is obtained as a result of subtracting color representations for normal vector fields. The pixel resultant color in the RGB model is

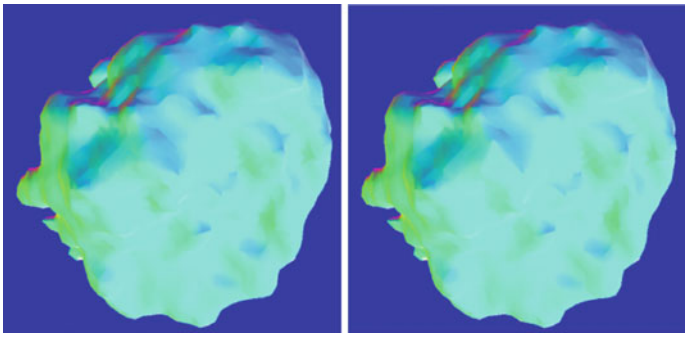


Fig. 5 The color representation of the vertex-normal vector field for the mandibular condyle: *left image*—original, *right image*—deformed

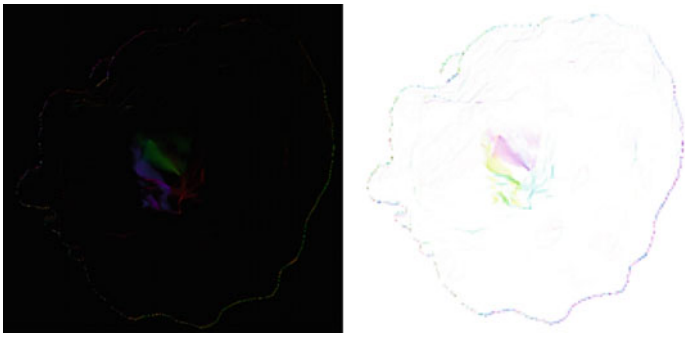


Fig. 6 The result of the subtraction of the vertex-normal vector field for the original and the deformed mandibular condyle *left image*—subtraction, *right image* inversed colors of subtraction

calculated as an absolute value of the difference between red, green and blue components of pixel colors from the left image of Fig. 5 and from the right image of Fig. 5. For better visualization, the color results of subtraction have been normalized.

As can be seen in Fig. 6, the presented method allows for quick visual comparisons of two surfaces (for example, an original and a damaged one). This approach can be useful, for example, for a quick evaluation of treatment results for TMJ syndrome.

4 Conclusions and Future Work

In recent years surface processing and analysis have been playing a more and more important role in a wide range of computer graphics applications. This is caused by both the popularization of 3D scanners and 3D printers and by the growing

popularity of medical 3D scanner devices (such as a CT and NMRI scanners). As a result, the demand for different types of analyses of multipoint 3D mesh structures, including surface damage identifications (especially in medical applications), increases.

In this paper the solution based on light reflections used to identify damages of a surface was presented. The proposed solution has one major advantage: it can be easily implemented with the use of typical and cheap computer hardware (e.g. a graphically-accelerated PC workstation, etc.). Vertex Shader-based 3D mesh processing and the proposed mathematically ideal rendering method offer a really fast and reliable surface damage comparison tool. It can be used in a real-time mode, processing complex 3D mesh objects taken directly from a 3D scanner or other similar devices.

On the other hand, it is important to assume that any precise 3D object surface analysis process would require multiple rendering passes—with zoom, angle or exposure shifts between them. In this case a 2D data will be harvested and passed to the external module for further analysis many times. Therefore, it is important to provide a process that will be able to generate results with extremely high speed—processing even hundreds of mesh "frames" per second (and subsequently producing results in real time). Thanks to 3D accelerated graphical platforms with vertex shading technology, it can be done natively. 2D images obtained will probably be compared with patterns in some domain-specific ways. For instance, comparative cross-sections from these images could be extracted. A comparison process itself can also be iterative, where each new image is rendered after a slight angle, zoom or exposure shift. This technique will be similar to a classical 3D rotary laser surface scanning [9], where a laser-head takes a 3D object reading with fixed angular resolution. Factory-ready procedures for such analyses can be developed in further work. For now we can be sure that a quick 3D to 2D rendering method suitable for such jobs is available—with extensive options opened.

References

1. Hedin, F., Klasén, K.: Terrestrial laser scanning: an investigation of 3D CAD model accuracy. Master's thesis, Royal Institute of Technology (KTH) (2003)
2. Smith, C.: On vertex-vertex systems and their use in geometric and biological modelling. Ph.D. thesis, Calgary, Alta., Canada (2006)
3. Lorensen, W.E., Cline, H.E.: Marching cubes: a high resolution 3D surface construction algorithm. In: Stone, M.C. (ed.) SIGGRAPH. pp. 163–169. ACM (1987). doi:10.1145/37401.37422. http://dblp.uni-trier.de/db/conf/siggraph/siggraph1987.html#LorensenC87. http://www.bibsonomy.org/bibtex/2d653a88c0bcd04018dbf0d67c1aa2e37/dblp. Light-reaction analysis method for 3D surface damage identification
4. NVidia Vertex Shaders. http://www.nvidia.com/object/feature_vertexshader.html (2015)
5. Nvidia: Cg Toolkit Release Notes. NVIDIA Corporation (May 2012)
6. Crow, F.C.: Shadow algorithms for computer graphics. In: George, J. (ed.) SIGGRAPH, pp. 242–248. ACM (1977). http://dblp.uni-trier.de/db/conf/siggraph/siggraph1977.html#Crow77

7. Kaplan, A., Assael, L.: Temporomandibular disorders: diagnosis and treatment. W.B. Saunders (1991). https://books.google.pl/books?id=VvNpAAAAMAAJ
8. Tadeusiewicz, R., Ogiela, M.R.: Medical image understanding technology—artificial intelligence and soft-computing for image understanding. In: Studies in Fuzziness and Soft Computing, vol. 156. Springer (2004). doi:10.1007/978-3-540-40997-7
9. ShapeGrabber. http://www.shapegrabber.com/sol-products-3d-auto-inspection-Ai310.shtml (2015)

Part III
Operation Research Applications

Applications of Operations Research and Intelligent Techniques in the Health Systems

A Bibliometric Analysis

Marek Lubicz

Abstract Operations research and quantitative methods, including the ones based on artificial or computational intelligence, have been applied in health care for decades. While looking at areas of applications and evolution of techniques one can notice trends, emerging problem areas, and also international disparities in the application-oriented research. Due to the scale of the problem, in this paper we explore these ideas by means of analysis of a limited, though comprehensive sample of the research results presented in the period 1985–2015 at annual conferences of the EURO Working Group Operational Research Applied to Health Services. The technical background for this research is a bibliometric analysis based on SciMAT—a Science Mapping Analysis software Tool, developed at the University of Granada.

Keywords Operations research · Analytics · Healthcare · Bibliometric analysis

1 Introduction

As observed in many publications (e.g. [1, 2]), the academic literature on applying operations research and other quantitative methods is vast and growing at a very high rate.[1] On the other hand the boundaries are very unclear between scientific disciplines using modelling for supporting solving decision problems, be them managerial, economic or clinical, arising at different levels of healthcare system. In an important monograph on modelling in healthcare [3] one can hardly find any single mention on the Operations Research (OR) domain in the explanation of either analytical approaches (specified as: Computer simulation, Mathematical

[1]Brailsford et al. [1] estimate the magnitude of the academic literature on healthcare modelling for 200.000 in 2007, with an expansion rate of about 300 papers per day.

M. Lubicz (✉)
Wroclaw University of Technology, Wrocław, Poland
e-mail: marek.lubicz@pwr.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_8

analysis, Numerical analysis) or modelling techniques (classified as population, behavioural, global models and operational models). In some publications, simulation, queueing or Markov models are juxtaposed with operations research models, understood as mathematical programming (MP). In contemporary academic textbooks (e.g. [4]), traditional Management Science (MS) techniques (simulation, MP) or modern Data Science techniques (Decision Trees, Neural Nets) are labelled Prescriptive or Predictive Analytics respectively, while OR is classified into Traditional Analytics domain, together with Business Intelligence and Data Mining. Also at official INFORMS web page [5], it is suggested that OR, MS, and Analytics are used as synonyms, which is further exemplified when looking at the scientific program of INFORMS Healthcare Analytics conferences, dedicated to "healthcare OR and analytics," while covering extremely wide spectrum of modelling approaches.

In the past few decades, many review papers have been written on healthcare modelling. The reviews have generally focused either on a specific modelling approach (e.g. Discrete Event Simulation DES, Fone et al. [6]), the use of modelling for a specific healthcare setting (e.g. clinics, Jun et al. [7]), or the use of a specific pre-selection methodology (e.g. stratified sampling, Brailsford et al. [1]). In most cases the reviewers followed a standard methodology of querying a recognized scientific bibliographic database (most often: Web of Science or Scopus), defining inclusion-exclusion rules, and performing manual analysis of the articles selected. However, in bibliometric reviews (see a discussion in [8]) a large number of articles analysed (occasionally over 100.000) precludes or at least significantly hinders a thorough analysis of full texts of selected articles.

In this paper, we aim to review the current state of and trends in healthcare modelling research, not restricted to particular approaches or healthcare problems, starting from a comprehensive selection of healthcare modelling papers presented in the period 1985–2015 at annual conferences of the EURO Working Group Operational Research Applied to Health Services (ORAHS) or published in the proceedings following such conferences. The total number of papers presented at ORAHS conferences since 1975 until 2015 is 1744, however due to the lack of the abstracts or full texts of many papers from the years 1975–1996, only 1363 papers have been included in this review. We also plan to confront the results of such a review with the findings of Brailsford and Vissers [2], who analysed a selection of 233 ORAHS papers from the period 1975–2008, and contrasted these papers with 342 papers analysed in a broad review of the research literature on healthcare modelling [1] in relation to both healthcare issues considered and modelling techniques applied.

Throughout the review we apply a novel multistage methodology of bibliographic analysis that comprises defining a unified hierarchical taxonomy of keywords used for bibliographic, and further also for bibliometric analysis. In our view, it is a necessary step when analysing documents from any domain, especially

published during a long period, and when the diversity of source keywords describing documents is considerable (or when there are no authors' keywords for a substantial proportion of documents, which was the case of ORAHS papers).

2 Methodology of the Analysis

A query to identify a core set of healthcare modelling academic articles was defined in [1] as follows: "(health-care OR health care) AND (modelling OR modeling OR simulat* OR (system AND dynamic*) OR markov*)." Having filling in the natural phrases or add-ins (health?care, systems?dynamic, agent?based?model, DES?model, where ? stands for any hyphen) and applying the query to the titles, abstracts and keywords of the set of 1363 ORAHS papers we realize that only 17 % would be selected (in addition only 33 % satisfied healthcare part and 49 % satisfied the modelling part). It justifies critical importance of homogeneity of terminology used in bibliographic analyses, particularly in relation to keywords. Other problems we are faced with consist in the diversity of terminology used by authors to denote similar objects, units or processes; a considerable number (almost 44 %) of papers with no authors keywords defined and also a number of important language-related ambiguities (e.g. "operational research" as synonym of field research or "system dynamics" in the sense: dynamics of a system).

To overcome some of the terminology related problems, a general structured taxonomy of the domain keywords has been developed in two language versions: English and Polish. The taxonomy is comprised of the following notions (a hierarchy from more detailed to general notions):

- WORD (most specific notions, used in articles)
- WORDGROUP (clusters of similar WORDS, also taking into account variants, inflection, punctuation, etc.)
- GENERAL and VERY GENERAL WORDGROUP
- Problem LIST (examples: ANALYTICS, DEA, ECONOMETRIC_MODEL, HEURISTICS, MARKOVIAN, QUEUEING, SIMULATION, STATISTICAL)
- Problem category (examples: OR, economics, general management, health care management, medical, modelling, intelligent techniques)

Tables 1 and 2 present respectively: sample items from all 10 problem categories for all levels of hierarchy and a part of third level keywords (General Word Groups) set for a class of heuristic optimization approaches in both languages. The first column to the right in Table 2 contains variant search terms for a specific keyword, determined in the process of semi-automated validation of the taxonomy, using keywords sets of three main academic journals in the healthcare modelling domain, namely Health Care Management Science, Operation Research for Health Care, and Health Systems, as well as healthcare modelling keywords from the articles

Table 1 Taxonomy of healthcare modelling keywords (sample items for each category)

Category	List	GeneralWordGroup	WORD-GROUPS=wG
OR	ORAHS	Patient flow modelling	PATIENT-FLOW-MODELLING
OR	SIMULATION	Agent simulation	AGENT-SIMULATION
Intelligent technique	INTELLIGENT_MAIN	Expert systems	EXPERT-SYSTEM
Intelligent technique	ANALYTICS	Prescriptive analytics	PRESCRIPTIVE-MODEL
Bio model	CLINICAL_DSS	Clinical DSS	CLINICAL-DSS
Bio model	EPIDEMIC_MODEL	HIV/AIDS model	HIV-AIDS-POLICY-MODEL
Economic model	ECONOMICEFFICIENCY	Economic efficiency	TECHNICAL-EFFICIENCY
Economic model	DEA	DEA cross efficiency	CROSS-EFFICIENCY-DEA
Modeling	MODELLING	Generic model	GENERIC-MODEL
Modeling	PROBABILISTIC	Phase type	PHASE-TYPE-DISTRIBUTION
Medical	CANCER	Cancer	CANCER
Medical	SURGERY	Vascular surgery	VASCULAR-SURGERY
Health care	AMBULATORY_CARE	Ambulatory care	OUTPATIENT-CARE
Health care	OTHER_HEALTHCARE	Walk in	WALK-IN-CENTRE
HC mgmt	HEALTHMANAGEMENT	HC facilities	HEALTH-CARE-FACILITY
HC mgmt	HC_PATHWAYS	HC pathway	PATIENT-PATHWAY
General mgmt	PERSONNEL	Personnel management	PERSONNEL-ALLOCATION
General mgmt	NEED_DEMAND	Inequalities	EQUITY-OF-ACCESS
General	GEOCOUNTRY	Poland	LOWER-SILESIA
General	POPULATION	Population	POPULATION-MODEL

Table 2 Sample hierarchy of keywords for heuristic approaches

GeneralWordGroup	WordGroup	WordGroupPL	WORDS=search terms
metaheuristic	meta heuristic	metaheurystyki	meta?heuristic
metaheuristic	meta heuristic	metaheurystyki	metaheuristic
hyperheuristics	hyper heuristic	hiperheurystyki	hyper?heuristic
hyperheuristics	hyper heuristic	hiperheurystyki	hyperheuristic
direct search	direct search	przeszukiwanie bezpośrednie	direct?search
direct search	direct search	przeszukiwanie bezpośrednie	directed?choice
direct search	scatter search	przeszukiwanie rozproszone	scatter?search
local search	greedy heuristic	algorytm zachłanny	greedy?algorithm
local search	greedy heuristic	algorytm zachłanny	greedy?heuristic
local search	greedy heuristic	algorytm zachłanny	greedy?search
local search	guided search	przeszukiwanie ukierunkowane	guided?search
local search	local search	lokalne przeszukiwanie	iterated?search
local search	local search	lokalne przeszukiwanie	local?search
local search	tabu search	przeszukiwanie Tabu	tabu?search

published in the European Journal of Operational Research in the period 1977–2014. The final taxonomy contains 2744 lowest level keywords (WORDGROUPS) and 7465 variant search terms. The set of search terms was constructed according to a hierarchy of keywords expressions lists (example: list1: mixed-integer-linear-programming; list2: integer-linear-programming; list3: linear programming) implemented in a software tool for automatic analysis of the title, authors keywords (if any) and abstract of each paper. The tool incorporates a module for preliminary conversion of keywords database using regular expressions. Outputs of the automatic bibliographic analysis, which assigns keywords to papers (in contrast to [2] multiclass keywords assignment is performed and papers are not clustered into disjunctive categories), comprise: a list of unified keywords assigned to each paper and divided into problem lists (Table 3) and a table report on articles to keyword assignment and distribution (for each keyword: identifiers and total number of papers published in each period) (Table 4).

The next step of the literature analysis consists in bibliometric analysis, i.e. automatic analysis of the documents, using only new unified keywords for each paper. We have applied a science mapping type of bibliometric analysis [8–11],

Table 3 Sample output of the bibliographic analysis: keywords to article assignment

idorahs	OR_PROBLEMS	MATHPROGRAMMING	SIMULATION	EMERGENCY_CARE	HEALTH_PERSONNEL	HOSPITAL
w928		MP integer			Physician	Hospital
w43		Optimization	Simulation	Emergency EMS		
w1235	Scheduling problem				Nurse planning	Hospital
w708	Knapsack model	MP dynamic		ED management		Surgery planning, surgery room, hospital admission

Table 4 Sample output of bibliographic analysis: articles to keyword assignment (category: OR, list: HEURISTICS)

General word group	ORAHS articles		1985–1990	1991–1995	1996–2000	2001–2005	2006–2010	2011–2015
Local search	32	[w707], [w49], [w556], [w803], [w318], [w438], [w597], [w640], [w926], [w622], [w938], [w1131], [w58], [w486], [w641], [w51], [w584], [w1169], [w270], [w642], [w1132], [w1211], [w1248], [w59], [w333], [w1051], [w974], [w323], [w577], [w1228], [w709], [w884]	0	0	0	1	14	17
Meta heuristic	17	[w87], [w121], [w556], [w555], [w100], [w289], [w304], [w486], [w641], [w682], [w51], [w32], [w792], [w52], [w1133], [w115], [w1228]	0	0	0	0	10	7

based on co-word analysis combined with standard performance analysis, based on the number of documents (due to the specificity of the sources no citation information was available). The science mapping part is used to analyse the problem domain to detect and visualize its conceptual subdomains (particular themes or general thematic areas), relationships between themes, and the domain thematic evolution [10].

To perform the analysis we used SciMAT—a science mapping analysis tool, developed at University of Granada [9, 12]. The software combines both performance analysis tools and science mapping tools to analyse the problem domain in a longitudinal framework. A bibliometric analysis with SciMAT includes four stages [10]:

- Detection of research themes. In each period of time examined the corresponding research themes are detected by applying a co-word analysis
- Visualization of research themes using strategic diagrams and thematic network
- Discovery of dynamics of thematic areas and visualization of evolution of research themes over a set of periods (Fig. 1)
- Performance analysis, i.e. relative contribution of research themes and thematic areas to the whole research field with use of several bibliometric indicators: number of documents published, number of citations and different types of h-index

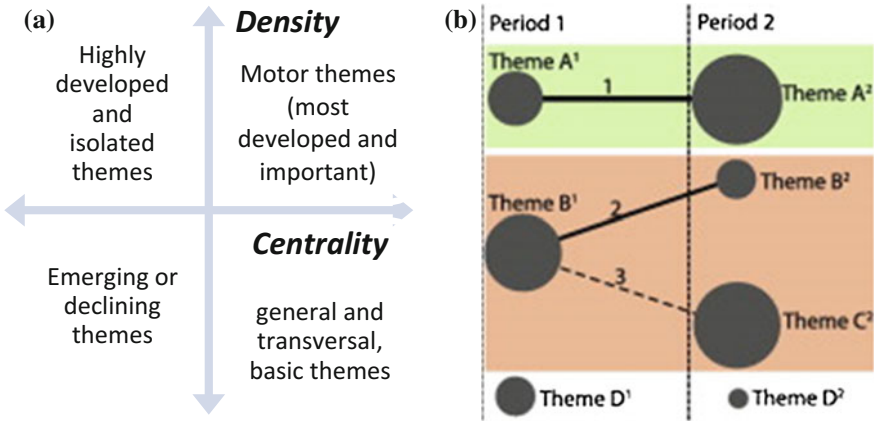


Fig. 1 Main outputs of SciMAT-based bibliometric analysis: **a** strategic diagram (*Source* based on [11]), **b** thematic evolution map (*Source* [10] with author's permission)

Before performing a SciMAT-based analysis, a secondary bibliographic database was developed using the keywords defined during the stage of bibliographic analysis. The database was validated using SciMAT tools for knowledge base cleansing.

3 Areas and Trends in ORAHS Research

The general bibliographic analysis comprised 1363 ORAHS papers presented at annual conferences or published in ORAHS proceedings in the period 1985–2015. Due to the lack of detailed information, papers from the years 1986–1989, 1991–1994, 1996 were excluded. Distribution of the papers by years and geography (countries defined for the first author) is presented on the Fig. 2.

Selected results of the bibliographic analysis of ORAHS papers are presented in Tables 5, 6, 7 and 8 for specific problem sub-domains (Fig. 3):

- general (Table 5) and more specific (discrete optimization—Table 6a), simulation modelling—Table 6b) modelling approaches,
- specific health services (Table 7a) or clinical issues (Table 7b),
- general keywords dealing with health management (Table 8a) or general management (Table 8b) subdomains.

Moreover, Fig. 4 illustrates in a more visible format relative frequencies of using detailed keywords concerning discrete optimization. We can conclude that

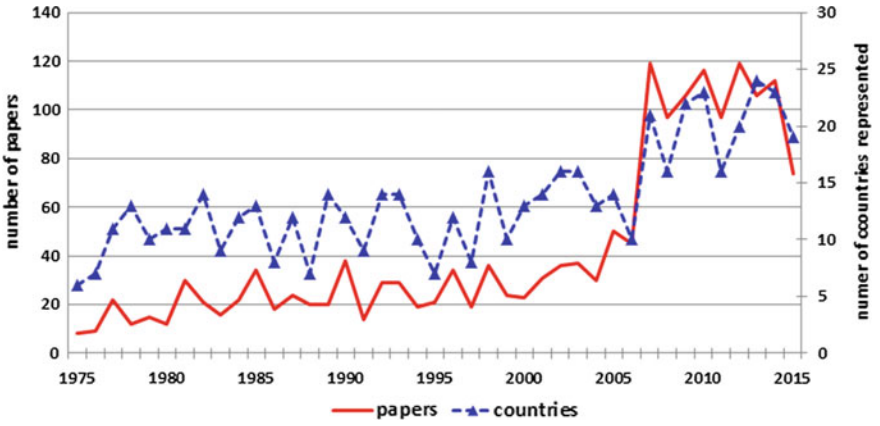


Fig. 2 Numbers of papers and countries represented at ORAHS conferences 1975–2015

Table 5 Number of ORAHS papers in which particular modelling approaches had been used

Category	Generalized keyword (VGwG)	Total	1985–1990	1991–1995	1996–2000	2001–2005	2006–2010	2011–2015
OR	Simulation	460	13	6	22	69	169	181
OR	Mathem. programming	202	4	2	7	16	80	93
OR	Optimization	177	5	2	3	8	76	83
OR	Heuristics	85	1	0	1	4	39	40
OR	Queueing model	83	0	2	1	6	39	35
OR	Location allocation	69	1	1	1	2	22	42
OR	Markovian model	69	2	0	4	6	28	29
OR	Scheduling problem	68	0	0	0	8	25	35
OR	Other OR problems	55	0	0	1	0	20	34
OR	Networks graphs	32	1	0	0	2	14	15
EM	DEA	25	0	0	1	8	9	7
OR	Multi objective	24	0	1	2	3	5	13
TI	Approximate reasoning	21	1	0	2	0	10	8
OR	Genetic algorithm	18	0	0	1	0	7	10
OR	Nature inspired other	14	0	0	0	2	5	7
OR	Transportation problem	7	0	0	0	1	3	3
OR	Neural network	6	0	0	2	1	3	0

Table 6 ORAHS papers dealing with particular detailed modelling approaches (selection)

(a) Keywords lists: MATHPROGRAM., NATURE_INSPIRED, HEURISTICS			(b) Keywords list: SIMULATION		
GeneralWordGroup	Number	%	GeneralWordGroup	Number	%
Optimization	84	6	Simulation DES	155	11
Integer programming	73	5	Simulation	149	11
Optimization method	64	5	Simulation model	142	10
Heuristics	53	4	System dynamics	70	5
Mixed-integer programming	40	3	Simulation modelling	49	4
Local search	32	2	Simulation study	31	2
Networks graphs	32	2	Computer simulation	25	2
Linear programming	28	2	Simulation approach	23	2
Mathematical programming	28	2	Monte Carlo simulation	17	1
Combinatorial optimization	27	2	Visual simulation	14	1
Stochastic programming	26	2	Simulation optimization	14	1
Optimization problem	21	2	Agent-based simulation	12	1
Metaheuristic	17	1	Simulation application	11	1
Dynamic programming	15	1	Simulation games	7	1
MP approaches	15	1	Simulation software	7	1

Table 7 ORAHS papers dealing with specific healthcare and medical issues (selection)

(a) Keywords lists: AMBULATORY_CARE, EMERGENCY_CARE, HOME_CARE, HOSPITAL, LONG_TERM_CARE, PRIMARY_CARE			(b) Keywords for particular medical issues, services and clinical specialties		
Generalized Keyword	Number	%	Generalized Keyword	Number	%
Hospital management	386	28	Surgery room	148	11
Hospital planning	241	18	Disease	116	9
Hospital type	192	14	Surgery planning	90	7
Hospital emergency department	172	13	Infectious disease	76	6
Ambulatory care	156	11	Surgery	66	5
Inpatients	155	11	Intensive care	64	5
Emergency medical services	126	9	Cancer	63	5
Hospital	124	9	Surgery other	56	4
Hospital unit	98	7	Cardiology	54	4
Primary care	64	5	Diagnosis	52	4
Long term care	63	5	Neurology	49	4
Home care	55	4	HIV/AIDS	43	3

Table 8 ORAHS papers dealing with specific health management and general management issues (selection)

(a) Keywords within health management			(b) Keywords within general management		
GeneralWordGroup	Number	%	GeneralWordGroup	Number	%
Patient	738	54	Management	1042	76
Medical personnel	306	22	Efficiency	512	38
Health system	252	18	Needs	360	26
Public health	223	16	Finance	352	26
Health management	165	12	Personnel management	311	23
HC efficiency	127	9	Economics	238	17
HC finance and economics	109	8	Population	208	15
HC need	94	7	Regional level	163	12
HC pathway	86	6	Quality	162	12
HC quality	59	4	Policy	142	10
HC information systems	51	4	Social issues	89	7
Health policy	41	3	Environment	52	4

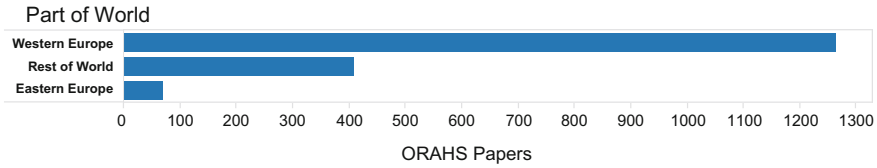


Fig. 3 Distribution of papers per countries represented at ORAHS conferences (broad groups)

simulation, in particular *discrete event simulation*, and *mathematical programming* are two most often used modelling approaches, primarily for modelling healthcare problems in *hospitals*, dealing with *surgical* services, or—especially in the last decade—dealing with *patient*-oriented healthcare issues.

4 Bibliometric Analysis of ORAHS Research

The second part of the literature review consisted in SciMAT-based bibliometric analysis of trends, areas of applications, evolution of techniques, and emerging problem areas in ORAHS research. The analysis was done using unified keywords, defined as separate WORD-GROUPS. Two (1985–2005, 2005–2015) or four (1985–2000, 2001–2005, 2006–2010, 2011–2015) periods were defined for longitudinal analysis. In each period, research themes were analysed by applying a

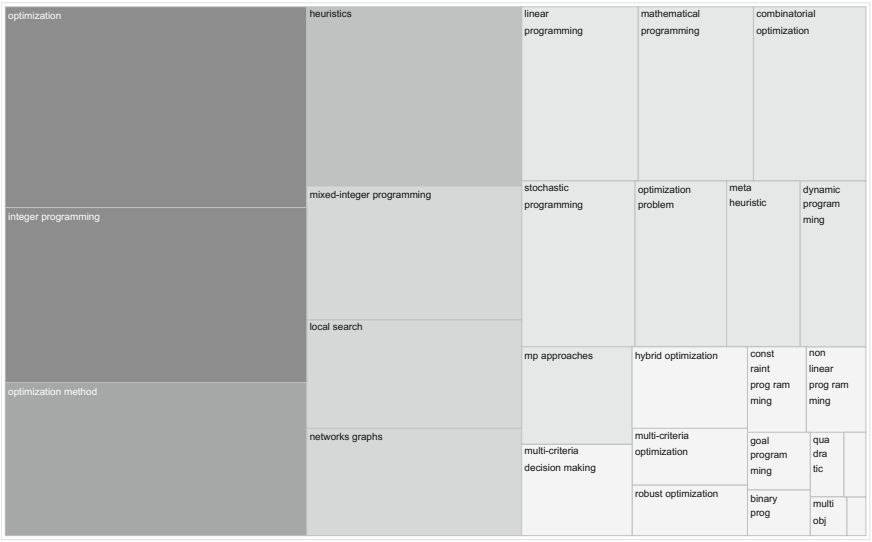


Fig. 4 Visualisation of magnitudes of modelling approaches as applied in ORAHS papers

co-word analysis followed by clustering of keywords to themes using the simple centers algorithm with standard parameters for analyses for most problem categories: minimum network size 3, maximum network size 12. No specific bibliometric quality measures (of the available: sum/min/max/average citations, h/g/hg/q2 index) were selected for ORAHS papers in addition to the standard measure (number of articles), as citation information was available only for very few ORAHS papers.

Resulting visualisations of the thematic evolution maps for particular problem sub-domains and sample strategic diagrams for modelling approaches are presented on Figs. 5, 6, 7 and 8. Figures 5 and 6 show strategic diagrams for two periods, visualising classifications of main research themes into four groups (the sphere size is proportional to the number of papers; [10]):

- motor themes for healthcare modelling domain (upper-right quadrant), i.e. the topics most developed and important for the structuring of the field; we may observe a great diversity of such topics in the period 1985–2005, which however are isolated in numbers, with the exception of generally labelled *optimization*, in most cases related to mathematical programming; in the period 2006–2015 we get unquestionably *DES* as the most important research topic, with additional major topics in that period: *heuristics*, *discrete optimization*, and *phase type* models;

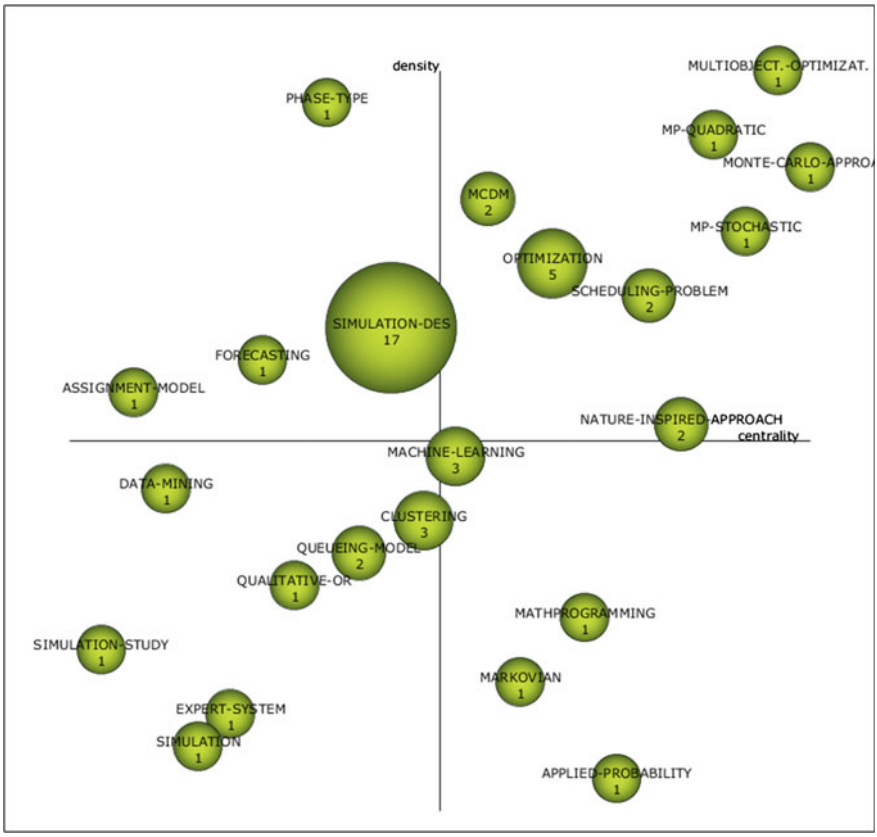


Fig. 5 Areas and trends in ORAHS research: strategic diagram for modelling approaches for the 1985–2005 period (created with SciMAT software [12], version 1.1.03)

- general and transversal, basic themes (lower-right quadrant), important for the research field and being in the development stage; in both periods this group contains *applied probability*, while in the second period it also contains simulation-related (*agent-based simulation*, *simulation optimization*) and discrete optimization-related (*linear MP* models, *scheduling problem*) notions;
- specialized, though peripheral so far, well-developed themes (upper-left quadrant); in fact *DES* was the major member of this cluster in the first period before becoming the most important motor theme in the second period when the group includes *queueing models* and *multi-objective optimization*;
- the lower-left quadrant of a strategic diagram includes the themes of (temporary) secondary importance for the whole domain, which mainly represent either emerging or disappearing themes (in both periods the group contains *qualitative OR*, *simulation study*, and *expert system*).

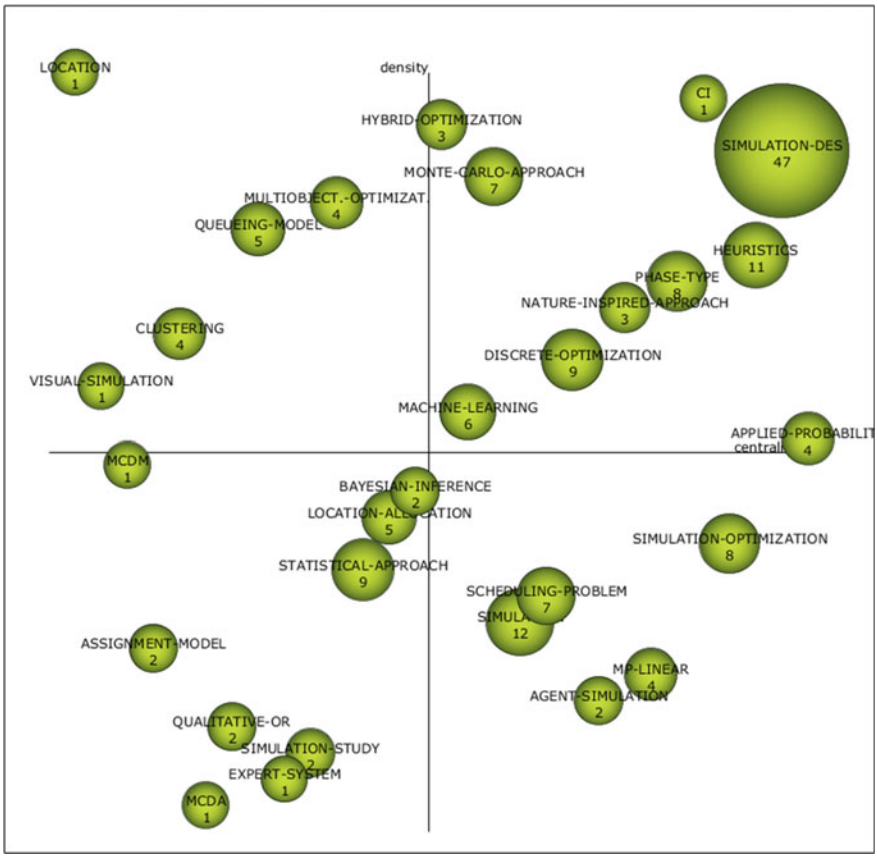


Fig. 6 Areas and trends in ORAHS research: strategic diagram for modelling approaches for the period 2006–2015 (created with SciMAT software [12], version 1.1.03)

Figures 7 and 8 present selected results of the longitudinal analysis of the evolution of main research themes across periods. Due to editorial reasons the figures illustrate only a selection of thematic areas detected in relation with modelling approaches (Fig. 7), healthcare management issues (Fig. 8a) and medical issues (Fig. 8b). Among modelling issues, two thematic areas were identified for the whole period 1985–2015: *Simulation* (evolving from general notions into most specific DES) and *Classification* (which had been applied, using also DEA and simulation models to classify healthcare units). There are also other distinguishable thematic nexuses in the period 2006–2015 connected with an evolution of more general terms (*MP*, *optimization*) into more specific approaches (*heuristics*, *network models*).

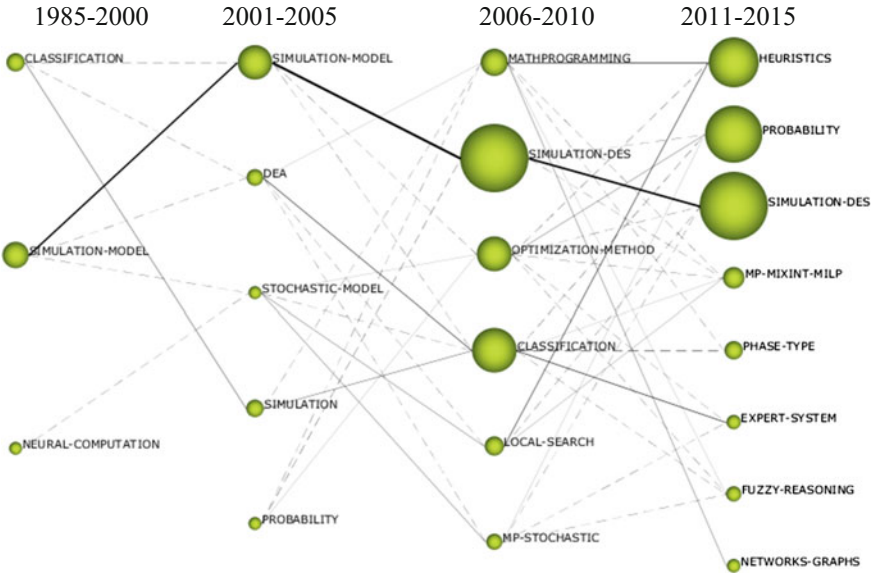


Fig. 7 Areas and trends in ORAHS research: thematic evolution of modelling approaches in ORAHS papers 1985–2015 (created with SciMAT software [12], version 1.1.03)

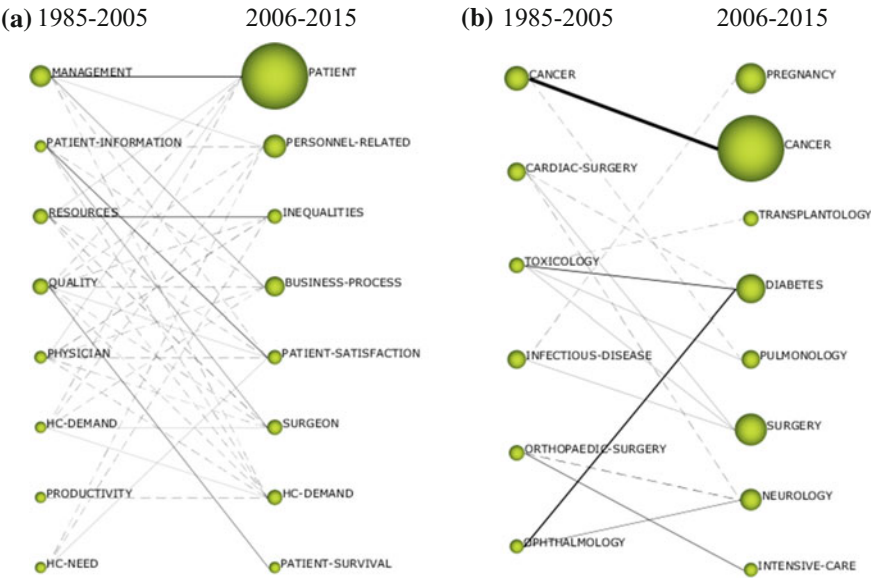


Fig. 8 Areas and trends in ORAHS research: thematic evolution of **a** healthcare management issues and **b** medical issues in ORAHS papers 1985–2015 (created with SciMAT software [12], version 1.1.03)

5 Concluding Remarks

We developed a novel methodology of bibliographic analysis of a large problem domain followed by a bibliometric analysis using SciMAT science mapping software. A general structured taxonomy of the domain's keywords was created and implemented for multiclass keywords to papers assignment. The methodology was applied to analyse 1363 ORAHS papers from the period 1985–2015. We have found that in relation to modelling approaches, ORAHS research was focused on the following thematic areas: Simulation, Mathematical Programming, and broadly interpreted Classification. The major trend in health management issues consisted in the transition from Management to Patient-centred issues. Amongst clinical management issues, Cancer remained the major topic, together with Surgery-related themes and developing relation between Ophthalmology (diabetic retinopathy) and general Diabetes-related issues.

Finally we have confronted our results (denoted ORAHS 2016 on Fig. 9) with those published in [2], concerning smaller samples of ORAHS research and general healthcare modelling literature (RIGHT project [1]). We found much larger proportion of mathematical models in the extended ORAHS sample. We have distinguished additional broad categories of modelling approaches, namely intelligent systems (e.g. machine learning, analytics, approximate reasoning, artificial and computational intelligence), and applied probability, which was differentiated from statistical modelling. We have identified a large set of other modelling approaches, e.g. economic and epidemic models; these were however not taken into account in our discussion in this paper. Our view is that a multiclass assignment of the

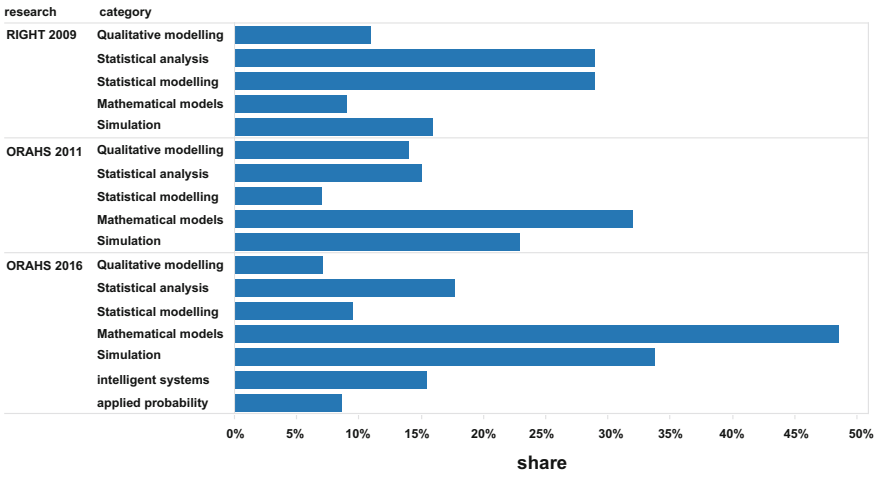


Fig. 9 Comparative analysis of healthcare modelling reviews: methodology by broad category

keywords, as applied in our review in contrast to separable classifications in [1, 2], enables a more comprehensive flexible depiction of a research field. Further research should consist in applying our methodology to a broader set of academic healthcare modelling literature.

References

1. Brailsford, S.C., Harper, P.R., Patel, B., Pitt, M.: An analysis of the academic literature on simulation and modelling in health care. J. Simul. **3**(3), 130–140 (2009)
2. Brailsford, S., Vissers, J.: OR in healthcare: a European perspective. Eur. J. Oper. Res. **212**(2), 223–234 (2011)
3. Complex Systems Modelling Group: Modelling in Healthcare. Simon Fraser University, American Mathematical Society, Providence (2010)
4. Asllani, A.: Business Analytics with Management Science Models and Methods. Pearson Education, Upper Saddle River (2015)
5. Institute for Operations Research and the Management Sciences (INFORMS). https://www.informs.org/About-INFORMS/What-is-Operations-Research
6. Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., Roberts, K., Coyle, E., Bevan, G., Palmer, S.: Systematic review of the use and value of computer simulation modelling in population health and health care delivery. J. Public Health Med. **25** (4), 325–335 (2003)
7. Jun, J.B., Jacobson, S.H., Swisher, J.R.: Application of discrete-event simulation in health care clinics: a survey. J. Oper. Res. Soc. **50**(2), 109–123 (1999)
8. Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: Science mapping software tools: review, analysis, and cooperative study among tools. J. Am. Soc. Inf. Sci. Technol. **62** (7), 1382–1402 (2011)
9. Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: SciMAT: a new science mapping analysis software tool. J. Am. Soc. Inf. Sci. Technol. **63**(8), 1609–1630 (2012)
10. Cobo, M.J., Martinez, M.A., Gutierrez-Salcedo, M., Fujita, H., Herrera-Viedma, E.: 25 years at knowledge-based systems: A bibliometric analysis. Knowl.-Based Syst. **80**, 3–13 (2015)
11. Lopez-Herrera, A.G., Cobo, M.J., Herrera-Viedma, E., Herrera, F., Bailon-Moreno, R., Jimenez-Contreras, E.: Visualization and evolution of the scientific structure of fuzzy sets research in Spain. Inf. Res. **14**(4), 4 (2009)
12. SciMAT—Science Mapping Analysis Tool. http://sci2s.ugr.es/scimat/

A Simulation Model of Aircraft Ground Handling: Case Study of the Wroclaw Airport Terminal

Artur Kierzkowski and Tomasz Kisiel

Abstract The article presents a system dynamic approach in the modelling of the aircraft ground handling process. The aim of the article is to use the developed simulation model to determine the operational ability of the aircraft. The article identifies subsequent activities performed during the ground-handling process and presents them in the form of graph and defines the dependence between them. An algorithm of the operation on the model and its verification was presented on the basis of data obtained from measurements conducted at the Wrocław Airport. Aircraft from category B according to ICAO was tested. On the basis of the model, the robustness of the ground-handling system was analyzed depending on the intensity of aircraft reports. The article discusses limitations of the model and the direction of work to implement further stages of tasks involving the development of the model of logistics support for the airport operation.

Keywords Airport · Ground handling · Simulation model

1 Introduction

The dynamic development of air transport requires accurate analysis of the entire system to ensure the planned implementation of the connection network. It is particularly important due to the transfer-based nature of the process. Passengers often use hub and spoke connections, during which they need to change planes. Such a trip may consist of even a few different connections served by various carriers.

A. Kierzkowski (✉) · T. Kisiel
Faculty of Mechanical Engineering, Department of Maintenance and Operation of Logistics, Transportation and Hydraulic Systems, Wroclaw University of Technology, Wrocław, Poland
e-mail: artur.kierzkowski@pwr.edu.pl

T. Kisiel
e-mail: tomasz.kisiel@pwr.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_9

The changing time is limited and, therefore, from the passenger's point of view, it is important that each part of the trip should be completed punctually.

The report, which was presented by Eurocontrol and drawn up by the Central Office for Delay Analysis [5], shows that in 2014, as many as 37.4 % air operations have been delayed upon the departure by more than 5 min. The average delay for delayed air operations was 26 min. Over 4 years, it was possible to reduce the value of the average delay by only 1.5 min. The percentage of delays caused by the Airline stays at the 30 % level. The classification of delays caused, amongst other things, by Aircraft and Ramp Handling as well as Passenger and Baggage Handling is surprising. A separate group of aircraft delays involves the so-called Airport group. Delays caused by this group constitute approx. 10 % of flights with a delayed departure. This group includes delays caused, amongst other things, by Restriction at Airport of Departure. The report also indicates that during an operational day, the percentage of secondary delays keeps growing to reach as many as 60 % towards the end of the day. Therefore, effects caused by the primary delay may bring about very disadvantageous consequences resulting from propagation of delays.

According to the [5] report, there are airports in the world, where as many as 55 % of air operations are delayed upon departure. The average delay can be over 30 min. The Lisboa Airport (LPPT), which serves approx. 16,000,000 passengers a year, is ranked first. In this way, LPPT causes more delays than the Heathrow Airport (EGLL), which is ranked second and which serves nearly 72,500,000 passengers a year. It is worth emphasising that Venezia Tessera Airport (LIPZ) serving only approx. 8,500,000 passengers a year was ranked 8. Therefore, it can be seen that the quality of services provided by airports does not depend mostly on the intensity of the passenger traffic. Thus, reasons or delays may vary, starting from the limited operational capacity of the airport to errors in the handling process management.

The results presented above are surprising, due to the fact that it is the airport that agrees to accept an aircraft within a given slot. It guarantees punctual ground handling and minimisation of risk of a delayed air operation. Timely handling by handling agents at a given airport is also imposed by internal national legal regulations [13], which state that the handling agent's main task involves, amongst other things, running their company in a manner, which guarantees proper functioning and elimination of interferences at the airport and ensures continuity of ground handling within the scope of the license.

The aim of this article is to develop and present the possibilities of using a simulation model supporting the ground-handling management and planning process at airports.

1.1 Ground Handling at the Airport

Ground handling of aircraft at the airport is a very broad notion. Ground handling tasks include [6]:

- ground administration and supervision
- passenger handling
- baggage handling
- freight and mail handling
- ramp handling
- aircraft services
- fuel and oil handling
- aircraft maintenance
- flight operations and crew administration
- surface transport
- catering services.

To develop a simulation model, the basic graph of activities performed at the stand during the aircraft ground-handling process (Fig. 1). The graph presents a full turnaround of an aircraft. The handling of some aircraft operations, for which the available slot is longer for an hour is divided into two stages: handling after the arrival (unloading-related activities) and handling before the departure (activities connected with the preparation of the aircraft to the next operation). In-between these stages, the aircraft is locked for some time.

Ground-handling operations at an airport are performed by ground-handling agents (GHA) who are authorised to perform activities in the aforementioned handling category (Airport Handling Agent Certificate). Activities performed by the GHA are based on a set of procedures contained in the Ground-Handling Manual introduced in a given company and approved by the President of the Civil Aviation Office. According to the GHM, the ground-handling agent performs activities connected with ground handling of an aircraft and the ground-handling process depends on the airline's standards and policy. Some airlines (especially budget ones) allow the performance of some activities independently (e.g. fuelling and passenger boarding, passenger boarding and catering services). The number of ground handling devices, which are necessary to perform the process also depends on the equipment and the size of an aircraft. Some Boeing 737–800 planes have an automatically unfolded airstair under the board. While planning the ground-handling process, the agent, that operates at the airport, must ensure one set of steps for such an aircraft. Due to the size of the aircraft, access to the hold can be direct or by means of a belt conveyor.

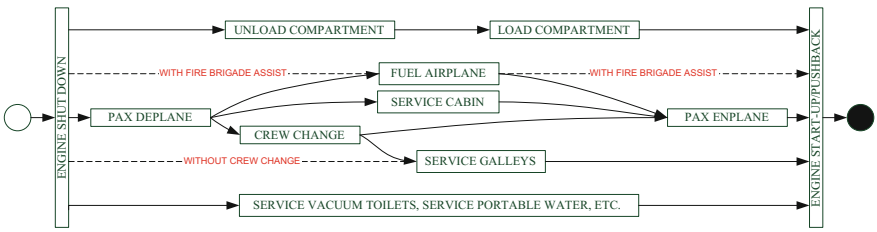


Fig. 1 General diagram of the ground-handling process at an airport

1.2 Current Status of Knowledge

There are many research studies devoted to the issue of ground handling at airports. These studies often focus on selected ground handling-related issues, which indicates the complexity of the process and the necessity of conducting a broader analysis of many aspects. Most studies focus on analysing the passenger boarding process. Authors [17] developed a simulation model of the passenger boarding process. This model allowed analysis of boarding scenarios, taking into account various types of passenger separation while boarding. On the basis of the analysis conducted, the authors showed the possibility of choosing an appropriate boarding strategy to reduce the boarding time. Author [4] developed a simulation model, which allows an analysis of identification of passenger congestion in the aircraft door, depending on various boarding scenarios. Author [28] concluded that passengers boarding in a manner allowing several passengers to load their luggage simultaneously, the boarding time can be dramatically reduced. This contradicts conventional wisdom and practice that loads passengers from the back of the airplane to the front. Study [10] investigated the issue of the influence of the boarding of disabled persons during the process. They specified what resources are needed to perform the process and the duration of the boarding of disabled persons was defined. Authors [27] on the other hand, estimated how long the waiting time for latecomers should be in the boarding process.

The conducted analyses also pertain to other component subprocesses. In the research conducted, Abdelghany et al. [1] also noticed that the aircraft ground-handling process should be optimised already at the terminal-handling stage. A model has been developed, which makes it possible to assign departing flights to available piers in the baggage sorting area. The model adopts the activity selection algorithm, which is modified to satisfy different operation requirements associated with baggage handling. Authors [19] focused on developing a scheduling method for de-icing device operation. The implementation of this Stockholm Arlanda Airport method reduced the number of delays and minimised the waiting time. Apart from logistics issues, energy consumption issues are often analysed. Authors [35] developed a solution, which makes it possible to minimise fuel consumption during an operating day of aircraft de-icing devices. Authors [7], on the other hand, developed a model for scheduling push-back device operation. The influence of the human factor on the ground-handling process is also taken into consideration. However, such analyses often focus on examining the influence of the human factor on aircraft damage [31].

Another group of publications includes articles devoted to ground handling in a broader sense. In the article [32], ground handling was analysed as a sub-process in aircraft rotations. By analysing historical data, the authors determined the probability of delays in the ground-handling process due to individual handling sub-processes (baggage loading, unloading, etc.). On the basis of research and observations, Nugroho et al. [21] developed a set of practical recommendations, which may improve the efficiency of the ground-handling process (e.g. providing

additional marking for passengers at the terminal, developing a simulation model for individual ground-handling stages, etc.). One of more interesting issues in the ground-handling planning was presented by Ansola et al. [3]. They presented a system operating on the basis of a three-level architecture composed of an end node level (RFID), a router level (RFID and Zigbee) and a coordinator level (Zigbee and Database). The system makes it possible to provide short-term support to process management. The objective function is based on three factors: the cost, energy and time consumption. Andreatta et al. [2] developed an algorithm for the allocation of technical and human resources to the implementation of individual ground-handling activities. The main assumption of the developed heuristic method is the division of resources into groups allocated to the ground handling of a given aircraft. This method focuses on one-day analysis. The quickness of the method is an advantage here. The implementation of the method made it possible to increase the efficiency of the process.

A separate group includes articles, which pertain to the stages of planning and managing air transport processes, which are directly or indirectly related to ground-handling processes. In one of such articles, Sinclair et al. [25] confirm that the continuation of using hub-and-spoke networks means that small disruptions can have a significant impact on operational costs. They propose solutions, which make it possible to correctly plan connections on the basis of data on interferences of air operations at given airports. Samà et al. [23], in the developed method for aircraft scheduling in the terminal control area using the rolling horizon approach showed that ground-handling planning according to the branch-and-bound strategy is the most effective. Such an approach makes it possible to achieve much higher efficiency than the first-come-first-served strategy. Optimal management of the stand allocation process for the ground-handling process is related to the estimation of the time of arrival and the taxi time of an aircraft, which is also the subject of research aimed at unleashing the power of soft computing methods, in particular fuzzy rule-based systems, for taxi time prediction problems [22]. Skorupski [26] presented a concept of airport traffic modelling using coloured, timed, stochastic Petri nets. Zając and Świeboda [33] developed an algorithm, which makes it possible to determine the order of tasks in the transport process. Meersman et al. [18] developed a method, which makes it possible to assess the optimal number of ground-handling agents at the airport. Methodology is based on the concept of economies of scale, leading to the hypothesis that the fact that an industry has no or few opportunities for economies of scale is an incentive to increase the number of providers of ground-handling services. Schmidberger et al. [24] organised the knowledge on the possibility of assessing the efficiency of ground-handling processes at airports on the basis of an overview of available methods. Action research was conducted to develop a holistic performance measurement system. The developed system represents a suitable basis for competitive benchmarking activities. Zhang et al., on the other hand, [34] presented the possibility of using the DEA-BCC method to assess the effectiveness of airport airside activities. For processes with time limitations, Kierzkowski and Kisiel [14, 15] proposed the possibility of using the availability coefficient to assess the reliability of processes.

In the literature, there is a significant gap as regards the development of models, which take into account the reliability of ground-handling processes. Models, which indicate optimal solutions, do not make it possible to estimate the risk connected with the possibility of interference in the form of delayed handling, due to the stochastic nature of the processes, reliability of human and technical resources. Reliability is commonly considered for other transport processes [11, 12, 16] or waste management [8, 9] taking into account the issue of deterioration [29]. Operational planning of a process can also be extended to include the development of a model for maintenance management [20]. From the point of view of system management, it also important that the system vulnerability to interference-inducing factors should be determined [30].

This study presents a simulation model that allows to determine the probability of a delay of the aircraft and to show its reason. The undoubted advantage is the ability to analyze the ground handling as a single process, not dividing it into separate parts.

2 Simulation Model

The developed simulation model operates in accordance with the algorithm presented in Fig. 3. The flight schedule with the identification of necessary ground-handling activities is entered as input data. Tasks were marked as $T(j)$, where j means the consecutive task number. A task assigned to a given aircraft is marked as $T(i, j)$, where the i index is allocated by assigning task $T(j)$ to an appropriate aircraft operations. 12 ground-handling tasks were identified. Each of these tasks has a defined random duration for a given aircraft, which is marked as $t(i, j)$, where i means the aircraft operation number (the consecutive number of an aircraft from the flight schedule), while j means the number of the consecutive task. The names of tasks are presented in Table 1. The notation of input data for the flight schedule is consistent with (1).

$$|i|s|l|d|T(i,1)|T(i,1)|T(i,2)|T(i,3)|\ldots|T(i,11)|T(i,12)| \tag{1}$$

where:

i	flight operation index,
s	type of aircraft,
l	landing time,
d	departure time,
$T(i, 1)$, …, $T(i, 12)$	state for tasks during handling (0—to omit, 1—to perform).

The beginning of simulation is initiated by two parallel algorithms (Figs. 2 and 3). The first of them (Fig. 2) makes a loop, the task of which involves a change in the current status of the possibility of ground handling for a given aircraft.

Table 1 Limiting conditions

Aircraft A(i)			Actual state of handling											
Task name	Task designations	TASK SIGN T(j)	GH(i)	T(i, 1)	T(i, 2)	T(i, 3)	T(i, 4)	T(i, 5)	T(i, 6)	T(i, 7)	T(i, 8)	T(i, 9)	T(i, 10)	T(i, 11)	T(i, 12)
Engine shut down	$f_{Besd}(t_{(i,1)})$	T(1)	h	1	a	a	a	a	a	a	a	a	a	a	a
Deplane passengers	$f_{Bdp}(t_{(i,2)})$	T(2)	h	3	1	a	a	a	a	a	a	1	a	a	a
Enplane passengers	$f_{Bep}(t_{(i,3)})$	T(3)	h	3	3	1	3	3	3	a	a	3	a	a	a
Cabin crew exchange	$f_{Bcce}(t_{(i,4)})$	T(4)	h	3	3	1	1	1	a	a	a	a	a	a	a
Service galleys	$f_{Bsg}(t_{(i,5)})$	T(5)	h	3	3	1	3	1	a	a	a	a	a	a	a
Service cabin	$f_{Bsc}(t_{(i,6)})$	T(6)	h	3	3	1	a	a	1	a	a	a	a	a	a
Unload compartment	$f_{Buc}(t_{(i,7)})$	T(7)	h	3	a	a	a	a	a	1	1	a	a	a	a
Load compartment	$f_{Blc}(t_{(i,8)})$	T(8)	h	3	a	a	a	a	a	3	1	a	a	a	a
Fuel airplane	$f_{Bfa}(t_{(i,9)})$	T(9)	h	3	3	1	a	a	a	a	a	1	a	a	a
Service portable water	$f_{Bspw}(t_{(i,10)})$	T(10)	h	3	a	a	a	a	a	a	a	a	1	a	a
Service vacuum toilets	$f_{Bsvt}(t_{(i,11)})$	T(11)	h	3	a	a	a	a	a	a	a	a	a	1	a
Engine start-up	$f_{Besu}(t_{(i,12)})$	T(12)	h	3	3	3	3	3	3	3	3	3	3	3	1

where:
1—to perform, 2—pending, 3—completed, a—any of 1, 2, 3 state, h—handling state

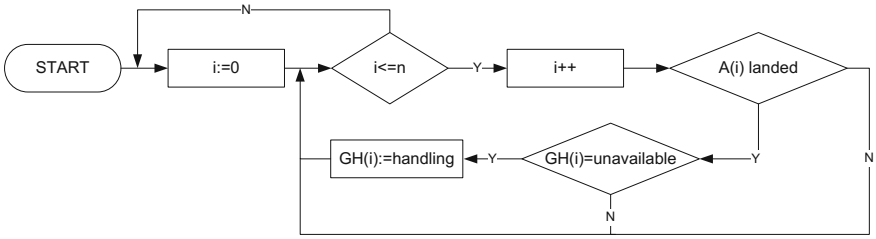


Fig. 2 Main algorithm for timetable

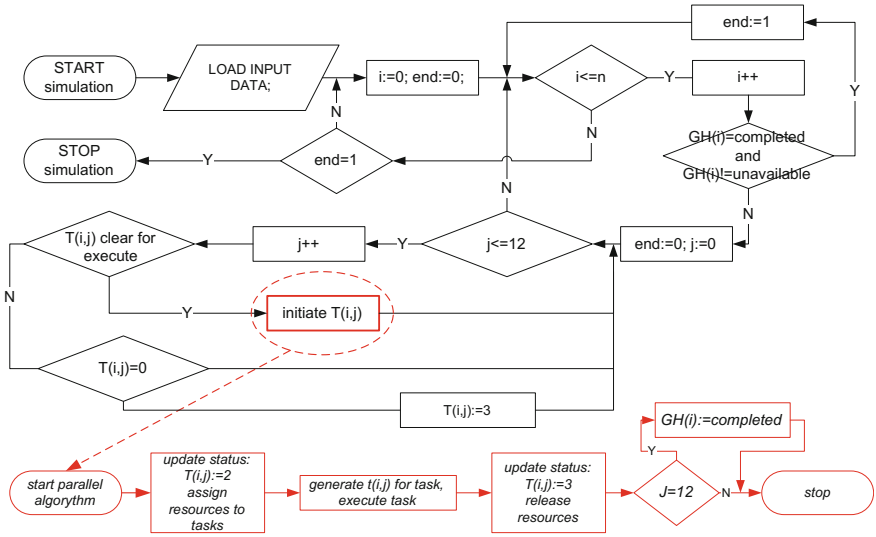


Fig. 3 Main algorithm for dispatcher

For subsequent aircraft operations $A(i)$, it is checked whether the aircraft is situated on the stand. If an aircraft has landed and its current ground-handling status (assigned during the input data import) is still in the "unavailable" position (Table 1, column $GH(i)$), the status is changed to *handling*.

The algorithm responsible for ground handling (Fig. 3) works for further aircraft operations from $A(1)$ *to* $A(n)$—the current ground-handling status. If it is not possible to perform a further task for a given aircraft or its ground handling has been completed, the handling status for the next aircraft is checked. For aircraft, for which ground handling has not been completed yet, conditions allowing the commencement of tasks $T(j)$ are checked in Table 1. If it is possible to perform task $T(j)$ for a given aircraft $A(i)$—the current handling status is consistent with the table of conditions (Table 1) and for $T(3)$, the passenger boarding at the gate has been completed (20 min. before departure) and for $T(8)$, the check-in process has been completed (30 min. before departure) and also technical resources for the

performance of the given task $T(j)$ are available, a parallel algorithm is initiated, which is responsible for the implementation of task $T(j)$. The task is assigned a status $T(i, j)$: = 2 (*pending*) and technical resources are allocated. The handling time $t(i, j)$ is generated. After task completion, its status is changed to completed $T(i, j)$: = 3 and technical resources are released. If the last task has been completed $j = 12$, the aircraft receives the completed status $GH(i)$: = *completed*. The value of 3 is assigned by the algorithm to tasks, which do not need to be performed, thus marking them as completed. The simulation ends the moment all aircraft operations have been completed.

3 Model Verification

The simulation model was verified on the basis of the operation of a real system. Research was conducted at the Wrocław Airport. Over 100 air operations of aircraft in category B were examined. B. CRJ-700 and CRJ-900 aircraft were handled. A detailed handling process of the examined aircraft was identified. Although the Fig. 1 shows that the process is complex, investigated flight operations take into account only some tasks (engine shutdown, deplane passengers, enplane passengers, service cabin, unloading compartment, loading compartment, fuel airplane and engine start-up). Regional jets perform air operations over short distances, so most of the tasks are performed only in the base airport. Ongoing research took place at the airport, which is not a base airport for these aircraft. The research work taking into account the handling of larger aircrafts like B737, A320 is still ongoing. The functionality of the model will be developed in the next steps upon completion of the extensive data.

On the basis of the research conducted, probability density functions were estimated, which characterise category B aircraft handling times for the following tasks: T(1) engine shutdown, T(2) deplane passengers, T(3) enplane passengers, T(6) service cabin, T(7) unloading compartment, T(8) loading compartment, T(9) fuel airplane and T(12) engine start-up. The functions are presented by the following formulae (2–9).

Estimating of the probability density function process was conducted using the estimator ExpertFit, in addition to the software FlexSim. Functions of best fit have been selected.

$$f_{Besd}(t_{(i,1)}) = \frac{1}{B(1.07, 1.33)} \frac{\left(t_{(i,1)} - 0.34\right)^{0.07} (1.99 - t_{(i,1)})^{0.33}}{1.65^{1.4}} \tag{2}$$

$$f_{Bdp}(t_{(i,2)}) = \frac{1}{B(1.60, 2.04)} \frac{(-1.07)^{0.6} (7.96 - t_{(i,2)})^{1.04}}{6.89^{2.64}} \tag{3}$$

$$f_{Bep}(t_{(i,3)}) = \frac{1}{B(1.66, 1.82)} \frac{\left(t_{(i,3)} - 0.97\right)^{0.66} (12.69 - t_{(i,3)})^{0.82}}{11.72^{2.48}} \tag{4}$$

$$f_{Bsc}(t_{(i,6)}) = \frac{1.83}{4.90} \left(\frac{t_{(i,6)} - 1.72}{4.90}\right)^{0.83} \exp\left(-\left(\frac{t_{(i,6)} - 1.72}{4.90}\right)^{1.83}\right) \tag{5}$$

$$f_{Buc}(t_{(i,7)}) = \frac{2.16}{4.59} \left(\frac{t_{(i,7)}}{4.59}\right)^{1.16} \exp\left(-\left(\frac{t_{(i,7)}}{4.59}\right)^{2.16}\right) \tag{6}$$

$$f_{Blc}(t_{(i,8)}) = \frac{1.73}{7.76} \left(\frac{t_{(i,8)} - 4.49}{7.76}\right)^{0.73} \exp\left(-\left(\frac{t_{(i,8)} - 4.49}{7.76}\right)^{1.73}\right) \tag{7}$$

$$f_{Bfa}(t_{(i,9)}) = \frac{1}{B(4.28, 0.95)} \frac{\left(t_{(i,9)} - 2.03\right)^{0.66} (6.09 - t_{(i,9)})^{-0.05}}{4.06^{4.23}} \tag{8}$$

$$f_{Besu}(t_{(i,12)}) = \frac{\left(\frac{t_{(i,12)}}{17.65}\right)^{27.12}}{17.65 \cdot B(28.12, 99.99)\left(1 + \frac{t_{(i,12)}}{17.65}\right)^{128.11}} \tag{9}$$

The consistency of the estimated functions with the real data was checked using the Kolmogorov–Smirnov test at the significance level of $\alpha = 0.05$. The test results below the critical value of 1.36 approve the consistency of empirical and real data. The presented implementation times of individual tasks contain the direct time related to the task implementation as well as times of positioning necessary technical resources before and after task implementation.

The functioning of the model was verified on the basis of an assumption that the system is up-state and technical devices are available. The turnaround times were compared for real data and data generated by the simulation model were compared. The results are presented in Fig. 4. The Kolmogorov–Smirnov test at the significance level of $\alpha = 0.05$ was performed for real and model data, from which the results are that the empirical data are consistent with the real data (statistic test equal 0.7615 and is below critical value 1.36). The determined probability density function of the ground-handling time has the following form:

$$f(x) = \exp\left(-\left(\frac{40 - 18.92}{18.36}\right)^{6.94}\right) \tag{10}$$

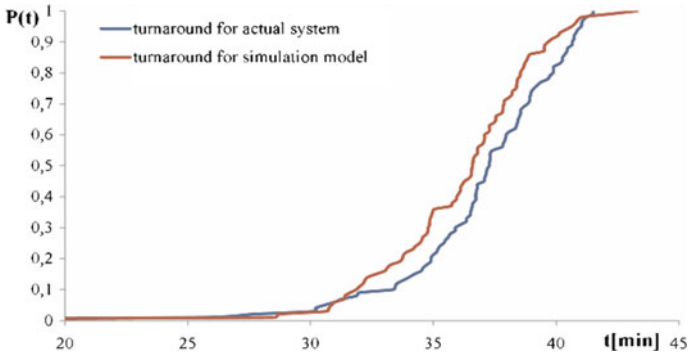


Fig. 4 Distributions for ground-handling times—verification of the model for aircraft operations performed for class B planes

4 Model Application: The Wroclaw Airport

Ground handling of aircraft makes it necessary to ensure relevant number of technical and human resources to perform the process. Erroneous process management may influence the aircraft handling time. Aircraft have different designs and, as a result, a different number of resources is often required for the implementation of the process. In the discussed case for category B aircraft, the engine stopping-and-starting process does not require any resources on the part of the handling agent. Handled aircraft are equipped with their own airstairs and, as a result, no airport airstairs need to be used. It is necessary to assign a cabin cleaning group. Fuelling requires the use of an airport cistern, one-belt conveyor, baggage cards and tugs are used to load and unload goods. In the simulation models, loading and unloading resources are treated as a group dedicated for handling a given aircraft operation. A list of requirements and availability of resources assumed for further analysis are presented in Table 2.

Considering the verification of the operation of the simulation model for data obtained from the sample, the probability of delays in handling category B airplanes

Table 2 Technical resources—assumptions

Task name	Task sign	Resources	
		Required	Available
Engine shut down	T(1)	–	–
Deplane passengers	T(2)	–	–
Enplane passengers	T(3)	–	–
Service cabin	T(6)	1	2
Unload compartment	T(7)	1	2
Load compartment	T(8)	1	2
Fuel airplane	T(9)	1	2
Engine start-up	T(12)	–	–

Table 3 Timetable

i	s	l	d	T(i, 1)	T(i, 2)	T(i, 3)	T(i, 4)	T(i, 5)	T(i, 6)	T(i, 7)	T(i, 8)	T(i, 9)	T(i, 10)	T(i, 11)	T(i, 12)
1	B	0	40	1	1	1	0	0	1	1	1	1	0	0	1
2	B	$dd(k) * 2$	$dd(k) * 2 + 40$	1	1	1	0	0	1	1	1	1	0	0	1
...	…	...	...	…	…	…	…	…	…	…	…	…	…	…	…
i	B	$dd(k) * i$	$dd(k) * i + 40$	1	1	1	0	0	1	1	1	1	0	0	1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
15	B	$dd(k) * 15$	$dd(k) * 15 + 40$	1	1	1	0	0	1	1	1	1	0	0	1

for the entire handled population (11) with the assumption that events (further handling task) are independent is:

$$P(x > 40) = 1 - P(x \leq 40) = \exp\left(-\left(\frac{40 - 18.92}{18.36}\right)^{6.94}\right) = 0.07 \qquad (11)$$

Using the simulation model, it is possible to analyze the resistance of the system, depending on the intensity of served flight connections (dependence between events —limited number of resources for the implementation of subsequent tasks). Therefore, the input flight schedule with varying intensity of reports in the subsequent scenarios. The time between consecutive reports is deterministic and marked as $dd(k)$, where k means the next simulation scenario. Tests for $dd(1) = 15$, $dd(2) = 12$, $dd(3) = 9$, $dd(4) = 6$ and $dd(5) = 3$ min were performed. Ground handling of 15 aircraft was assumed. A general flight schedule, according to notation (1) for all scenarios is presented in Table 3.

The simulation conducted made it possible to determine the probability of delays in the ground-handling process with a 40-min handling slot. The results are presented in Fig. 5. The analysis performed shows that for the assumed number of technical resources, the process is independent for report times between individual aircraft within a range of more than 9 min. The availability of technical resources makes it possible to comply with the assumed planned while keeping the probability of aircraft delays in accordance with the process assumption (11). For time intervals between individual reports which are shorter than 9 min, the handling of the following aircraft begins to depend on the availability of technical devices used for handling the previous aircraft, which increases the probability of delays in handling and extends their duration. An additional analysis allows for determining values of average delays in the handling for delayed implementations of the process. The results are presented in Table 4.

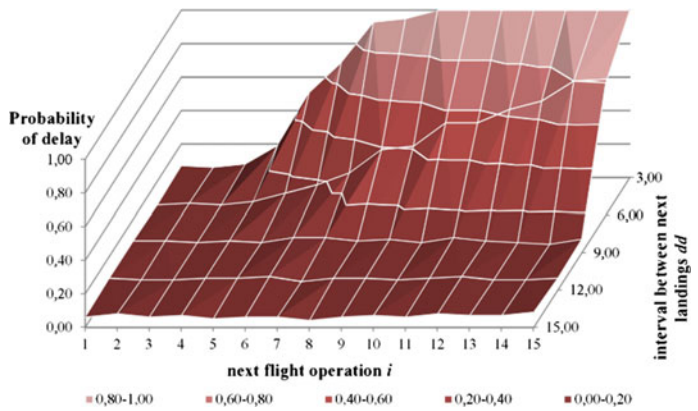


Fig. 5 Probability of handling delays for subsequent aircraft operations i, taking into account time intervals dd between the reports of individual planes

Table 4 Delay time analysis

i	Mean value of delay [min]				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
	$dd(1) = 15$	$dd(2) = 12$	$dd(3) = 9$	$dd(4) = 6$	$dd(5) = 3$
1	1.8	1.2	1.0	1.9	1.5
2	0.9	1.4	0.8	0.5	1.4
3	1.2	1.5	1.6	1.8	4.2
4	0.9	1.0	0.8	2.4	3.7
5	0.8	0.8	0.7	4.4	6.5
6	1.5	1.9	0.9	3.9	7.8
7	1.4	0.9	0.7	4.0	11.2
8	1.9	1.4	1.7	5.7	14.9
9	1.3	1.6	1.3	4.9	19.4
10	0.9	1.1	1.1	6.4	23.3
11	0.8	1.1	1.5	7.7	28.2
12	0.8	1.6	1.2	8.7	33.3
13	1.2	0.8	1.9	9.5	36.7
14	0.5	0.9	2.0	10.5	43.3
15	1.1	1.9	1.4	11.6	46.7
Mean	1.1	1.3	1.2	5.6	18.8

The simulation model does not cover a broader approach to the process to take into account dependent subprocesses. Various destinations are characterised by varying numbers of passengers and different types of baggage. The extension of the simulation model to include the possibility of entering data concerning the planned number of passengers and amount of baggage will allow for making the handling time dependent on this input data, which will increase the accuracy of the forecast operational capacity of the airport. Moreover, the other terminal handling processes also depend on this. For example, the results obtained for operations, for which low probability is forecast, may show that the baggage unloading process is moved closer towards the end of the ground-handling process of a given aircraft as no resources are available. In this case, the baggage claim wait time will be considerably extended for arriving passengers and, as a result, the handling process will need to be regarded as unreliable.

5 Conclusions

The article presents a method of analysing the functioning of the ground-handling system at the airport using computer simulation. The structure of the operation on the developed simulation model and advantages of its applications have been presented. Also, limitations, which have not been taken into account at the given stage of the development of the model, were also identified.

Further research work in the presented area will be aimed at determining basic time characteristics for aircraft from other categories. The research will also be targeted at determining dependence between the duration of individual handling activities and the characteristics of the connection served (the number of passengers, the number of goods handled, flight destination, etc.).

Developmental work on the extension of the simulation model will also be targeted at the possibility of assigning handling priorities to aircraft and also at handling chance events interfering with the process (reliability of technical devices, external factors).

Acknowledgments The project is co-financed by the National Research and Development Centre under the Applied Research Program. This publication presents the results of research conducted in the project: "Model of logistical support for the functioning of the Wroclaw Airport" realized by the Wroclaw University of Technology and Wroclaw Airport consortium

References

1. Abdelghany, A., Abdelghany, K., Narasimhan, R.: Scheduling baggage-handling facilities in congested airports. J. Air Transp. Manage. **12**, 76–81 (2006)
2. Andreatta, G., Giovanni, L., Monaci, M.: A fast heuristic for airport ground-service equipment-and-staff allocation, AIRO Winter 2013. Procedia—Soc. Behav. Sci. **108**, 26–36 (2014)
3. Ansola, P., Morenas, J., Garcia, A., Otamendi, J.: Distributed decision support system for airport ground handling management using WSN and MAS. Eng. Appl. Artif. Intell. **25**, 544–553 (2012)
4. Bazargan, M.: A linear programming approach for aircraft boarding strategy. Eur. J. Oper. Res. **183**, 394–411 (2007)
5. CODA: All-Causes Delay and Cancellations to Air Transport in Europe—2014 (2015)
6. Council Directive 96/67/EC of 15 October 1996 on access to the groundhandling market at Community airports (1996)
7. Du, J., Brunner, J., Kolish, R.: Planning towing processes at airports more efficiently. Transp. Res. Part E **70**, 293–304 (2014)
8. Giel, R., Plewa, M.: The assessment method of the organization of municipal waste collection zones. In: Theory and Engineering of Complex Systems and Dependability: Proceedings of the Eleven International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, Brunów, Poland (2016)
9. Giel, R., Plewa, M., Mlynczak, M.: Evaluation method of the waste processing system operation. In: Safety and Reliability: Methodology and Applications—Proceedings of the European Safety and Reliability Conference, ESREL (2016)
10. Holloway, C., Thoreau, R., Petit, P., Tyler, N.: Time and force required for attendants boarding wheelchair users onto aircraft. Int. J. Ind. Ergon. **48**, 167–173 (2015)
11. Jodejko-Pietruczuk, A., Plewa, M.: Components' rejuvenation in production with reused elements. Int. J. Perform. Eng. (2014)
12. Jodejko-Pietruczuk, A., Plewa, M.: Reliability based model of the cost effective product reusing policy. In: Safety and Reliability: Methodology and Applications—Proceedings of the European Safety and Reliability Conference, ESREL, pp. 1243–1248 (2014)

13. Journal of Laws of the Republic of Poland No 83, item 695. The Regulation of the Minister of Infrastructure concerning ground handling in airports (2009)
15. Kierzkowski, A., Kisiel, T.: Functional readiness of the security control system at an airport with single-report streams. In: Zamojski, W., et al. (eds.) Theory and Engineering of Complex Systems and Dependability, pp. 211–221. Springer (2015)
15. Kierzkowski, A., Kisiel, T., Functional readiness of the check-in desk system at an airport. In: Zamojski W., et al. (eds.) Theory and Engineering of Complex Systems and Dependability, pp. 223–233. Springer (2015)
16. Koucký, M., Vališ, D.: Reliability of sequential system with restricted number of renewals. In: Risk, Reliability and Social Safety ESREL 2007. Taylor & Francis London (2007). ISBN: 978-0-415-44786-7
17. Landeghem, H., Beuselinck, A.: Reducing passenger boarding time in airplanes: a simulation based approach. Eur. J. Oper. Res. 294–308 (2002)
18. Meersman, H., Pauwels, T., Struyf, E., Voorde, V., Vanelslander, T.: Ground handling in a changing market. The case of Brussels Airport. Res. Transp. Bus. Manage. **1**, 128–135 (2011)
19. Norin, A., Granberg, T., Yuan, D., Värbrand, P.: Airport logistics—a case study of the turn-around process. J. Air Transp. Manage. **20**, 31–34 (2012)
20. Nowakowski, T., Tubis, A., Werbińska-Wojciechowska, S.: Maintenance decision making process—a case study of passenger transportation company. In: Zamojski, W., et al. (eds.) Theory and Engineering of Complex Systems and Dependability, pp. 305–318. Springer (2015)
21. Nugroho, I., Riastuti, U., Iridiastadi, H.: Performance improvement suggestions for ground handling using lean solutions approach, international congress on interdisciplinary business and social science 2012. Procedia—Soc. Behav. Sci. **65**, 462–467 (2012)
22. Ravizza, S., Chen, J., Atkin, J., Stewart, P., Burke, E.: Aircraft taxi time prediction: comparisons and insights. Appl. Soft Comput. **14**, 397–406 (2014)
23. Samà, M., D'Ariano, A., Pacciarelli, D.: Rolling horizon approach for aircraft scheduling in the terminal control area of busy airports. Transp. Res. Part E **60**, 140–155 (2013)
24. Schmidberger, S., Bals, L., Hartmann, E., Jahns, Ch.: Ground handling services at European hub airports: development of a performance measurement system for benchmarking. Int. J. Prod. Econ. **117**, 104–116 (2009)
25. Sinclair, K., Cordeau, J.-F., Laporte, G.: A column generation post-optimization heuristic for the integrated aircraft and passenger recovery problem. Comput. Oper. Res. (2015). doi:10.1016/j.cor.2015.06.014
26. Skorupski, J.: Airport traffic simulation using petri nets. Commun. Comput. Inf. Sci. CCIS **395**, 468–475
27. Skorupski, J., Wierzbińska, M.: A method to evaluate the time of waiting for a late passenger. J. Air Transp. Manage. **47**, 79–89 (2015)
28. Steffen, J.: Optimal boarding method for airline passengers. J. Air Transp. Manage. **14**, 146–150 (2008)
29. Vališ, D., Koucký, M., Žák, L.: On approaches for non-direct determination of system deterioration. In: Eksploatacja i Niezawodnosc - Maintenance and Reliability, vol. 14, no. 1, pp. 33–41 (2012). ISSN: 1507-2711
30. Vališ, D., Pietrucha-Urbanik, K.: Utilization of diffusion processes and fuzzy logic for vulnerability assessment. Eksploatacja i Niezawodnosc - Maintenance and Reliability **16**(1), 48–55 (2014)
31. Wenner, C., Drury, C.: Analyzing human error in aircraft ground damage incidents. Int. J. Ind. Ergon. **26**, 177–199 (2000)
32. Wu, C.-L., Caves, R.: Modelling of aircraft rotation in a multiple airport environment. Transp. Res. Part E **38**, 265–277 (2002)

33. Zając, M., Świeboda, J.: An unloading work model at an intermodal terminal. In: Zamojski, W., et al. (eds.) Theory and Engineering of Complex Systems and Dependability, pp. 573–582. Springer (2015)
34. Zhang, B., Wang, J., Liu, C., Zhao, Y.: Evaluating the technical efficiency of Chinese airport airside activities. J. Air Transp. Manage. **20**, 23–27 (2012)
35. Zilio, C., Patricelli, L.: Aircraft anti-ice system: evaluation of system performance with a new time dependent mathematical model. Appl. Therm. Eng. **63**, 40–51 (2014)

Analysis of the Impact of Changes in the Size of the Waste Stream on the Process of Manual Sorting of Waste

Robert Giel and Marcin Plewa

Abstract Recycling centers are one of the most important elements of the waste management system. The achievement of required levels of recovery for installations is largely dependent on the proper organization of processes, in particular processes of sorting of waste materials. In most recycling center the key role is played by the processes of manual sorting of waste. The main goal of this article is to analyze the impact of changes in the size of the waste stream on the value of indicators used to evaluate the process of manual sorting of waste. For the analysis the authors used a simulation model that describes the process of manual sorting of waste in the WPO ALBA S.A. recycling center. Sensitivity analysis of simulation model will allow to improve the process by adjusting the size of the stream to the line ability.

Keywords Waste management · Waste processing · Reverse logistics · Simulation

1 Introduction

Recycling centers are one of the most important elements of the waste management system. The achievement of required levels of recovery for installations is largely dependent on the proper organization of processes, in particular processes of sorting of waste materials. In most recycling center the key role is played by the processes of manual sorting of waste.

The main goal of this article is to analyze the impact of changes in the size of the waste stream on the value of indicators used to evaluate the process of manual sorting of waste. The evaluation method for manual picking lines was described in [3].

R. Giel (✉) · M. Plewa
Wroclaw University of Technology, Smoluchowskiego 48, Wrocław, Poland
e-mail: robert.giel@pwr.edu.pl

M. Plewa
e-mail: marcin.plewa@pwr.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_10

The second chapter describes only the basic assumptions of the method. For the analysis the authors used a simulation model that describes the process of manual sorting of waste in the WPO ALBA S.A. recycling center. Similar use of computer simulation can be found at [5–8]. Sensitivity analysis of simulation model will allow to improve the process by adjusting the size of the stream to the line ability.

In the literature, there are four main groups of the assessments systems of waste processing systems:

- safety (impact of harmful substances on the health of installation workers) [9, 10];
- impact on the environment (LCA models) [2, 11];
- economic factors (models for comparing investment in terms of economic efficiency) [1];
- operational measurements (productivity, efficiency, capacity).

The method used in this article is a part of the last shown group. It includes methods for dealing with the overall assessment systems [4]. The methods described in the literature are based mainly on a single evaluation criterion.

Further work assume the development of an algorithm which will allow for an improvement of the process of manual waste sorting by amending the other parameters of the process line (e.g. Number of waste fractions selected from the stream on each workstations). The algorithm will use a computer simulation to analyze the process and the evaluation of parameter values which allow to indicate the directions improvement.

2 Description of the Real System

2.1 Waste Flow

The data used in the article were collected during research that has been performed in a WPO ALBA S.A. recycling center. The research involved sampling the weight of the waste stream in a given period of time. The waste stream has been described with probability distribution. In order to verify the distribution the Kolmogorov test has been carried out (significance level $\gamma = 0.95$). The test confirmed the correctness of distribution fitting.

Waste mass flow was described by equation:

$$f(s_m) = \frac{1.64}{0.69} \cdot \left(\frac{s_m - 0.2}{0.69}\right)^{0.64} e^{-\left(\frac{s_m - 0.2}{0.69}\right)^{1.64}} \quad (1)$$

Waste sorted by hand on the picking line installed in the analyzed sorting cabin may be divided into seven groups. The morphology of the waste is described in the Table 1. Waste fractions which are not picked by any stations in the line were classified as ballast.

Table 1 Waste morphology of S_1

Fraction name	Composition (%)
Blue PET	7
Green PET	2
White PET	8
Aluminium	2
Tetrapak	4
HDPE	8
Ballast	69
Sum	100

Table 2 Waste fractions mass

Fraction name	Item mass [kg]
Blue PET	0.04
Green PET	0.04
White PET	0.03
Aluminium	0.02
Tetrapak	0.07
HDPE	0.05

Because the productivity of the workstation is in the model expressed in the pieces, there was a need for the conversion of waste mass flow into pieces. Based on the research conducted in the WPO ALBA S.A. recycling center the average weight of one piece of a fraction of waste was determined (Table 2).

2.2 *Sorting Cabin*

Sorting cabin consist of ten workstations and conveyor belt characterized by a constant speed. For each of the 10 workstations the amount of items (waste) taken from a conveyor per unit time was tested. The opposite workstation picks up the same waste fractions (Fig. 1).

Type of waste picked at each workstation are described in the following table (Table 3).

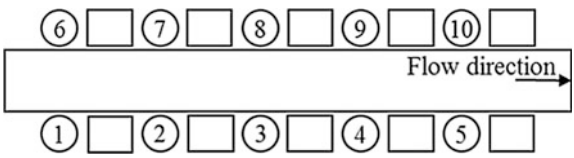


Fig. 1 Diagram of a real waste sorting line

Table 3 Waste picked up on defined workstation

Workstations	Picking fractions
1 and 6	Blue PET
2 and 7	Blue PET, White PET, Aluminium, Tetrapak
3 and 8	Blue PET, White PET, Aluminium, Tetrapak
4 and 9	Blue PET, Green PET, White PET, Aluminium, Tetrapak
5 and 10	Blue PET, Green PET, White PET, Aluminium, Tetrapak, HDPE

3 Model of Sorting Line

In this chapter authors describe the model of the waste stream on sorting line in waste recover facility. Total waste mass flow S_j at j-th workstation [kg/s] can be described by equation:

$$S_j = \sum_{i=1}^{m} s_{i,j}, \quad \text{for } j = 1, \ldots, n, \tag{2}$$

where: m—number of fraction in flow, n—number of workstations, where n-th workstation is container or next conveyer, $s_{i,j}$—waste mass flow of i-th fraction at j-th workstation [kg/s].

The waste morphology for $j = 1$ is known from morphological composition analyzes:

$$S_j = S_{j-1} - O_{j-1} \cdot B_{j-1}, \quad \text{for } j = 2, \ldots, n, \tag{3}$$

where: O_j—waste mass flow separated from main stream at j-th workstation [kg/s], B_j—task vector describing fractions to be separated at j-th workstation $B_j = [b_{j,i}]$,

$$b_{j,i} = \begin{cases} 0 & i\text{-th fraction not separated} \\ 1 & i\text{-th fraction separated} \end{cases}.$$

Waste mass flow of i-th fraction at j-th workstation $s_{i,j}$ [kg/s] can be described by equation:

$$s_{i,j} = s_{i,j-1} - o_{i,j-1} \cdot b_{j-1,i} \tag{4}$$

where: $o_{i,j}$—waste mass flow for fraction separated from main stream at j-th workstation [kg/s].

$o_{i,j}$—waste mass flow of fraction separated from main stream at j-th workstation [kg/s] can be described by equation:

$$o_{i,j} = \alpha_{a_{i,j}} \cdot P_j, \tag{5}$$

where: P_j—workstation productivity; $\alpha_{a_{i,j}}$—productivity distribution coefficient for i-th fraction on j-th workstation, which depends on $a_{i,j}$—i-th fraction percentage share of total flow at j-th workstation [%].

$a_{i,j}$ is described by following equation:

$$a_{i,j} = \frac{s_{i,j}}{S_j} \cdot 100\,\%, \tag{6}$$

where:

$$\sum_{i=1}^{m} a_{i,j} = 100\,\%.$$

O_j—waste mass flow separated from main stream at j-th workstation [kg/s] is described by following equation:

$$O_j = \sum_{i=1}^{m} o_{i,j}, \text{ for } i = 1, \ldots, m. \tag{7}$$

Possible fractions denoted by i are: 1—Blue PET, 2—Green PET, 3—White PET, 4—HDPE, 5—Foil, 6—Small plastic, 7—Glass, 8—Wastepaper, 9—Iron metals, 10—Aluminum, 11—Tetra pack, 12—Ballast.

On the basis of above assumptions the model of the single workstation has been developed (Fig. 2). The following figure shows the relationship between the parameters of the model.

Conducted research indicates that the maximum employee productivity is equal to $P_j = 28.8$ units/min, for $V_c = 0.4$ m/s. The model was developed in FlexSim software. The model enables the sensitivity analysis of manual picking line to changes of input parameter values. The parameter whose influence was studied was the size of a stream that goes to the sorting cabin. Waste stream was increased

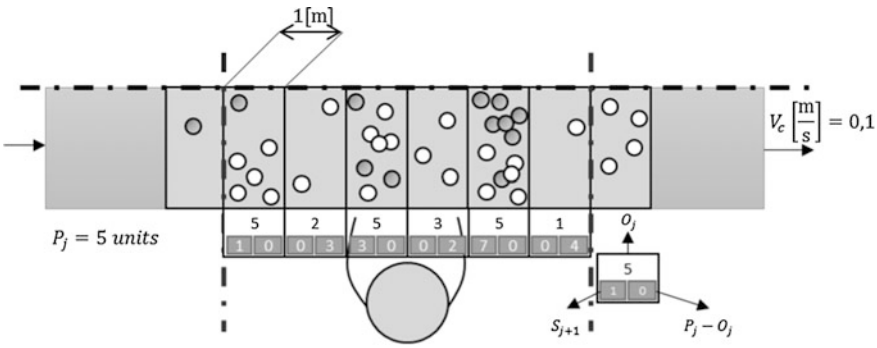


Fig. 2 Workstation graphical model

proportionately, as described at the beginning of the chapter. During the sensitivity analysis the impact on the evaluation indicators was assessed. Indicators used in evaluation method are described in the next section.

4 Method of Waste Processing System Evaluation

The analysis of the waste sorting process, where most operations are performed manually, should include an assessment of the entire process line as well as single workstation. It should be based on specific criteria. The following chapter describes a method of evaluation of the manual waste sorting process.

The first evaluation criterion for the technological line is its productivity. The maximum achievable productivity of the line is calculated by multiplying the number of stations and the maximum productivity of a single workstation (in the article it is the highest productivity achieved during research conducted in the real system).

Developed method is based on the assumption that all employees have the same theoretical productivity (maximum achievable). Volatility of the current productivity depends on the frequency of disturbances. Determination of impact of disturbances on productivity is the direction of further research and will not be analyzed in this article. The productivity of the line can be described by the ratio of actual and theoretical productivity by formula:

$$LP = \frac{\sum_{j=1}^{n} O_j}{\sum_{j=1}^{n} P_j} \cdot 100\,\%, \quad \text{for } j = 1, \ldots, n; \tag{8}$$

The second assessment criterion is the uniformity of the workstations workload determined by the coefficient of variation (LV_{TP}) calculated for the current productivity (P_j^R).

The third assessment criterion is the efficiency of the line defined as the ratio of the waste mass separated in the sorting process to the total waste mass. The efficiency of the sorting process can be determined by formula:

$$LE = \frac{\sum_{j=1}^{n} O_j}{S_1} \cdot 100\,\%, \quad \text{for } j = 1, \ldots, n; \tag{9}$$

Sorting line is characterized by three factors defined above and in real life one may observe various situations depending on values of criterions: LP, LVTP and LE. Assuming dichotomy of the criterions, they can valuated as low and high.

Table 3 and the following description include the interpretation of the obtainable value of the evaluation criteria. The situation where LP and LV_{TP} achieve high status didn't exist. When LP is high, that blocks coefficient of variation from high level.

States that system could be in, are described in the table above (Table 4).

Table 4 Evaluation of the technological line

States	Criteria values		
	LP	LV_{TP}	LE
A	High	Low	Low
B	High	Low	High
C	Low	Low	High
D	Low	Low	Low
E	Low	High	Low
F	Low	High	High

5 Sensitivity Analysis

5.1 Assumptions

Model of manual picking line was developed in the FlexSim software. The data for the preparation of the simulation model were collected in waste recovery facility in Wrocław (WPO ALBA S.A. recycling center). Authors used the self-made application for android system which collected data of workers productivity during one work shift.

During the sensitivity analysis the effect of increasing size of the waste stream (from 0.4 to 1.8 of the waste mass flow described in the Sect. 2.1), on the evaluation indicators for the entire line and individual workstations was studied. In this article the 8 scenarios were tested. Equation describing the size of the waste stream was presented in the second chapter. The probability distribution that describes the size of the waste stream was described in the Sect. 2.1. In the article the eight scenarios of changes in the waste stream were analyzed (Table 5).

In the presented model, the number of workstations has been reduced to five (opposite workstation picks up the same waste fractions). This makes it necessary to assume that the waste stream is distributed proportionally between the opposing sides of the sorting line. Productivity for each workstation in the model is equal to $P_j = 28.8$ units/min, for constant conveyor speed equal to $V_c = 0.4$ m/s. Diagram of sorting line is shown in Fig. 3.

Table 5 Simulation scenarios

Scenarios	Stream control parameter
1	0.4
2	0.6
3	0.8
4	1.0
5	1.2
6	1.4
7	1.6
8	1.8

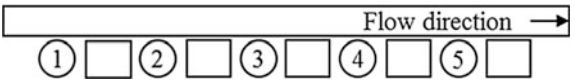


Fig. 3 Graphical model of waste sorting line

5.2 *Results of the Analysis*

At first the entire sorting line was evaluated. The values of evaluation parameters were varied. As shown in the chart below (Fig. 4).

To evaluate the manual picking line the following limit values for evaluation indicators were used (Table 6).

- LP High—when LP ≥ 70 %
- LV_{TP} High—when LV_{TP} ≥ 30 %
- LE High—when LE < 70 %

Table 6 shows the values of the indicators used to assess the sorting line. For value of the control indicator, defining the size of the stream, less than 0.8 the state F was observed. This means that the efficiency of line is achieved at the expense of its productivity. High value of the coefficient of variation inform that the waste does not reach the next workstations. For such situation, it is necessary to reorganize workplaces and/or reduce the number of workstations. Stream size between 0.8 and 1.0 gives the state C. This means that the size of the stream is too small for the line

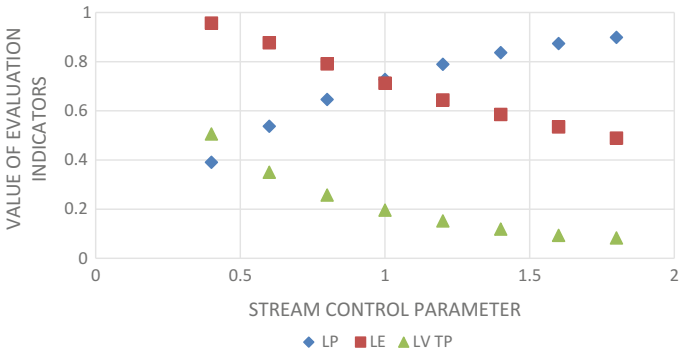


Fig. 4 Analysis of results from simulation

Table 6 Evaluation of the technological line

Evaluation measures	Stream control parameter							
	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
LP	Low	Low	Low	High	High	High	High	High
LV_{TP}	High	High	Low	Low	Low	Low	Low	Low
LE	High	High	High	High	Low	Low	Low	Low

to be able to achieve its maximum performance. However, the increase in productivity, will be at the expense of the efficiency of the line. The next state of the system was reached after crossing the value of the waste stream control index over 1.2 (state A). Employees are equally burdened and the line is closer to reaching its maximum capacity. Unfortunately, this happens at the expense of efficiency. Summing up optimal situation for the described sorting line is located between the states in which the value of the ratio determining the waste mass flow is between 1.0 and 1.2.

6 Summary

The authors of this article tried to analyze the impact of changes in the size of the waste stream on the value of indicators used to assess the process of manual sorting of waste. These indicators were part of authorial method of assessment of such systems. The described method makes it possible to identify weaknesses in the system and provide directions for improvement.

Data for the analysis came from the simulation model developed in the FlexSim software. The authors conducted 100 replications of simulation for 8 different scenarios. Scenarios studied the effects of changes in the value of the parameter controlling the size of the stream.

In the last chapter the authors pointed out the optimal size of the stream for the described system. Which does not change the fact that the situation described in the article is far from ideal. Therefore, in future works the authors will undertake to try to optimize the sorting line by developing an algorithm for assigning fractions of waste to each workstation.

References

1. Cimpan, C., Maul, A., Wenzel, H., Pretz, T.: Techno-economic assessment of central sorting at material recovery facilities—the case of lightweight packaging waste. Elsevier J. Clean. Prod. 112 (2016)
2. Gentil, E., et al.: Models for waste life cycle assessment: review of technical assumptions. Waste Manage. (30), 2636–2648 (2010)
3. Giel, R., Plewa, M., Mlynczak, M.: Evaluation method of the waste processing system operation. In: Safety and Reliability: Methodology and Applications—Proceedings of the European Safety and Reliability Conference, ESREL (2016)
4. Hryb, W.: Modernization of waste management plants exemplified by installation in. In: Archives of Waste Management and Environmental Protection, Tom 15, pp. 47–56 (2013)
5. Kierzkowski, A., Kisiel, T.: The simulation model of logistic support for functioning ground handling agent, taking into account the probabilistic time of aircrafts arrival. In: CLC 2013: Carpathian Logistics Congress—Congress Proceedings, pp. 463–469 (2013)
6. Kierzkowski, A., Kisiel, T.: Functional readiness of the check-in desk system at an airport. In: Theory and Engineering of Complex Systems and Dependability: Proceedings of the Tenth

International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, Brunów, Poland. Springer, cop. 2015 (2015)
7. Kierzkowski, A., Kisiel, T.: Functional readiness of the security control system at an airport with single-report streams. In: Theory and Engineering of Complex Systems and Dependability: Proceedings of the Tenth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, Brunów, Poland. Springer, cop. 2015 (2015)
8. Kierzkowski, A.: The use of a simulation model of the passenger boarding process to estimate the time of its implementation using various strategies. In: Advances in Intelligent Systems and Computing. Springer (in press)
9. Kozajda, A., Sowiak, M., Piotrowska, M., i Szadkowska-Stańczyk, I.: Waste sorting plants—recognition of exposure to biological agents (Moulds). Medycyna Pracy 6(60), 483–490 (2009)
10. Krajewski, J., Szarapińska-Kwaszewska, J., Dudkiewicz, B.: Assessment of exposure to bioaerosols in workplace ambient air during municipal waste collection and disposal. Medycyna Pracy 6(52), 417–422 (2001)
11. Pressley, P.N., Levis, J.W., Damgaard, A., Barlaz, M.A., DeCarolis, J.F.: Analysis of material recovery facilities for use in life-cycle assessment. Waste Manage. 35 (2015)

The Extension of User Story Template Structure with an Assessment Question Based on the Kano Model

Grażyna Hołodnik-Janczura

Abstract The aim of this study was to verify whether a simple method of customer satisfaction survey can aid Agile project management in an effort to provide customers with an IT solution that is well-matched to their actual needs. A pilot study was made to confirm the frequent occurrence of the problem under consideration. Based on literature review, an analysis of the relationship between user requirements and final properties of the software was made, starting from user goals, and ending with ways to ensure user satisfaction. Discussed was also the impact of customer business value on customer satisfaction from the use of the resulting IT solution. The described importance of the business value resulted in pointing to the need for combining Agile and Lean principles, leading to effective elimination of waste. Product backlog was characterized as one of the key artifacts to the Agile approach. This backlog became the foundation for the draft of a new user story template structure, as well as for its application in waste elimination as soon as at the stage of formulating the needs and requirements of the customer/user. The end of the study presents an example usage of a template designed in this manner.

Keywords Customer satisfaction · Elimination · Feature · Product backlog · Template · User story · Waste

1 Introduction

Taking up the issue of elimination of unnecessary features of computer application was inspired by author's own observations and other published data. Users of various computer applications often complain that software contains many features that they do not use. It was even found that only about 20 % functions on typical client software are frequently used, and 2–3 % are rare features. Thereby, the utility index of such products reduces [1, p. 24]. What occurs here is a waste of effort in

G. Hołodnik-Janczura (✉)
Wrocław University of Science and Technology, Wrocław, Poland
e-mail: Grazyna.Holodnik-Janczura@pwr.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_11

successive phases of a complex process, including coding, testing, documenting and, consequently, working code maintenance. Therefore, studying the adequacy of product features to the real needs of the customer plays an increasingly important role. To meet such expectations, a pilot study to confirm or reject this opinion was designed and conducted.

Additionally, knowledge of the application of the Kano model for customer satisfaction analysis in Agile projects reinforced with Lean principles was the source of the concept behind the attempt at solving this issue. The solution is designed as a combination of the two sides of the problem under examination, that is, the formulation of customer/user needs and requirements in the form of user story with their simultaneous value assessment from the point of view of the customer/user.

2 The Relationship Between Customer Satisfaction and the Degree to Which Customer Needs Are Met

Modern approach to ensuring customer satisfaction puts customer goals—that the customer intends to achieve by purchasing or ordering a new product—at the top of the pyramid of knowledge about customer needs. The methods of customer satisfaction study in the Agile approach are often based on the Kano model (Fig. 1) or MoSCoW method, both based on prioritization. Other methods, such as the very detailed and time-consuming QFD method, are not recommended by the generally agile and economical Agile approach. Noriaki Kano, consultant of the Japanese TQM, developed a customer satisfaction model in relation to the product features, distinguishing three categories of features based on customer preference assessment [2, p. 79]:

- must be: "expected" features, the presence of which the customer takes for granted, their absence is inevitably followed by customer dissatisfaction; e.g. inoperative part of the software, missing instructions. These features are alleged, the customer does not feel the need to express them, they are "obvious" and their absence causes irritation;
- linear: the features that the customer asks for and tells about; the more these features in a product, the greater customer satisfaction; e.g. greater efficiency, lower cost, higher reliability, easier to use;
- attractive: these product features give the customer a pleasant surprise, yet if they are absent, the client will not be dissatisfied; the customer does not express the desire to have such extras, or may not know how to express such needs, e.g. because of a lack of technical knowledge.

Customer satisfaction is inextricably linked with the concept of product quality. Product quality equals the degree to which customer needs are met; it is, namely, the core element of a universal product quality definition: "Quality is the degree to which a set of inherent features fulfills the requirements" [3]. Comparing the

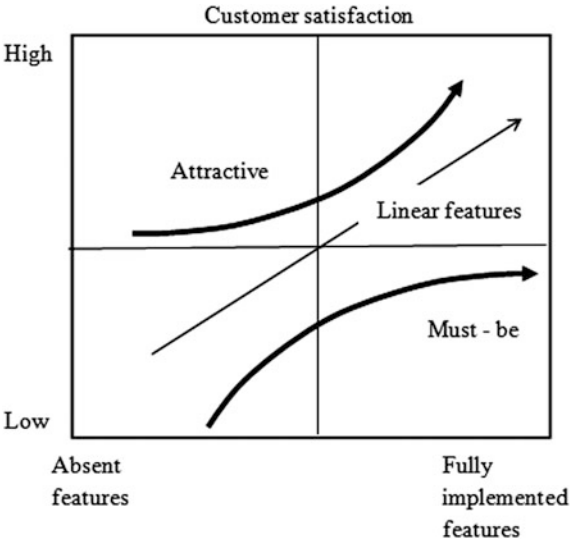


Fig. 1 Customer satisfaction model according to Kano. *Source* Author, based on [8, p. 113]

classification of Kano against this ISO definition, it is necessary to elaborate on the meaning of an inherent feature. It is a permanent feature of the product, process or system. Sample inherent features of a car are: amount of fuel consumed, maximum engine power. The term "requirement" means a need or expectation that is: stated, mandatory or customary.

Stated requirements are formulated by the customer in various ways, e.g. through negotiations with the supplier or as a result of marketing research. Sample requirements fixed by the customer in a car showroom are: the brand and model of car, number of doors, and engine power.

Mandatory requirements are such requirements that the supplier must meet regardless of whether the customer formulates them or not. They result from legal provisions, e.g. in the field of safety, health protection, fire protection. For example, the car must be made in accordance with applicable safety regulations or traffic admission regulations.

Customary requirements usually are not formulated by the customer, because of a general assumptions that the provider should fulfill them due to the current level of technical culture, and that they will be obviously met. However, an example of such a requirement may for one customer group be a built-in car radio, while another group may find such a feature excessive. This is a group of features subjectively classified as common, depending on the habitual use of facilities.

Thus, in the context of the Kano model, the mandatory requirements are perceived as "must be" and do not require further study. In contrast, the stated requirements can be assessed differently by various customers, likewise the customary requirements. Some customers will recognize such features as unexpected attractions, while other will see them as universal, which means they should be subject to further studies in order to be qualified as waste and eliminated.

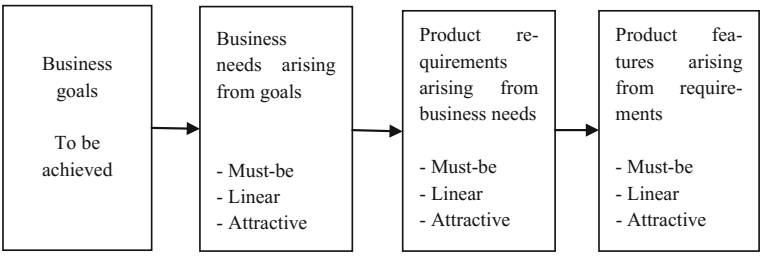


Fig. 2 Relationship between business goals and product features expressed through needs and requirements and their classification. *Source* Author, based on [4]

Due to the fact that the above definition of ISO is of universal character, quality models specific to IT solutions should be mentioned here. There are quite many, including: FURPS (+), IEEE standards, ISO/IEC 9126, ISO/IEC 15504. These models and standards are rooted in the original two: McCalla model (1977) and Boehm model (1978). In each of them, quality evaluation of an IT solution depends, similar to ISO standards, on the fulfillment of customer/user needs and requirements, and is measured by means of a corresponding set of features. In order to emphasize the existing relation between requirements quality and the quality of the final product, an organized group of IT quality standards known as SQuaRE [4] should be brought up. The SQuaRE model represents these significant relations (Fig. 2).

The SQuaRE model supports the occurrence of full dependence between the goals and business needs of project stakeholders, business needs and requirements, and consequently, the requirements and features of the product. The second conclusion is that the classification of these elements follows the same lines. Such an approach to customer satisfaction is considered the basis for deliberations on providing customers and software users only such properties that meet their exact needs. In contrast, providing a product with excess or inadequate features with respect to the actual needs is considered waste, causing higher costs and longer implementation deadlines.

3 The Use of Application Features Provided for the User: Study Results

The questionnaire study focused on the suitability of features of computer applications in use by the respondents. The questionnaire was posted on the website[1] of the author of this study, who received 12 filled questionnaires. Respondents had different, though generally long (between 9 and 36 years) experience of work with

[1] www.ioz.pwr.wroc.pl/Pracownicy/holodnik.

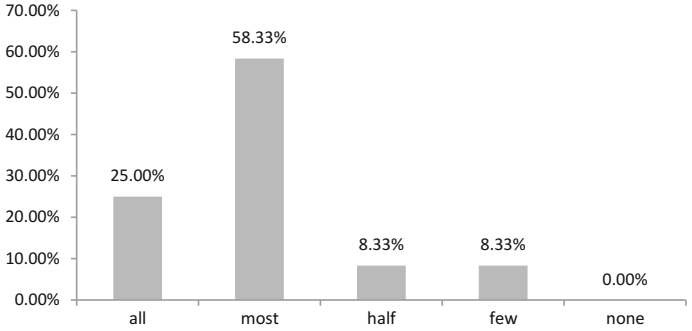


Fig. 3 Descriptive assessment of all used applications based on the content of unnecessary features. *Source* Author

the use of computer software for purposes such as work, education, entertainment, and handling of official matters. All respondents use between 15 and 100 various applications daily. The assessment of the satisfaction level on a 6-point scale, from very high level to total lack of satisfaction with the software used, was quite varied and depended on the type of software. Majority of responses pointed to high (9 respondents) and medium (7 respondents) levels of satisfaction, while the minority of responses pointed to the lowest level, or lack of satisfaction (2 respondents). All respondents claimed that significant parts of the software were unnecessary, that is, unused—e.g. 25 % of respondents think that all applications contain unnecessary features, and over 58 % think that most applications contain unnecessary features (Fig. 3).

When asked about how many of the features of used applications were unnecessary, 23 % of respondents said it was 60, 30.77 % of respondents said 30 %, and only 7.69 % of respondents think 10 % were unnecessary features (Fig. 4).

This short pilot study, although small, but conducted on a representative group of software users, confirmed other published and observed claims that the software market is filled with applications that have a significant number of unnecessary features.

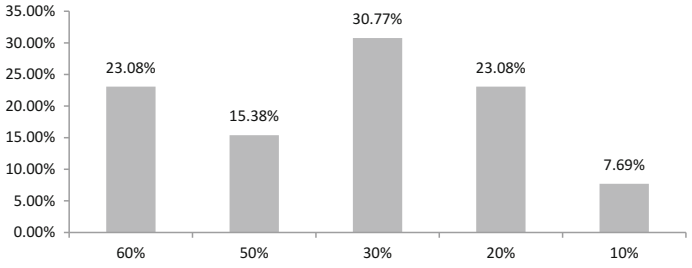


Fig. 4 Application features considered unnecessary in the software used by respondents. *Source* Author

4 Business Value as a Way to Increase Customer Satisfaction

Managing customer value is one of the marketing concepts from the wide area of management processes. In the 80s, Jack Welsh popularized the idea of value management [5] as one way to achieve customer satisfaction. In subsequent years, this concept has been well developed and increasingly used. The tools used to review value are primarily quantitative, but also semi-quantitative methods of financial analysis, since value can also be perceived subjectively, allowing a subjective assessment of product features by specifying: this is something I like, dislike, or feel neutral about.

In the management of IT projects, which are also subject to investment efficiency, the concept works in a similar way. Customer value, in IT projects referred to as business value, reflects the financial benefits expected from the implementation of a computer application, and resulting from growth in revenue or cut in spending. Financial indicators—such as ROI, NPV, IRR—used to determine the cost-effectiveness of alternative project drafts allow to take into account the concept of business value also in terms of IT project efficiency. On the client side, business value of a computer application is difficult to estimate, because it is not seen directly —contrary to income from sales. Business value is reflected in how it impacts the outcome of company operations, and is perceived as indirect and supporting the sales. Estimation of this impact is only possible with respect to a certain period of time, i.e. months or years.

On the other hand, the development of an IT project should also be subordinate to its needs and taking into account the perspective of business value. Hence, the question arises about the value of the features of the expected software. In Agile projects, when the software is delivered in iterative-incremental cycles, it is possible to decide on the order of delivered features, depending on their value to the customer. While the business value of a software investment calculated as a whole can be conducted with the use of methods accounting investment profitability, splitting this value between each release of the software in the following increments is a separate issue. The problem lies in the search for relationships between different parts, and in delivering them to the expected customer value and correct sequence from the viability point of view. The issue of relationship between releases is solved through the use of prioritizing. The priority is determined in direct relation to customer value.

In Agile projects, where elimination of all kinds of waste is necessary, the priority value should be determined quickly and cheaply, without involving complex and expensive methods. Commonly used are simple and often subjective methods, which is not an obstacle, but a benefit, as it is the customer who determines the final priority value, best from his point of view.

5 Strengthening Value Managements by Connecting Agile and Lean Approaches

There is a natural relationship between value and need: if something is needed, it is because it has value. Therefore, intending to introduce the concept of eliminating unnecessary software features, the basic principles of Agile and Lean approaches will be briefly discussed. These approaches have one common goal: to increase customer satisfaction. The guiding principle of the Agile Manifesto: "Our highest priority is to satisfy the customer through early and continuous delivery of valuable software" clearly and explicitly indicates the key element of this approach [6].

The concept of value has been highlighted in the Lean Thinking approach, which adopted the following basic principles in production management: Value, Value Stream, Flow, Pull and Perfection [7]. The notion of value was clarified by the introduction of its opposite—waste; hence, "everything that is not value is waste" [7, p. 14]; the principle of waste elimination was accepted as the basis of all activities leading to improved customer satisfaction. Lean Approach was adopted to project management in the wake of production systems, since software development processes evolve similarly. Changes made since the second half of the 80's have resulted in the establishing of first the Agile approach, whose basis[2] is consistent with Lean ideas to a degree that led to creating the concept of Lean Software Development.

For an Agile project to be considered Lean, it is necessary to adopt the five Lean principles into project management [7, p. 13]:

1. Identify value,
2. Map the value stream,
3. Allow the continuous flow of activities identified as value stream,
4. Enable value pull to the customer,
5. Seek perfection.

Hence, defining and understanding customer value has become a key element in Agile projects. [8, p. 80]. It should be emphasized that rules such as "do one thing from beginning to end," drop "batch-and-queue" processes, use the advantages of the "pull" over "push" and the requirement of action transparency for the whole team and all project stakeholders are successfully implemented in Agile projects [1].

Following the authors of Lean Thinking [7, p. 549, 9, pp. 19–20] slightly modified seven categories of waste occurring in IT projects were introduced to Lean Software Development concept. Given the problematic issues noted above, two of those categories should be highlighted: the second one, "over-production of useless goods", called by the authors of Lean Software Development the "extra properties",

[2]Agile principles were modeled after the pillars of the Toyota Production System relating to processes, people, tools and technologies: 1. Individuals and interactions over process and tools. 2. Working code over comprehensive documentation. 3. Customer collaboration over contract negotiation. 4. Responding to change over following a plan [6].

and the eighth one, "design of goods and services that do not meet customer needs," which was omitted [1, p. 74] because it includes this type of loss that is associated with unnecessary software features that—as a result of changes in business or poorly recognized needs—have become unnecessary [10]. Given the high cost of creating such outdated or extra features, it can be said that both software functionalities that do not match current needs and redundant software functionalities lead to the production of overly expensive software, or losses stemming from creating unnecessary applications.

6 Product Backlog and Business Value as Primary Priority Factors

The traditional set of functional and non-functional requirements as the basis for software code in the Agile approach is called a product backlog. Its elements should correspond to the actual work to be done, that is, above all, the current needs of the client, but also other elements, necessary from a technical point of view [2, p. 235, 11, p. 79]. Depending on the complexity of the application, backlogs can vary in sizes, but it is recommended that they do not exceed more than 100–150 items [2, p. 331].

Product backlog is subject to a two-way action: developing and updating [11, p. 79]. The whole team, and if necessary also other stakeholders, are involved in its development. The backlog is created on the basis of a product roadmap or vision. Developing means adding new elements, and raising their level of detail from a predefined set to explicit quantitative indicators, called acceptance criteria. Each new item must be analyzed according to the evaluation criteria and then positioned or prioritized. Items arising from the technical point of view of the developer are also added to the backlog. Items added by the developer as facilities or technical innovations, unreported by the user yet raising the cost of development, may also appear on the backlog. Updating requires item organization based on the removal of items already implemented or those who completely lost value, as well as making changes, most often resulting from customer's needs.

In the context of the problem under consideration, an important activity associated with each backlog, regardless of its level of detail, is to organize all the items by priority. Priority refers to both high level positions, corresponding to epics, essays, or themes,[3] or medium or low level positions, corresponding to medium or small user stories [11, p. 80]. Prioritizing is a rather complicated decision-making process, during which criteria such as value, cost, uncertainty and risk and technological dependencies [8, pp. 80–86] undergo examination. Quantitative or semi-quantitative methods, e.g. Wiegers method [12] are possible to apply. The

[3]Terms vary depending on team customs or adopted method, e.g. Scrum.

responsibility for determining the priority lies on product owner as a representative of the business.

Priority determines the order of providing the customer with operating software gains. In the case of many high-priority backlog items constitutes a threat to supplying what the customer needs the most at a given time[4] [11, p. 87]. Consequently, it is necessary to provide an additional factor, enabling the elimination of such ambiguities and support of informed decisions about priority. It is believed that the elimination of unnecessary backlog items should include features without which the application will work properly, and contain desirable features in sufficient quantities, as well as those expected, the lack of which causes the inevitable dissatisfaction of both customer and user.

It is assumed that in the case of technical restrictions, unknown to the business side of a project, the developer can participate in the process of prioritization. In such a case, a two-way exchange of information is necessary, leading to the understanding of the decision-making problem, which should not be solved alone neither by the customer, nor by the developer. Only after the values of these restrictions are known and considered by the customer can he make an informed decision about how to qualify the features or how to eliminate them.

Apparently, in difficult decision-making cases, leaving the decision to the customer alone results in 85 % of backlog items receiving high priority, 10 % medium, and only 5 % low [16], which certainly does not facilitate the management of requirements. For rapid software development, Steve McConnell proposes to filter these requirements by eliminating those non-essential [13].

7 Applying the Kano Model to Classification and Elimination of Product Backlog Items

In order to solve the problem posed above, the following assumptions were adopted:

- simplicity of the method,
- possibility of good communication between stakeholders
- language understandable to all stakeholders,
- identification of application features unnecessary to the studied segment of users.

The Kano model allows for the classification of product features into three specified categories related to varied customer satisfaction. As a result, it is possible to specify which of the described product features immediately affect the increase or decrease in customer satisfaction. The upper-left quadrant of the model (Fig. 1) presents the fastest growth of satisfaction—these are so-called attractions. In

[4]Use case applied in UML language may be another artifact.

contrast, the slowest increase in satisfaction, further limited to a maximum average level, is shown in the lower-right quadrant. The middle part of the diagram shows a linear increase in satisfaction, which means that the more such features, the greater the chance for satisfaction growth. The model shows that various strategies in the pursuit of customer satisfaction are possible, but it should be a conscious decision of the client.

Learning about customer and user preferences in relation to meeting their needs requires an interview. Two questions for each of the tested features are asked, and then, based on the analysis of the distribution of their assessments, a category is assigned to each feature. Research questionnaire for this study was constructed with the principle of assessment of satisfaction level from two opposite sides: the presence and absence of a quality [2, pp. 79–120]. In each question, the respondent is to choose one out of five options on an assessment scale:

1. I like it
2. I expect it
3. I am neutral
4. I can live with it
5. I dislike it

The evaluation of these answers based on the analysis of both sides, giving 25 combinations, allows to assign the studied features into three model categories:

- must have
- linear—linearly correlated with user satisfaction, the more/less the better/worse
- exciter—providing great satisfaction.

The other possible combinations of response pairs are further assigned into the following three categories: reverse, questionable and indifferent.[5]

8 New Template for User Story Narrative and Evaluation

The above method is suggested for categorization of product backlog items to reduce the burden of establishing their priority, and, above all, to eliminate unnecessary features which will be considered waste in the development process. Product backlog items are written on slips of paper using a specific template and natural language, which ensures that the text of user story narrative is clear to all project stakeholders. In accordance with the Lean principle of transparency, visualization of work is achieved with the use of a system known as “card walls” [15].

[5]Description of matrix interpretation rules for cross-reference assessment of functional and dysfunctional questions used to categorize the responses and to calculate the indicators of their distribution can be found in [14, pp. 79–120, 2, pp. 114–117].

Table 1 Structure of the new user story template, combined with the user assessment question. *Source* Author

First part of user story phrase	First part of Kano method question	Second part of user story phrase	Second part of Kano method question	Response scale	Mark the answer with an X
As a <type of user>	*If I can or If I can not*	*I want <some goal> so that <some reason>*	*How do I feel?*	*I like it*	
				I expect it	
				I am neutral	
				I can live with it	
				I dislike it	

In order to apply the Kano method to product backlog, it is proposed to extend the user story narrative template presented below [2, p. 239]:

$$As\ a < type\ of\ user >\ ,\ I\ want < some\ goal >\ so\ that < some\ reason > \quad (1)$$

where:

- *type of user*—the person (role), which expects the product, profit,
- *some goal*—awaited feature, property of the product,
- *some reason*—profit or value of the feature

with the two questions posed to the respondent, as well as the scale of possible answers to choose from. Due to the discrepancies in personal forms used in the user story template and Kano questions, it is proposed to standardize them and adopt the forms characteristic of user story, substituting "you" with "I" in Kano method questions. In case of functional questions the full phrase will be "if I can," and in dysfunctional—"if I cannot"; the phrasing of Kano questions will also gain a second part: "how do I feel?". The extended template for both functional and dysfunctional questions is presented in Table 1. The structure of these questions is created in natural language and does not contain technical terms incomprehensible to the customer.

As a representative of the business, the owner of the product who manages product backlog should ensure that an opinion poll is conducted among a substantial group of users in the course of writing the content of user stories. After completing the records in line with the specified template (Table 1), it is possible to analyze the distribution of the responses. The outcome is a list of registry items classified into one of six Kano model categories. This allows to take a reasoned decision on the elimination of unnecessary user stories, in line with the adopted elimination strategy. An example strategy for user stories from the "linear" category may read as follows: items with an distribution index value <90 % should be eliminated, in the "questionable" category elimination should apply to all user story regardless of the distribution index value.

Table 2 Usage example of the proposed template. *Source* Author

First part of user story phrase	First part of Kano method question	Second part of user story phrase	Second part of Kano method question	Response scale	Mark the answer with an X
As a lecturer (1)	*If I can*	*I want* to change the allocated course *so that* the schedule is correct	*How do I feel?*	*I like it*	*X*
				I expect it	
				I am neutral	
				I can live with it	
				I dislike it	
As a lecturer (1)	*If I can not*	*I want* to change the allocated course *so that* the schedule is correct	*How do I feel?*	*I like it*	
				I expect it	
				I am neutral	
				I can live with it	
				I dislike it	*X*
As a lecturer (2)	*If I can*	*I want* to watch course allocations on a graphic table *so that* the search is faster	*How do I feel?*	*I like it*	
				I expect it	*X*
				I am neutral	
				I can live with it	
				I dislike it	
As a lecturer (2)	*If I can not*	*I want* to watch course allocations on a graphic table *so that* the search is faster	*How do I feel?*	*I like it*	
				I expect it	
				I am neutral	
				I can live with it	
				I dislike it	*X*
As a lecturer (3)	*If I can*	*I want* to accept all course allocation *so that* I am not surprised with it	*How do I feel?*	*I like it*	*X*
				I expect it	
				I am neutral	
				I can live with it	
				I dislike it	
As a lecturer (3)	*If I can not*	*I want* to accept all course allocation *so that* I am not surprised with it	*How do I feel?*	*I like it*	*X*
				I expect it	
				I am neutral	
				I can live with it	
				I dislike it	

The topic "Effective course planning" serves as an example that illustrates the concept of creating an enlarged structure of the user story template (Table 2).

Analysis of the responses conducted in line with the rules of the cross-reference matrix in Kano model indicates that the user story no. 1 belongs to the linear category, user story no. 2 to must-have category, while no. 3 goes in the category of questionable assessment, suggesting a mistake in assessment process the presence or lack of interest in such a possibility.

Therefore, a simple analysis of responses from representative users will allow for a reliable categorization of their requirements, while the product owner responsible for prioritizing product backlog items can make a correct decision that follows the strategy arising from the business needs. This solution will allow a release of products in different versions, depending on the actual needs of users.

9 Conclusion

Reinforcing the principles of Agile approach through Lean and the establishment of Lean Software Development has become the source of the presented concept, focused on early elimination of unnecessary software features. Waste of time and resources on redundant or obsolete features of IT solutions is not infrequent, but quite common—as shown in references and in author's pilot study. The tools existing in the studied approach have been used to design an attempt at solving this problem. Simplicity, both of the Kano model and of the user story template structure, customer-friendly environment of the Agile approach which facilitates the required communications between customer and developer, the "card walls" method of work visualization and, above all, the guiding principle of waste elimination should all be considered favorable circumstances in the application of this project.

It is emphasized that the two aspects of product backlog management—the development and update of backlog items with simultaneous value assessment—combined in this concept are subject to the principles of Agile and Lean approaches. Their aim is to reduce the amount of work in progress and the opportunity to collect records, which consequently leads to a waste of time on unnecessary product features and failure to provide the software valued by customer on time.

However, the proposed extension in the structure of user story recording with the assessment question on individual user experience requires further testing, in order to verify its effectiveness and possible nuisances in practical application. It is believed that teams accustomed to the currently used user story template will not be surprised by this extension, and when the extended recording structure will become generated automatically by Agile project management tools, it will become a practical way to support the elimination of waste. Employing computers to aid this solution will open up to a much easier use of the more complex structure, at the same time providing documentation of its practicality, which will be the target of further studies of the author, and will allow to gain confidence in the solution.

References

1. Poppendieck, M.T.: Implementing Lean Software Development. From Concept to Cash. Addison-Wesley, New York (2007)
2. Cohn, M.: Succeeding with Agile. Software Development Using Scrum. Addison-Wesley, Michigan (2010)
3. PN-EN ISO 9000:2001: Systemy zarządzania jakością. Podstawy i terminologia. Polski. Komitet Normalizacyjny, Warsaw (2001)
4. http://www.uio.no/studier/emner/matnat/ifi/INF5181/h11/undervisningsmateriale/reading-materials/Lecture-06/04ZubrowISO25000SWQualityMeasurement.pdf. Accessed 20 July 2016
5. https://en.wikipedia.org/wiki/Shareholder_value#Shareholder_Value_Maximization, Accessed 20 July 2016
6. The manifesto for Agile software development. www.agileAlliance.org. Accessed 20 July 2016
7. Womack, J.P., Jones, D.T.: Lean Thinking—Szczupłe Myślenie. Eliminowanie marnotrawstwa i tworzenie wartości w przedsiębiorstwie. ProdPress, Wrocław (2008)
8. Cohn, M.: Agile Estimating and Planning. Addison-Wesley, Michigan (2006)
9. Womack, J.P., Jones, D.T., Roos, D.: The Machine That Changed the World. Rawson Associates (1990)
10. Hołodnik-Janczura, G.: Kategorie marnotrawstwa w projektach IT. Ekonomika i organizacja przedsiębiorstw, No. 9, pp. 31–40 (2013)
11. Pichler R.: Zarządzanie projektami ze Scrumem. Twórz produkty, które pokochają klienci, Helion (2014)
12. Wiegers, K.E.: First Thing First: Prioritizing Requirements. Software Development, September 1999
13. McConnell, S.: Rapid Development. Microsoft Press (1996)
14. Cohen, L.: Quality Function Deployment. Addison-Wesley Pub. Comp, USA (1995)
15. Anderson, D.J.: Kanban: Successful Evolutionary Change for Your Technology Business. Blue Hole Press, Washington (2010)
16. Wiegers, K.E.: First things first: prioritizing requirements. http://www.processimpact.com/articles/prioritizing.html. Accessed 20 July 2016

Part IV
Manufacturing Systems

Comparison of Discrete Rate Modeling and Discrete Event Simulation. Methodological and Performance Aspects

Jacek Zabawa and Edward Radosiński

Abstract The best known modeling approaches used in the simulation are discrete event simulation (DES) and system dynamics (SD). Discrete rate modeling (DRM) has been proposed by the Extendsim software manufacturer and combines both approaches but also provides new opportunities. The purpose of the paper is to compare implementation issues in DRM and DES approaches. The first part presents the use of the simulation package Extendsim for all approaches. The next part presents the assumptions of several models developed to compare the elements of methodology for discrete event modeling and discrete rate modeling. The one source and one-way structure as well as structures with many sources, streams and discrete attributes in both approach were proposed. The results of models were experimentally evaluated and discussed. Performances of approaches were compared. We pointed out the benefits of the DRM approach. DRM is particularly useful in high-intensity input streams.

Keywords Discrete rate modeling · Software performance · Hybrid system · Simulation modeling · Manufacturing systems

1 Introduction

In simulation approach we design a model of the system, conduct experiments with the model and observing the behavior of interesting elements. The model offers in-sight into the operation of real system under different conditions and test a variety hypotheses for a fraction of costs that one would ordinarily incur if testing was per-formed on a real system. Simulations (individual simulation runs) are

J. Zabawa (✉) · E. Radosiński
Department of Operation Research, Wroclaw University of Science and Technology, Wroclaw, Poland
e-mail: jacek.zabawa@pwr.edu.pl

E. Radosiński
e-mail: edward.radosinski@pwr.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_12

conducted in a simulated time, which is an abstraction of real time. One of the main advantages of the modelling procedure is that we can start with simple approximation of the process and gradually refine the model with the progress of understanding of the process as well as the changes in the structure of the process. Incremental model improvement enables sufficient (but approximated) solutions to very complex problems very quickly. On the other hand, adding improvement to the model allows more accurate imitation of the real processes.

The simulation approach with software packages (for example Extendsim, a simu-lation software that has been used for research in this paper) [7] is performed in the following manner: instead of running expensive tests on a real system we can con-struct the logic of the model – corresponding to real system in two dimensions: first, the model structure (relationship between components of the model) and second, relationships between model input and output. The simulation model may therefore imitate the action or the dynamics of the system and then aspects of a selected range of model parameters. Suitability of simulation approach in decision making involves risk reduction, reduction of uncertainties and accelerating of the delivery of research material.

2 Modeling Approaches

2.1 Classification of Methodologies and Modeling Techniques

The fundamental criterion for the classification of simulation methodologies is the manner in describing the system in the models [7]:

- Continuous; in the Extensim software implemented through the blocks from the "Value" library
- Discrete Event System/Discrete Event Modeling; in the Extensim software implemented through the blocks from the "Item" library
- Discrete Rate Modeling; in the Extensim software implemented through the blocks from the "Rate" library

Three main methodologies, mentioned above, are used in a variety of modeling and carrying experiment techniques:

- Monte Carlo
- Agent Based
- Mixed, hybrid
- Various types design of experiments (in this work we used experiments with constant steps)

2.2 Continuous Approach

In continuous approach we assume that time step is constant (in the specific experiment), therefore of the state variables and output values can change at regular intervals only: a simulated time changes abruptly. The input values may be subject to changes as planned experiment. We use continuous approach for carrying out calculations based on the results of other approaches and preparing input data (including initial data) using blocks from the "Value" library (Extendsim). However, we do not use the method of integration over time (accumulation of input values), occurring in comprehensive continuous models. One exception may be using information about the size of collectivity of something, for example: cohort in population (calculated in accordance with the approach of continuous) to control data input module for example discrete event modeling the demand for services whose volume is dependent on the size of the age cohort. In the initial stage of this system dynamics approach (continuous approach) structure of the modeled system is mapped using a loop connections of information and decision: loops can be of positive feedback, negative feedback or neutral [11]. The basic elements of this approach are the resources, flows and arithmetic equations. At the introduction stage to the continuous methodology most frequently used examples are the water reservoir together with the input and output streams (flows) [1].

2.3 Discrete Event Approach

In discrete-event simulation approach (DES) model changes state only when an event occurs. The passage of time does not per se a direct impact on the model. The time of an event occurs in a model is determined by the order of previous events and the status of model elements. In contrast to the continuous methodology, simulated time progresses from one event to another. Constant interval between events is possible when the model has fixed intervals between input events (induced into the system under test environment), or the characteristics of delay elements in the system is de-terministic and intensity of the input stream is higher than the possibility of its service (the existence of bottleneck). Usually, however, in non-trivial models, time interval between events is not fixed.

2.4 Discrete Rate Approach

The need for this approach emerges when a large number of fast moving objects in a process for reasons of efficiency calculation fails discrete approach. The situation is similar when instead focus on modeling the flow rates (like in continuous

approach) we consider the modeling of discrete control objects. This causes problems with the adoption of units of measurement and calculation accuracy [6].

Discrete rate modeling (DRM) approach from the developers' point of view is suit-able for modeling high-speed streams of uniform matter or non-uniform raw materials, products, processing, mixing and separation of the various items of a "continu-ous", e.g. different kinds of ores [2], liquid [12], chemical processes [10], data streams - for example in computer networks [4].

In this approach, we also often need to take into account sudden changes in flow rates caused by e.g. achievement of a certain level of filling of the so-called "tank". In the DRM approach we define a constant stream structure "liquid" which flows through "tube" but the flow rate (may be even equal zero) and the route can vary when a specific event occurs. In the case of the handling of individual variants of the structure we have to develop and run separately created models.

DRM approach reminds some aspects of DES approach because the intensity and even the flow direction may vary as a result of the occurrence of a specific event and relevant calculations are performed only when something changes within the model. On the other hand, it is much easier to continuously observe the current state of the stream and the resource (level) as well as the system dynamics (continuous approach) models than e.g. the degree of completion of the operation on the service station in DES models. Problems using DRM approach rely on the need to use a fairly high degree of abstraction and aggregation laborious process descriptions [9].

The basic elements of the models in the DRM approach are flows (rates) and levels (tanks). They recall the concept of system dynamics, but they are differ e.g. way of flows controlling (based on the assumption of maximum flow at the assumed constraints, i.e. that connections are always filled by "things" which are subject to "pressure") [8] and the opportunity to take the presence of additional characteristics or attributes (numeric or text). It is also possible to combine parts of the model con-structed in accordance with DRM approach and modules that use DES approach (interaction between movements "ongoing matter" and the discrete objects). Loading and unloading operations can be combined with system dynamics: accumulating (summing) the calculated values of "integrating the output over time" and calculation of the source data coming from the DRM module.

2.5 Modeling Techniques

In the Monte Carlo technique we generate repeated random numbers and use they to assign values to input parameters. The series of simulation results created in this way are undergo statistical processing. Monte Carlo technique is used in organizing both experiments in models which not take into account the impact of real time (e.g. modeling investment performance in oil pool) as the division experiment into separate fragments (e.g. calculation of aggregated statistical parameters such as mean and variance).

In the agent-based approach basic elements of the model are individually active actors (agents), equipped with a set of rules that describe individual behavior (the model is composed of many agents belonging to specific classes). Usually it is assumed that local agents can influence the behavior of its neighbors (this is described in the rules). Agent has a certain degree of autonomy (storage, interaction with the environment, information about the values of its parameters - including the position in the coordinate system using the concept of context and may even be programmed with intelligent behavior and a corresponding learning ability).

3 Introduction to Simulation Modeling Techniques in Extendsim

In the current section implementation of very simple and basic models in Extendsim are discussed. These models were built using the most known simulation approaches: continuous and discrete-event. Implementation consistent with DRM approach will be presented in the next section.

Extendsim modeling consists in placing in the created model blocks with ready-made (or created by user) libraries, assigning the initial parameters to the dialog fields in blocks, connecting blocks via connectors of various types (input and output, continuous, discrete or discrete rate) and - if necessary - the construction of mathematical equations and programs in the form of text. Libraries may contain elementary blocks (described in this paper) as well as groups of blocks (hierarchical blocks), created using graphical tools (drag and drop), according to the object-oriented approach (from software engineering domain). Each of the approaches has a separate dedicated library, while the model can use blocks of different libraries and we could create a hybrid model. What's more, Extendsim is equipped with an integrated tool to optimize the parameters of the test model using a genetic algorithm (OPTIMIZER block), as postulated in [3]. It is also possible to introduce own code (in language ModL) in blocks and even to automate the addition of blocks and relations between them.

Continuous approach in Extendsim environment is implemented by the blocks included in the "Value" library. Resource (level) is modeled with a Holding Tank block. This block is equipped with input connectors "InputIn" (input flow control), "want" (output flow control), "init" (starting value), "reset". Output connectors are "Contents" and "Get". There is no dedicated block designed for modeling stream: the concept of flow is represented by a mathematical equation by which information is sent to the control connectors. In the Fig. 1 we can see graphical representation of model resulting in the S-shaped grow – one of the patterns of behavior [5]. We can see two blocks HOLDING TANK (Potential Customers and Actual Customers), four CONSTANT blocks and two MATH blocks (with Multiply and Subtract icons).

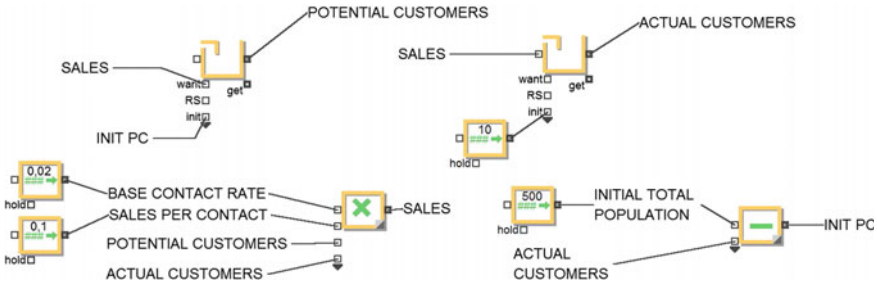


Fig. 1 An example of system dynamics model in Extendsim

Discrete event modeling in the Extendsim is realized by blocks included in the "Item" library, and usually supplemented by the blocks from the library "Value" for input and output data processing and from library "Plotter" for plots or charts making. Basic blocks of the "Item" library used to model the following components: source, object generator at intervals: "Create"; queue or the buffer: "Queue", delay, server: "Activity", exit, erase objects: "Exit"; resource objects: "Resource Item".

In the Fig. 2 is a schematic model of the system consisting of source object (Create), resource objects from the block delaying the arrival of the object (Activity 1), buffer (Queue), the service station (Activity 2) and system output (Exit) supplemented by a block for combining two sources (Select Item In - Merge) and a random number generator (Random Number) (for delay time on the Activity block).

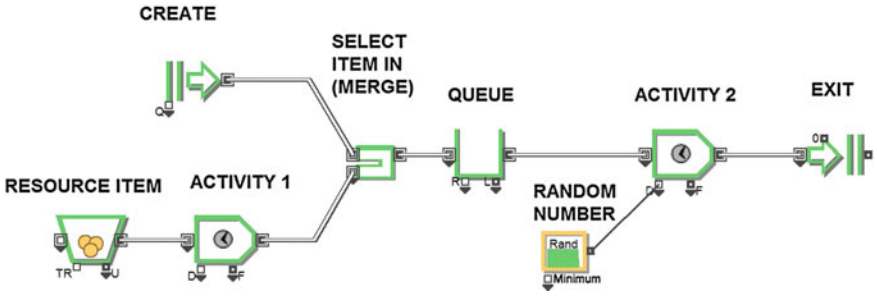


Fig. 2 An example of discrete event model in Extendsim

4 Proposed Assumptions and Implementation of the Models

4.1 Simple One-Way Model

Assumptions for the first test model are as follows. The model shows the one-way flow of content (fluid) from between the two reservoirs. It was assumed the presence of an intermediate tank and convey flow. Initial data contents of the tanks are as follows: TANK1-10,000 units, TANK2 - a zero units, TANK3 - 10,000 units. Maximum flow between TANK1 and TANK2 (VALVE maximum rate): 100 units per time unit. CONVEY FLOW is 1 slot operation, the residence time in the CONVEY FLOW is 1 time unit. The duration of the simulation run: 100 time units. We presented two equivalent models: in the Fig. 3 is DRM model and in the Fig. 4 is DES model. In the Fig. 5 we present output of the experiment (range of horizontal scale 0–100 time units: and vertical scale 0–11000 units). Under the assumptions (high-bandwidth from TANK1 to TANK2, low-bandwidth from TANK2 to TANK3), 10000 resource units should flow from TANK1 to TANK2. TANK1 should become empty, final state of TANK2 should be just below initial content TANK1. TANK3 should increase by this difference (reduced by the current contents of CONVEY FLOW). Figure 5 confirms these predictions - it should be noted that it has been made on the basis of the above Tank's state checked at fixed periods.

The both models have almost the same module (consistent with DES approach) enabling the registration of observed values of variables (contents of TANK1.. TANK3) and developing an appropriate graph (blue-TANK1, red-TANK2, green-TANK3). Conclusion: in the analyzed system, we notice the presence of

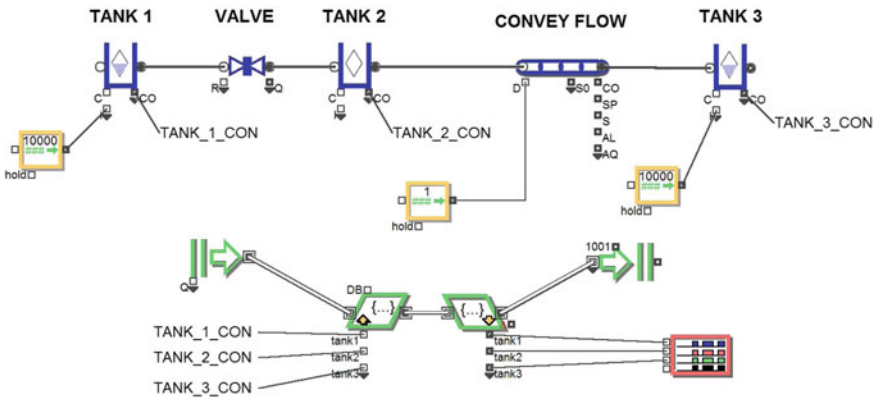


Fig. 3 Simple one-way flow model. An example of discrete rate modeling approach

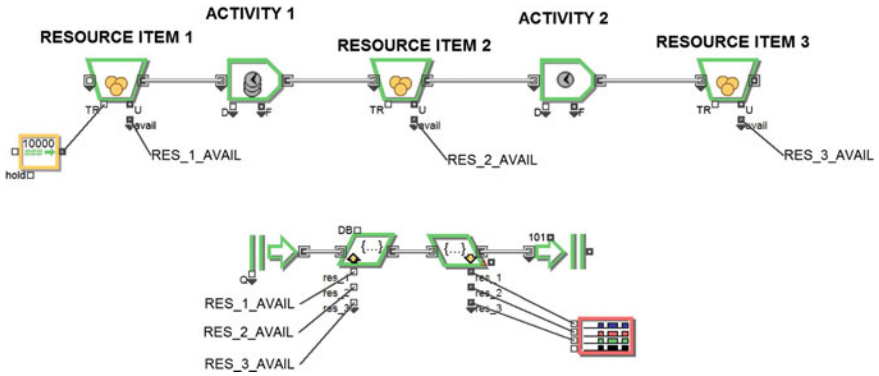


Fig. 4 Simple one-way flow model. An example of discrete event modeling approach

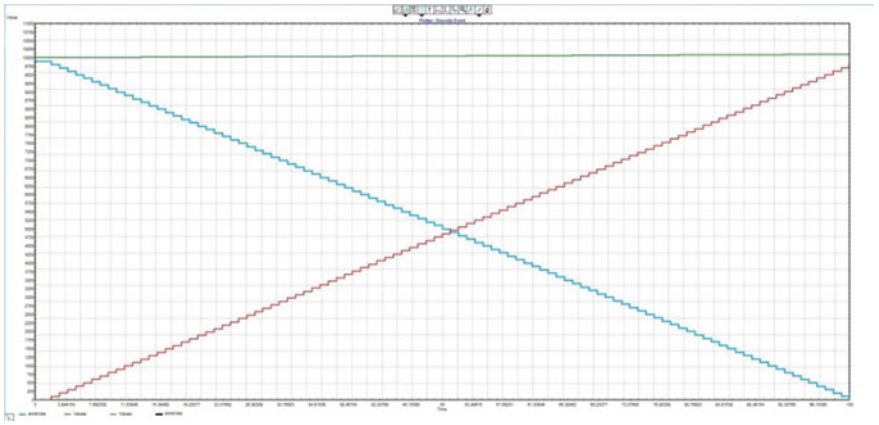


Fig. 5 Simple one-way flow model. Experiment output from DRM approach. *Blue curve* on the graph indicates contents of Tank1 and *red curve*—Tank2

bottlenecks at CONVEY FLOW, thereby TANK3 state increased by approx. 100 units. We tested also the execution time of 100 runs for several variables of TANK1 initial contents in the model made in accordance with the DRM and DES approaches. Then we increased throughput of the VALVE and ACTIVITY1 in proportion to the initial (maximum) TANK1 content or the number of objects RESOURCE ITEM 1. The results are shown in the Table 1. There are visible: the independence of computing time on the parameters of the model in the approach of DRM and rapid increase in computation time in the approach DES (with increasing intensity of objects).

Table 1 A comparison of experiment duration in ticks (Extendsim internal duration time units)

Initial contents TANK1/RESOURCE ITEM1, VALVE/ACTIVITY1 = 100	DRM	DES	Proportion to initial number/contents	DRM	DES
10000	9.09	12.45	10000	10.05	118
20000	9.24	12.39	20000	10.39	489
100000	9.17	12.31	100000	10.13	extra - large
200000	9.85	12.34	200000	10.22	extra-large

4.2 Attribute Modeling in Both Approaches

Assumptions for the second model are as follows. The model shows the use of flow control "things"/objects between a number of multi-stream reservoirs in the struc-ture. In the Fig. 6 we presented attributes using in DRM approach and in the Fig. 7 - DES approach. In these (two) models, there are two input streams starting from TANK1/RESOURCE ITEM 1 and TANK2/RESOURCE ITEM 2, respectively. The first source has almost unlimited capacity and infinite content (DES 10,000 objects). This is a homogeneous source from the text attribute point of view: block SET (R) 1 sets value (RED) but the value of the numeric attribute "att1" increase twice during the simulation run, from 1 to 2. Maximum flow of stream arise from the settings VALVE block of 10 units in a unit time. The first stream is then mirrored in the block DIVERGE in Unbatch operating mode. This enables us to record in the block TANK 3 stacked flow of "things"/objects from the first stream

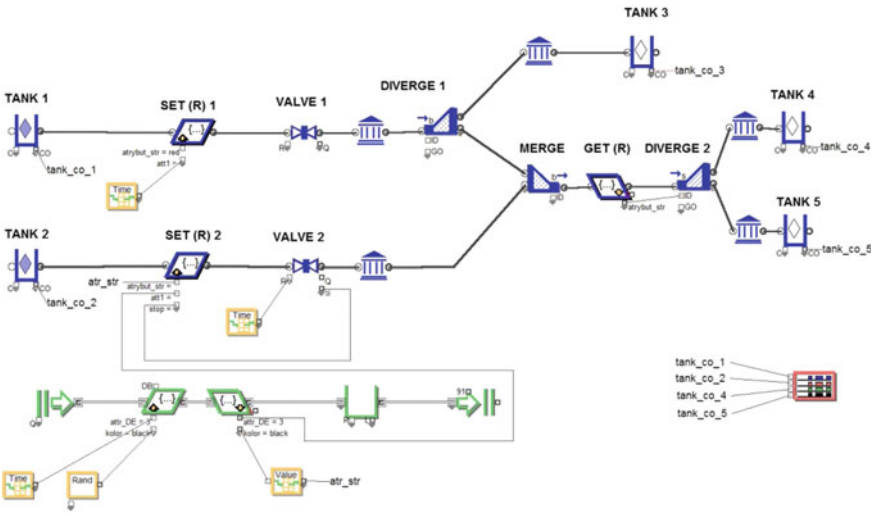


Fig. 6 The model with attributes. An example of DRM approach with DES module

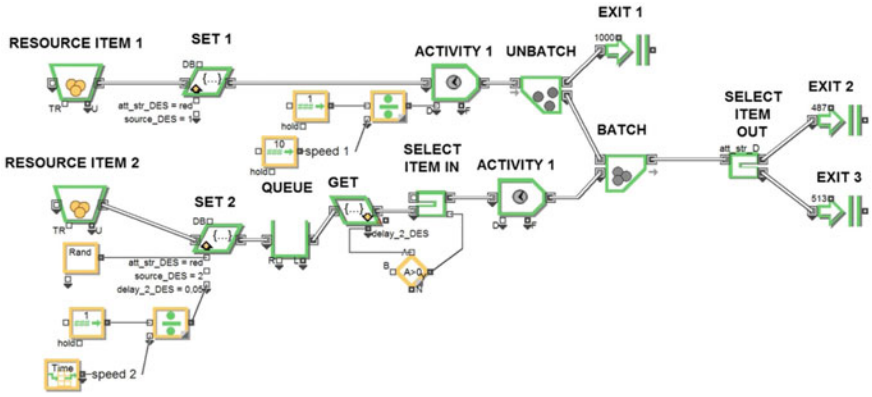


Fig. 7 The model with attribute handling. An example of pure DRM approach

and mix (a copy of) the flow of things/objects with the second stream and recognize the characteristics of the mixture. The second source of stream is represented by TANK2 and it is heterogeneous: the text (string) attribute may take one of the two values set (red/black) and its value is generated in "random number" block in accordance with a DES module. It is worth to pay attention to the texts as "atr_str" These are the so-called "named connections" that enable the transmission of objects, signals and flows between parts of the model without the tangle of connections, obscuring picture.

This is a hybrid model in which the parent process is managed according to DES approach. The value of a numeric attribute "att1" changes leaps and bounds twice during the course of the simulation in the range 3–4. With this combination of attribute values "att1" we can identify components of the mixture when combined streams of the first and second block MERGE. The second stream is assumed that the maximum intensity of the stream resulting from the VALVE block settings will vary during the run (this could be due to the changing availability of labor throughout the day) increments in the range 0–20. The real flow stream is then stored in an additional value attribute "stop" (the link between the blocks VALVE 2 and SET (R) 2. The rule creation mixture (block MERGE) is a Batch, means that the same "stuff"/objects must come from both streams that the flow was sent to the rest of the model. The purpose of blocks GET (R) and DIVERGE (2) is the separation of the mixture on the basis of the text attribute (source of the stream therefore has some impact on reaching which the TANK 4 or TANK 5). In the DES version of the model we have full information about the characteristics of objects, because the objects retain their individuality even after passing through the block "Select Item In" or a pair of blocks "Batch" and "Unbatch". Objects do not lose their attributes even after mixing streams.

In History (R) blocks located between DIVERGE and TANK we can see the arrival times of "stuff" in the streams and the values of their attributes. On this basis one can deduce whether the stream is homogeneous in terms of source of its origin

(because when the intensity of the second stream is equal to zero the value of the attribute "STOP" is equal to 0). Additional information will be available when selecting the appropriate mode of aggregation attribute value "att1" (i.e. "sum only con with non zero-rate" or "sum"). Ten runs (simulations) by 100 time units in the model made in accordance with a DRM (and hybrid control) took 4.55 ticks (duration time units), and in an equivalent (in terms of initial values) 38.7 ticks; one should mention the fact that they had to accept finite (10,000), the number of objects at the initial moment in RESOURCE ITEM 1 and RESOURCE ITEM 2. We used a quad-core computer (AMD Phenom II X4 925, with 2.8 MHz, 16 GB Ram) with Extendsim 9.2 in this study. On the other hand, the DES model does not need hybrid approach: a model was constructed using blocks from the library, and only processing ITEM data input and control of internal communications were accomplished using a continuous approach (VALUE library).

5 Conclusions

The paper presents some practical problems, in which modeling in accordance with the DRM approach is superior over DES in terms of calculation speed and consequently the possibility of using flow models of greater complexity (number of operating steps). We developed a number of assumptions for several models for the purposes of this study. Models were made using a competitive approaches. Some of the models and assumptions were described in this work. It was found that the discrete rate modeling approach allows us to perform simulations much faster, especially when dealing with high flow rate in the considered system objects. Moreover, duration of the simulation rum in DRM approach is independent of source streams intensity and initial resource values. In DES approach these are crucial factors determining the experiment duration and may reduce usefulness of the models if we want achieving results quickly.

It can be argued that DRM should be applied whenever is possible. In contrast, if the model assumptions contain representation of heterogeneous objects - for example flowing through a common channel (this can cause mixing of the objects) - then recommended approach is DES. The next step should be building the equivalent model in accordance with a DRM. In other words, it seems that if you can build equivalents of the model assumptions (implementing the same degree), it will run faster model built in accordance with a DRM. Further work using the discrete rate modeling approach will include hybrid structures of different approaches and submodels in healthcare and accounting.

Acknowledgments Extend blocks copyright 1987–2015 Imagine That, Inc. All rights reserved. This work, *Comparison of discrete rate modeling and discrete event simulation. Methodological and performance aspects*, was financed by the grant from the National Science Centre, Poland, which was awarded based on the decision 2015/17/B/HS4/00306

References

1. Banks, J., Carson II, J.S., Nelson, B.L., Nicol, D.M.: Discrete-Event System Simulation, 5th edn., Pearson, Upper Saddle River, New Jersey (2010)
2. Béchard, V., Côté, N.: Simulation of mixed discrete and continuous systems: an iron ore terminal example. In: Proceedings of the 2013 Winter Simulation Conference (2013)
3. Górski A.: Wspomaganie symulacji komputerowej w zakresie poszukiwania rozwiązań suboptymalnych poprzez integrację z algorytmem genetycznym; Komputerowo zintegrowane zarządzanie. Zbiór prac, T. 1/ Ryszard Knosala. (eds.) (2003) [in Polish]
4. Imagine That Inc. Solutions: Simulation Software Overview, http://www.extendsim.com/sols_simoverview.html (2002–2013))
5. Imagine That inc, Extendsim 9, http://www.extendsim.com/ (2013)
6. Kind, A., Metzler, B.: Rate-Based Active Queue Management with Token Buckets; High-Speed Networks and Multimedia Communications, pp. 176–187. Springer, Berlin (2003)
7. Kirkwood, C.W.: System Dynamics Methods: A Quick Introduction, Chapter 4: Basic Feedback Structures, http://www.public.asu.edu/∼kirkwood/sysdyn/SDIntro/ch-4.pdf
8. Krahl, D.: Extendsim advanced technology: discrete rate simulation. In: Rossetti, M.D., Hill, R.R., Johansson, B., Dunkin, A., Ingalls, R.G. (eds.) Proceedings of the 2009 Winter Simulation Conference (2009)
9. Reggelin, T., Tolujew, J.: A mesoscopic approach to modeling and simulation of logistics processes. In: Jain, S., Creasey, R.R., Himmelspach, J., White, K.P., Fu, M. (eds.) Proceedings of the 2011 Winter Simulation Conference (2011)
10. Sharda, B., Bury, S.: A discrete event simulation model for reliability modeling of a chemical plant. In: Mason, S.J., Hill, R.R., Moench, L., Rose, O. (eds.) Proceedings of the 2008 Winter Simulation Conference, pp. 1736–1740. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey (2008)
11. Simonovic, S.P.: Managing water resources: methods and tools for a systems approach. Routledge (2012)
12. Vasudevan, K.: Simulation of high speed bottle manufacturing lines—software evaluation, techniques and results. In: Proceedings of the IIE Conference and Expo (2009)

Simulation Tool for Effective Tasks Subcontracting in Manufacturing Networks

Joanna Gąbka

Abstract The article presents the simulation tool for effective task subcontracting in clusters, manufacturing networks or competence centers. The simulation model described is a fully functional prototype of an application that serves as decision support system. It was developed on base of knowledge about the specific requirements of the new business structures operation. The proposed tool comprises four modules. The most important one is the Module of Enterprise Selection for Virtual Organisation. The algorithm designed for this module of the system was verified on the real data concerning typical order that may be realized by an exemplary cluster and it proved to be effective and efficient. The big advantage of the presented solution is its flexibility that gives user a chance to define preferences towards criteria set and perform sensitivity analysis for the generated solution. It also offers readable user friendly presentation of the results in form of dynamic visualization elements such as graph and solution tree.

Keywords Simulation · Algorithm · Subcontracting · Manufacturing networks

1 Introduction

Dynamic changes on the competitive market caused development and increasing popularity of a new business models, [1]. Contemporary manufacturing enterprises, to meet client requirements, have to deliver customized products developed on the base of advanced technologies in a very short time, providing high quality and decent price. Gaining strong position in such an environment induces enterprises' managers to operate in such structures as manufacturing networks, industrial clusters and centers of excellence. Recent researches indicate that this organiza-

J. Gąbka (✉)
Faculty of Mechanical Engineering, Centre for Advanced Manufacturing Technologies CAMT, Wroclaw University of Technology, Łukasiewicza 5, 50-371 Wrocław, Poland
e-mail: joanna.gabka@pwr.edu.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_13

tional form are often adopted to enforce single producer potential, flexibility, cooperation competences, innovation level and obviously to achieve broadly understood synergy effect, [2, 3]. The conventional linear sequence of tasks from one order, which used to be performed by one production unit, is now replaced by dynamic links between several manufacturers created only for the purpose of a given contract. Manufacturing networks integrate from over a dozen to hundred organization. Each unit has its competences, capacities, a set of technologies and willingness to use them in a process of value creation for gaining profits. These implies high level of complexity and diversity in field of data gathering, processing and exchanging. Effective operating in the manufacturing networks requires applying dedicated ICT tools. It concerns not only issue of data organizing and unification but especially its processing for the purpose of decision making. A key success factor for the agile manufacturing in cooperation is quick and proper choice of subcontractors. The decision-maker configuring virtual organization (VO) has to consider many variables, conditions and should not do it intuitively. This article presents simulation tool designed and developed to support decision process of proper task assignment to cooperating manufactures with consideration of customer requirements. The implemented algorithms and reasoning rules ensure effective task subcontracting. The simulation tool framework used, enable to implement user friendly visualization of generated solutions and flexibility.

2 Algorithms of Partner Selection in Manufacturing Networks/Clusters

The approaches to the issue of selecting subcontractors in manufacturing networks may be divided into four groups. Each of these methods, regardless of the type of algorithm and selection criteria set applied, relays on the assumption that there exist data base enabling fast acquisition of necessary parameters and variables values.

In the first group, [4–6] one can find selection methods based on the expanded cost function. In this case all factors used in process of configuring VO are transformed into costs. The objective function with uniformed variables simplifies the problem because does not require multi-criteria analysis. The exemplary criteria set consists of the direct manufacturing costs, transportation costs (between the cooperating factories) and cost resulting from delayed or too early order realization. The problem of time is considered in cost category. The disadvantage of this approach is limited criteria set. It also does not analyse issue concerning capacities of factories from the network. However these group of method is relatively easy to implement in practice.

Another category could be made of models using graphs in the subcontractor selecting procedure, [7–10]. The solutions from this group also focus on cost, time and distance (transportation) analysis. They ignore such an elements as risk of the order realisation in the configured VO or subcontracted producers past experience

in cooperation. The big advantage of algorithms based on graphs is readable visualisation of the possible solutions in form of a paths in graph.

The third group is represented by methods applying genetic algorithm in selection process of appropriate producers to the order realization in cooperation, [11–13]. In this kind of methods cost is treated as a constant value fixed previously during network/cluster members negotiation. All potential contractors agree to be paid certain sum of money in case they are chosen to realize part of an order. They are also obligated to declare risk level of exceeding deadline. Criteria set includes risk factor related to the production's unit capacity and required deadline of the entire work completion. The strength of this method is thorough search of the solution space and risk factor consideration. Its disadvantage is not considering costs as variable that would enable extra benefits for a principal. It also requires data about risk which are hard to acquire from a standard ERP/PDM/PLM systems.

The fourth group consists of the methods with expanded criteria set which gives them the biggest flexibility and complexity [14]. The analysed criteria in this case are: cost of the order realization, cost of transport, risk related with the lack of capacity and manufacturer experience in cooperation. The method comprises several objective functions so an additional criteria may be added or some of them can be omitted if there is such a need. The method is also adjustable in field of multi-criteria analysis. It can be either goal-programming or such methods as SMART or SWING depending on the user's preferences and other conditions related with the process analysis.

The proposed tool for effective task subcontracting in manufacturing networks comprises good points remarked in relation to each of approaches mentioned above. It is easy to implement, relies on data easy to obtain from the common ERP/PDM/PLM systems used by companies, enables adjustments in criteria set and offers sensitivity analysis to give the user insight into the solution space. Additionally it offers three options of the generated result visualisation.

3 The Proposed Tool Structure

The proposed simulation tool dedicated to tasks subcontracting is a functional prototype of an application with a structure build from four element (Fig. 1). Two of them standard and commonly used:

1. The Intranet or Internet platform connecting all manufacturing networks/clusters members and saving basic data about the participants and their quantity.
2. ERP/PLM/PDM system used by each cooperating business unit, which registers and manages, among others, data concerning orders.

Beside these two well-known elements the proposed approach assumes implementation of two additional modules:

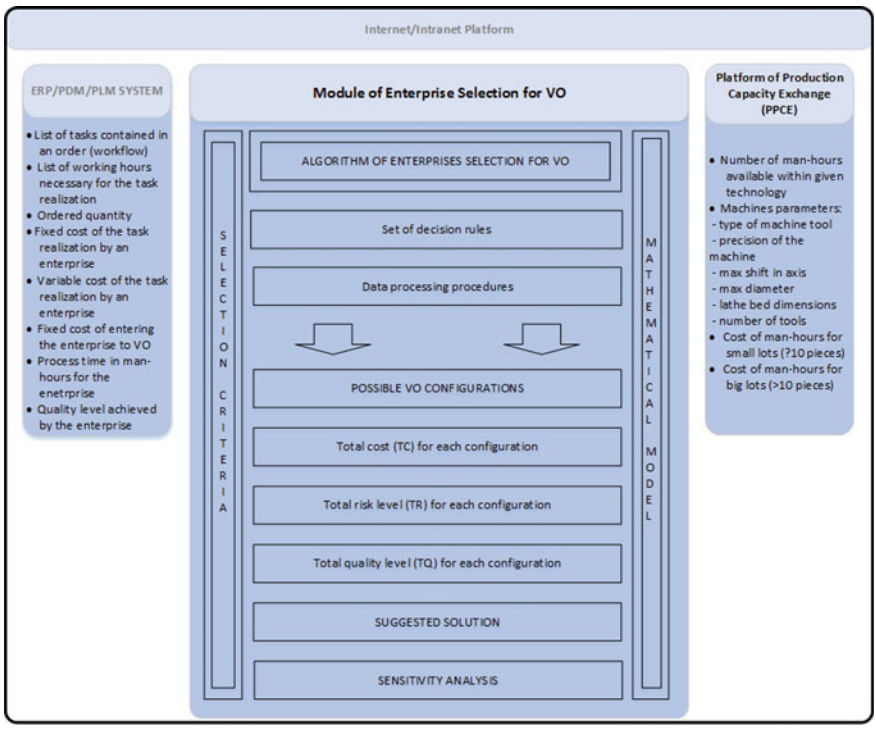


Fig. 1 Structure of the proposed tool

3. Platform of Production Capacity Exchange (PPCE) which registers information about spare capacities of each manufacturer altogether with necessary details concerning machine and tool parameters.
4. Module of Enterprise Selection for VO equipped in the selection criteria, set of reasoning rules and data processing procedures resulting from the dedicated algorithm developed for this module. Module of Enterprise Selection for VO gives the entire application characteristic of an artificial intelligence and enables aiding the decision process.

The module of Enterprise Selection for VO is crucial for the proposed tool. Its development required defining several areas. The framework for its implementation were fixed by choosing criteria set and building mathematical model. The criteria set consists of cost, capacity (qualitative, quantitative), risk factor, quality level, time of the order realisation. The risk level is connected with the failure frequency of machines in given factory and weight of the subcontracted tasks. Weight of the subcontracted task, in turn, depends on two other factors. One of them is competence type needed for the task realisation (standard/unique). The competences are unique if there is no other production unit in the cluster/network which has such a technology to replace the subcontracted in case of unexpected events occurrence.

Second element influencing the weight is relation of the working hours included in a task to the working hours of entire order.

The operation of the algorithm may be briefly described in the following steps:

Step 1 covers acquiring data about the preferences for the given order. The system user has an opportunity to express his/her preference in field of criteria (cost, risk, quality). If any of the criteria is not important for the customer then the weight of this criterion equals 0. The system only lets to choose numbers for the weights which sum equals 1. This information is crucial for the functions implemented in the algorithm and executed in step 8.

Step 2 consists in preparation a space for the further operations. During this phase the system collects list of tasks needed to the order realization and list of potential candidates for each task. It checks which member of the cluster has necessary machine/tools and enough man-hours to realize the task.

In step 3 man-hours are assigned to the potential contractors in accordance to the determined inference rules. The system checks whether there is possibility of realizing one task in one production unit. If it is not possible to the lack of qualitative capacity, then the task is divided between several contractors. The weight of task is an important variable in this place.

Step 4 enables to convert numerical data into graph. This form of presenting solution is assumed to be clear and readable for the system user.

In step 5 the system verifies which of the found alternative configuration does not fulfill necessary condition. This condition is deadline for order realization and general short time of an order realization.

Step 6 enables to see on GUI reduced solution space with the solutions which meet the fixed deadline.

Step 7, the three values: cost, risk and quality will be calculated in accordance with defined rules and conditions. In this place the system gather arguments that enable to asses and rank the configurations. It analysis solutions left after the preliminary elimination made in step 5 in a logical way.

In step 8 the system indicates preferred solution. It is obviously chosen with consideration of the criteria weights determined by the system user in step 1.

Step 9 enables to check what would happen if other values of weights were chosen. This last part of the procedure enable to take a more rational decision and find out the best alternatives in case of unexpected events like machine failure in one of the chosen units, [15].

The algorithm presented in this section enables to effectively use a significant number of data. If the decision process is not supported by any artificial intelligence then the decisions are usually taken intuitively and there is not possibility to explain them logically. Here the solution search is transparent. Additionally the system characterizes big flexibility because user decides which criteria are taken in the analysis and what is their importance. It is big advantage because in practice for some orders the highest priority is quality while for the others cost.

The proper output of the analysis made by the Module of Enterprise Selection for VO is dependent not only on the correctness of the implemented algorithm but

also on the completeness of the database used. The fragment of UML class diagram of database build for the simulation tool is presented in Fig. 2.

The solution tree class is connected with the visualisation of the solution generated by the system. The final result is displayed in form of graphic which is active during simulation and shows progress in the contract realisation (Fig. 4).

4 The Algorithm Verification on the Exemplary Order

The algorithm developed for effective task subcontracting was implemented into the simulation tool build in Anylogic environment. The simulation tool enable to test the function of the developed solution. The verification was conducted on the exemplary order typical for the group of producers from cluster CINNOMATECH integrating manufacturers from the metal processing industry, production of machine parts and cutting tools sector [16]. The order concerns twenty five pieces of the hydraulic cylinders. The ordered item consists of the following parts: cylinder tube (barrel), piston rod, rod wiper, head, flange, gland, end cup. Beside the operations from the operation sheet of each element there are two made on entire product which is assembly and painting. The database contained data about capacities of nine enterprises at the time of testing. The system suggested VO configuration with five companies. The fragment of solution is presented in Fig. 3. The simulation tool displays information about developed solution generated with use of the implemented algorithm in form of tables with information about time

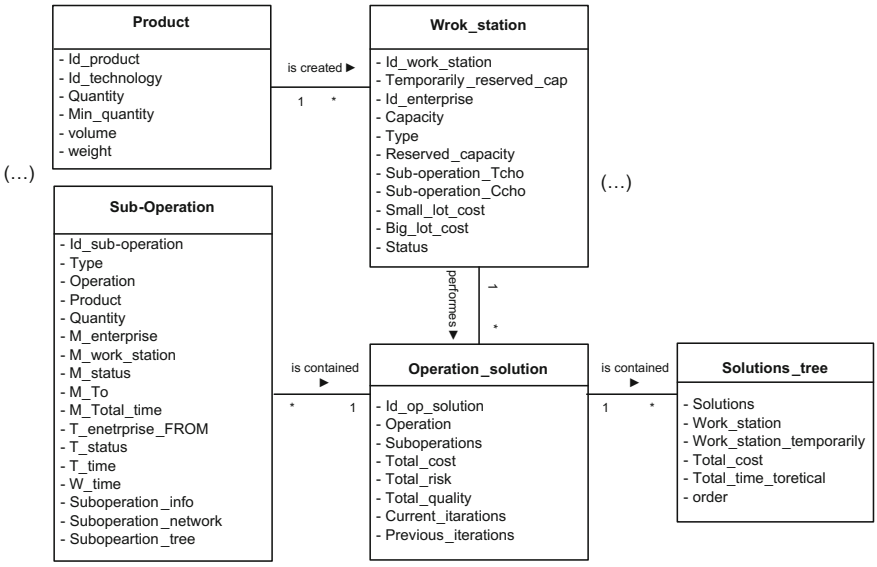


Fig. 2 Fragment of UML diagram of the simulation tool database

when the work should start, number of part to be processed, volume and weight of the piece, time and cost of work as well as manufacturer name and contracted operations list. Beside this standard way of information gathering which requires from the user a lot time to be analyzed the system offers more friendly way of presenting the solution in form of a graph marked on the map (Fig. 3). It helps to observe an issue of transportation between enterprises by showing scaled distances. The simulation tool gives opportunity to observe progress during the process realization in the chosen configuration of VO. As it is visible in the Fig. 3. Manufacturers involved in the order realization are marked by circles. The grey colored indicate operations or sub-operations that are already done. Green circle indicates operations/sub-operations with status Active that are being done in the given moment of time. Black dotted circles indicate operations that are to be done. After the operations in the given location is finished the yellow line suggests that a track is send there to ship the parts to the next manufacturer listed in the solution. All this information are confirmed on the panel on the right side of the map.

Another dynamic element which serves to the solution presenting and monitoring progress of work is a solution tree. Figure 4 shows a fragment of the tree for the analyzed order used in the process of algorithm testing.

The solution tree uses colors analogically to the graph described above. The tree in the Fig. 4 shows that none of the manufacturers available in the database declared enough capacity to make the operation number 90 in one location. In this case system suggested sharing capacity of the three manufactures as it is shown on the tree. All the methods of visualization implemented in the model are compatible and enable user better understand the processes which are occurring during simulation. The simulation itself is obviously a result of the previous analysis of the order made thanks to the algorithm in accordance with the inserted criteria. The

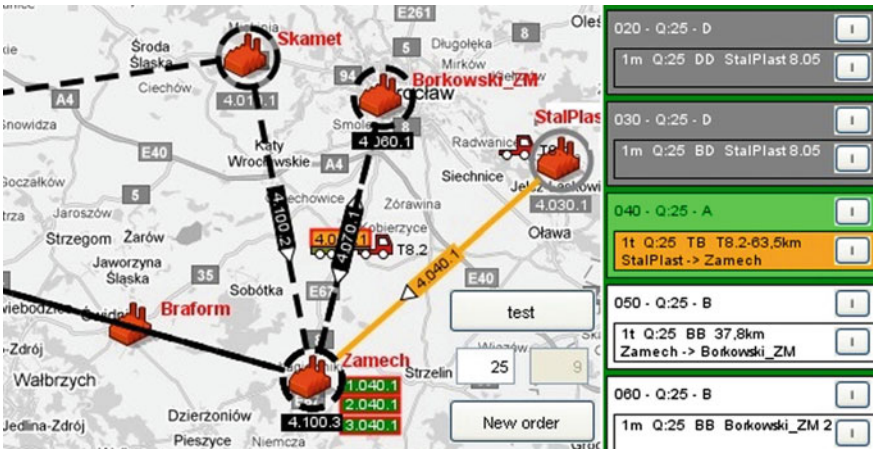


Fig. 3 Fragment of solution visualised in the system

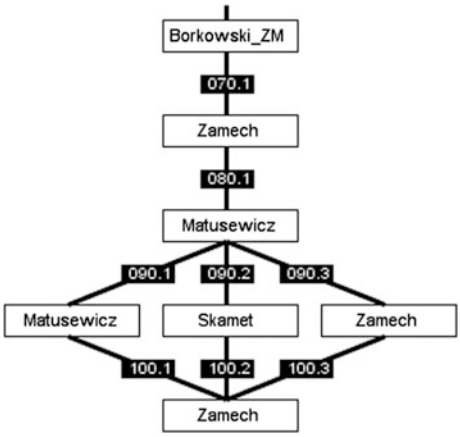


Fig. 4 Fragment of solution tree generated in the simulation tool

procedure of the best solution search is also available and explained in steps in the simulation tool so one can follow entire decision process.

The developed algorithm proved to be very useful in effective subcontracting task in the manufacturing network. In the test shown above it found configuration of the VO providing the order realization optimal regards to cost. The next phase of the verification concerned efficiency of the developed solution. In order to indicate benefits resulting from applying the simulation tool dedicated to order subcontracting there was made a measure of time of the VO configuration building in two circumstances:

1. The randomly chosen enterprise from the database is assigned to the work realization.
2. The system with the implemented algorithm searches for the solution.

Figure 5 shows results of analysis conducted for six different order sizes: 10, 20, 50, 100, 500, 1000 pieces. The ordered object was again hydraulic cylinder.

Applying the simulation system with implemented algorithm provides significant benefits regarding time of an order realization. Only in case of order with volume 10 the time of realization in configuration selected by the algorithm is 73 % shorter than in Virtual Organization build on base of random choices. In the event of 1000 volume order the time saving reaches 78 %. Saving of time and quick response to the customer is extremely important for entrepreneurs operating in dynamic, competitive environment. There is great probability that a customer would prefer to have his order done during three days as it have place in volume 10 order from the given example realized in configuration suggested by the simulation tool instead of ten days as it happens when the contractors are picked randomly. Thanks to the fast order realization the competitive advantage of enterprises belonging to cluster/network increases. Moreover the cost of lost opportunities are decreased because the enterprise may realize more contracts at the same time and possibly generate more profits.

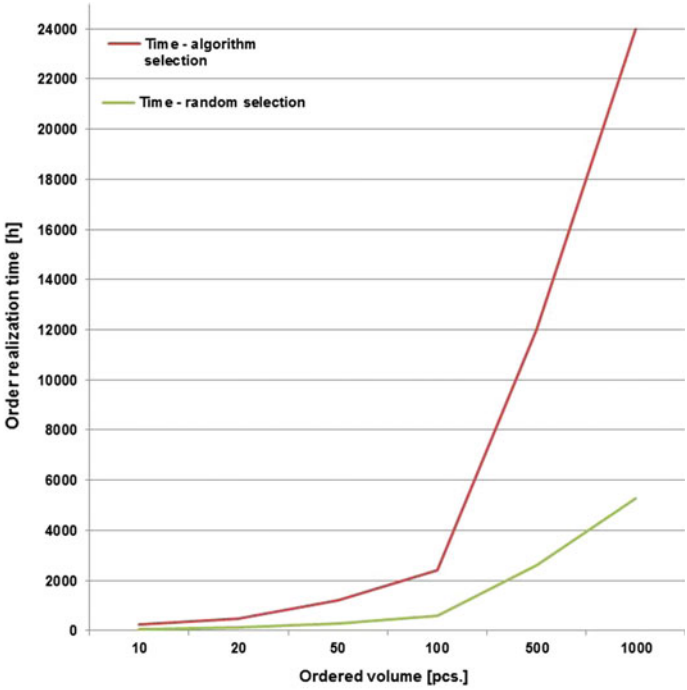


Fig. 5 Comparison of order realization time with and without the algorithm for VO configuration

The conducted verification proved the proposed tool efficiency and its effectiveness in indicating optimal choice of subcontractors for the given order with considering of criteria set.

5 Conclusions

A new business structures adopting is nowadays significant success factor especially in operation of small and medium enterprises. The dynamic environment which requires flexibility and fast reaction to the changes enforced trends connected with tight cooperation of business units in order to broaden their competence range and innovative potential. The new circumstances imply need for a dedicated aiding tools. The key aspect of effective operation in dispersed manufacturing environment is choosing proper subcontractors who will best meet the customer demand. The complex criteria set and spacious database has to be analyzed with decision support system. The simulation tool presented in the article is a prototype of such tool with implemented algorithm which during test proved to be very effective and efficient in field of configuring optimal virtual organization for a given order realization. It is

user friendly thanks to the readable way of solution displaying. It is also very flexible because enable to insert map of any region and analyze data from different networks/clusters. The user has full control over the selection process because may trace the solution search process and modify the solution if the sensitivity analysis shows better options the one suggested by the system. Applying the proposed tool enables to check how the order realization process proceeds in virtual environment before actual activating it in real production systems.

References

1. Dekkers, R., Bennett, D.: Industrial networks of the future: review of research and practice. In: Dekkers R. (ed.) Dispersed Manufacturing Networks. Challenges for Research and Practice pod red, pp. 13–34. Springer, London (2009)
2. Villa, A., Antonelli, D.: A view of SME clusters and networks in Europe. In: A Road Map to the Development of European SME Networks, pp. 23–60. Springer, London (2009)
3. Kuroiwa, I., Heng, T.M.: Production Networks and Industrial Clusters, Integrating Economies in Southeast Asia. Seng Lee Press, Singapore (2009)
4. Feng, D.Z., Yamashiro M.: A pragmatic approach for optimal selection of plant-specific process plans in a virtual enterprise. J. Mater. Process. Technol. **173**, 194–200 (2006)
5. Ko, C.S., Kim, T., Hwang, H.: External partner selection using tabu search heuristics indistributed manufacturing. Int. J. Prod. Res. **39**(17), 3959–3974 (2001)
6. Zeng, Zhi-Bin, Li, Yan, Zhu, Wenxing: Partner selection with a due date constraint in virtual enterprises. Appl. Math. Comput. **175**, 1353–1365 (2006)
7. Wu, N., Su, P.: Selection of partners in virtual enterprise paradigm. Robot.Comput. Integr. Manuf. **21**, 119–131 (2005)
8. Ipa, W.H., Yunga, K.L., Wang, D.: A branch and bound algorithm for sub-contractor selection in agile manufacturing environment. Prod. Econ. **87**, 195–205 (2004)
9. Jarimo, T., Salo, A.: Optimal Partner Selection in Virtual Organisations with Capacity Risk and Network Interdependencies, 2007. Materiały z projektu ECOLED (2007)
10. Talluri, S., Baker, R.C.: Quantitative framework for designing efficient business process allience, In: Proceedings of 1996 International Conference on Engineering and Technology Management, Piscataway, pp. 656–661 (1996)
11. Ip, W.H., Huang, M., Yung, K.L., Wang, D.: Genetic algorithm solution for a risk-based partner selection problem in a virtual enterprise. Comput. Oper. Res. **30**, 213–231 (2003)
12. Wang, D., Yung, K.L., Ip W.H.: A heuristic genetic algorithm for subcontractor selection in a global manufacturing environment. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **31**(2). 189–198 (2001)
13. Dinu, S., Pacuraru, R.: Intelligent modeling method based on genetic algorithm for partner selection in virtual organizations. Bus. Econ. Horiz. **5**(2), 23–34 (2011)
14. Jarimo, T., Salo, A.: Optimal Partner Selection in Virtual Organisations with Capacity Risk and Network Interdependencies, 2007. Materiały z projektu ECOLED (2007)
15. Gąbka, J., Susz, S., Chlebus, E.: A concept of advisory system supporting partner selection in virtual organizations. In: Innovations in Production Engineering, Part V: Knowledge Management and Decision Support Systems, pp. 349–357 (2012)
16. Chlebus, E., Chrobot, J., Gąbka, J., Susz, S.: Clusters as a modern pattern of running business supporting innovation. Manag.Prod. Eng. Rev. **2**(2), 71–79 (2011)

Distribution of Inversions and the Power of the τ- Kendall's Test

Mariusz Czekala and Agnieszka Bukietyńska

Abstract In this paper we firstly present the so far known result of the distribution of the number of inversions in the sequence of random variables. We say for the sequence $(X_1, X_2, \ldots, X_n)$ the inversion is given for i, j and X_i, X_j when $i < j$ and $X_i > X_j$. Under independence we show the exact distribution of the number of inversions in the permutation (equivalent to τ- Kendall distribution). The difference is in normalizing constants. Considering inversions is more convenient. The aim of this paper is to provide the exact distribution respectively for the dependence case.

Keywords T- kendall · Dependence · Permutation · Inversion · Power of the test

1 Introduction

In the classical Kendall's test, the number of disorders is a basis to build well known - τ -Kendall's statistics. This statistics is a popular measure of dependence. Indeed Kendall's correlation provides a distribution free test of independence. Unfortunately it is difficult to interpret the case of rejecting null hypothesis. There exist many arti-cles and ideas which have been trying to find the probability of the second type error. The authors usually use Monte- Carlo method and some other simulation ideas. This case was researched by Ferguson, Genes and Hallin [4, 5]. They present a Monte Carlo study comparing the τ- Kendall test for independence against serial dependence. In the paper we show an exact distribution of the number of inversions. This result ensures finding the power of the test based on τ- Kendall's statistics under some alternative hypotheses.

M. Czekala (✉) · A. Bukietyńska
Wroclaw School of Banking Ul, Fabryczna 29-31, 53-609 Wroclaw, Poland
e-mail: mariuszczekala@windowslive.com

A. Bukietyńska
e-mail: agnieszka.bukietynska@wsb.wroclaw.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_14

2 A Formula for Number of Inversions

In the paper we consider two samples $(Z_1, Z_2, \ldots, Z_n)$ and $(X_1, X_2, \ldots, X_n)$ each of size n. Now consider ordering the pairs by the $(Z_1, Z_2, \ldots, Z_n)$ values. Then after ordering we obtain $\left(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)}\right)$ for some permutation σ. We can now consider the number of inversions for this permutation.

Consider specifying a sequence of inversion by recursion. Let $N_n = \frac{n(n-1)}{2}$ be the maximal number of inversions in permutation σ.

Let $\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\}$ be the number of permutation having exactly k inversions. Put $N_1 = 1$ and $\left\{ \begin{matrix} N_1 \\ 0 \end{matrix} \right\} = 0$ by definition. Then for $N_2 = 2$, we have $\left\{ \begin{matrix} N_2 \\ 0 \end{matrix} \right\} = 1$ and $\left\{ \begin{matrix} N_2 \\ 1 \end{matrix} \right\} = 1$. For $N_3 = 3$ we have $\left\{ \begin{matrix} N_3 \\ 0 \end{matrix} \right\} = \left\{ \begin{matrix} 3 \\ 0 \end{matrix} \right\} = 1, \left\{ \begin{matrix} 3 \\ 1 \end{matrix} \right\} = 2,$ $\left\{ \begin{matrix} 3 \\ 2 \end{matrix} \right\} = 2, \left\{ \begin{matrix} 3 \\ 3 \end{matrix} \right\} = 1.$

Analogously, we can obtain for the following relations for $N_4 = 6$:

$$\left\{ \begin{matrix} N_4 \\ 0 \end{matrix} \right\} = 1, \left\{ \begin{matrix} N_4 \\ 1 \end{matrix} \right\} = 3, \left\{ \begin{matrix} N_4 \\ 2 \end{matrix} \right\} = 5, \left\{ \begin{matrix} N_4 \\ 3 \end{matrix} \right\} = 6,$$
$$\left\{ \begin{matrix} N_4 \\ 4 \end{matrix} \right\} = 5, \left\{ \begin{matrix} N_4 \\ 5 \end{matrix} \right\} = 3, \left\{ \begin{matrix} N_4 \\ 6 \end{matrix} \right\} = 1$$

In the general case we get the formula:

$$\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} = \sum\nolimits_{i=\max(0,k-n+1)}^{k} \left\{ \begin{matrix} N_{n-1} \\ i \end{matrix} \right\} \tag{1}$$

On the right hand side of this formula we always have k + 1 components. Therefore, the number $\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\}$ can be calculated using the previous numbers. Note that $\sum_{k=0}^{n} \left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} = n!$ and $\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} = \left\{ \begin{matrix} N_n \\ n-k \end{matrix} \right\}$. So, we have decomposition $n!$ for N_n components. As a result, the probability of the number of inversions has the form:

$$p_{n,k} = \frac{\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\}}{\mathrm{n!}} \tag{2}$$

The following Figs. 1 and 2 show the number of inversions and probabilities for n = 2, 3, 4, and 5.

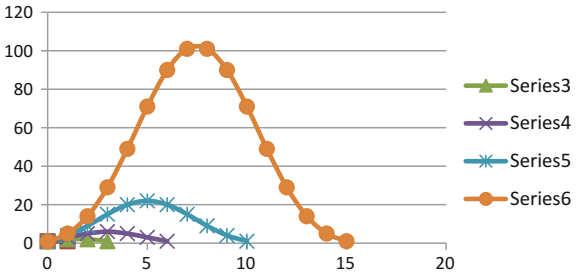


Fig. 1 The number of permutations with k inversions for n = 3, 4, 5 and 6

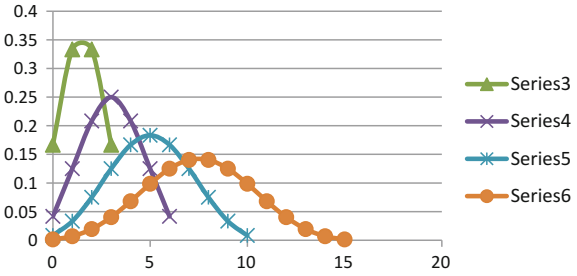


Fig. 2 Probabilities of permutations with k inversions for n = 3, 4, 5 and 6

The Considered sequence is well known from the On-Line Encyclopedia of Integer Sequences as the sequence A008302 [8].

In Table 1 the idea of the recursive formula (1) is explained. For $n = 4$ we have $6 = 1 + 2 + 2 + 1$. Similarly $101 = 9 + 15 + 20 + 22 + 20 + 15$, or $101 = 15 + 20 + 22 + 20 + 15 + 9$ (Table 2).

This distribution can be provided in a different way using random variables $\{Y_k\}$, Feller [3]. For $k = 1$ we put $Y_1 = 0$ almost surely. For $k > 1$, we define Y_k in the form:

Table 1 Sequences A008302 for chosen n and a number of inversions

	0	1	2	3	4	5	6	7	8	9
1	1									
2	1	1								
3	1	2	2	1						
4	1	3	5	6	5	3	1			
5	1	4	9	15	20	22	20	15	9	4
6	1	5	14	29	49	71	90	101	101	90
7	1	6	20	49	98	169	259	359	455	531
8	1	7	27	76	174	343	602	961	1415	1940
9	1	8	35	111	285	628	1230	2191	3606	5545
10	1	9	44	155	440	1068	2298	4489	8095	13640
11	1	10	54	209	649	1717	4015	8504	16599	30239

Table 2 Distribution of Y_k

Y_k	0	1	2	3	….	….	k-1
Probability	$\frac{1}{k}$	$\frac{1}{k}$	$\frac{1}{k}$	$\frac{1}{k}$	…..	…..	$\frac{1}{k}$

The values of Y_k are the additional numbers of inversions in permutation $(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(k)})$ having the number of inversions in permutation $(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(k-1)})$. Therefore, the number of inversions can be written in the form: $I_n = \sum_{k=1}^{n} Y_k$.

It was proved in Feller [3] that:

$$I_n \sim N\left(\frac{n^2}{4}, \frac{n^3}{36}\right)$$

This case was previously considered by David et al. [2]. The sequence $\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\}$ was firstly presented probably by Netto [7]. The recursive formula was given by Bukietyńska [1].

3 Associations

Let us assume p is a constant (by assumption) probability of inversion. So, in the presented model we assume constant probability of inversion and the distributions of random variables $(X_1, X_2, \ldots, X_n)$ are continuous. Therefore $P(X_i = X_j) = 0$ for $i \neq j$. So we consider the sequence $(X_1, X_2, \ldots, X_n)$ is without any ties.

We show the analysis of the sequence of random variables Y_k and their distributions (Table 3).

Let, as usual, Y_k be the additional number of inversions. Firstly, assume $k = 2$. Then $p_0 = p$ and $p_1 = 1 - p = q(say)$. By definition the probability of inversion equals p (Tables 4 and 5).

Table 3 Distribution of Y_k (the case of associations)

Y_k	0	1	2	3	….	….	k-1
Probability	p_0	p_1	p_2	p_3	…..	…..	p_{k-1}

Table 4 Distribution of Y_3 (the case of associations)

Y_3	0	1	2
Probability	$\frac{q^2}{w_3}$	$\frac{qp}{w_3}$	$\frac{p^2}{w_3}$

Table 5 Distribution of Y_4 (the case of associations)

Y_4	0	1	2	3
Probability	$\frac{q^3}{w_4}$	$\frac{q^2p}{w_4}$	$\frac{qp^2}{w_4}$	$\frac{p^3}{w_4}$

Table 6 Distribution of Y_k (the case of associations)

Y_k	0	1	2	3	….	k-1
Probability	$\frac{q^{k-1}}{w_k}$	$\frac{q^{k-2}p}{w_k}$	$\frac{q^{k-3}p^2}{w_k}$	$\frac{q^{k-4}p^3}{w_k}$	…..	$\frac{p^{k-1}}{w_k}$

The case of $k = 3$ is more difficult because probabilities for Y_k need to be defined in a different way. Assuming the probability of i inversions is proportional to $p^i q^{3-i}$ with constant $1/w_3(say)$ we can get the distribution of Y_k.

where $w_3 = \frac{q^3-p^3}{q-p}$.

In case of $k = 4$ we get:

where $w_4 = \frac{q^4-p^4}{q-p}$.

In the general case we have:

where $w_k = \frac{q^k-p^k}{q-p}$ (Table 6).

Now we can consider random variables $I_n = \sum_{k=1}^{n} Y_k$. Let $P_k(s)$ be the ordinary generating function [3]. We have:

$$P_1(s) = 1, \qquad P_2(s) = q + ps,$$

and

$$P_3(s) = \frac{q^2 + pqs + p^2s^2}{w_3}.$$

Therefore:

$$P_{I_3}(s) = \prod_{i=1}^{3} P_i(s) = \frac{q^3 + 2pq^2s + 2p^2qs^2 + p^3s^3}{w_3}.$$

Thus, the distribution of I_3 has the form presented in Table 7:

Note that the coefficients in the above probabilities are (1, 2, 2, 1), according to the number of inversions in case of $N_3 = 3$.

Table 7 Distribution of I_3 (the case of associations)

I_3	0	1	2	3
Probability	$\frac{q^3}{w_3}$	$\frac{2q^2p}{w_3}$	$\frac{2qp^2}{w_3}$	$\frac{p^3}{w_3}$

Table 8 Distribution of I_4 (case of associations)

I_4	0	1	2	3	4	5	6
Probability	$\frac{q^6}{w}$	$\frac{3q^5p}{w}$	$\frac{5q^4p^2}{w}$	$\frac{6q^3p^3}{w}$	$\frac{5q^2p^4}{w}$	$\frac{3qp^5}{w}$	$\frac{p^6}{w}$

In case of $k = 4$ we have:

$$P_{I_4}(s) = P_{I_3}(s)P_4(s) = \frac{(q^3 + 2pq^2s + 2p^2qs^2 + p^3s^3)(q^3 + q^2p + qp^2 + p^3)}{w_3w_4}$$

Therefore:

$$P_{I_4}(s) = \frac{\left(q^6 + 3q^5ps + 5q^4p^2s^2 + 6q^3p^3s^3 + 5q^2p^4s^4 + 3qp^5s^5 + p^6s^6\right)}{w_3w_4}$$

Thus, the distribution of I_4 has the form:

Where $w = w_3w_4$. The coefficients (1, 3, 5, 6, 5, 3, 1) are the number of inversions (according to the case of $N_4 = 6$). The main theorem of this paper provides the probability of the number of inversions in the general case (Table 8).

Theorem *Under the assumptions above*

$$P(I_n = k) = \frac{\left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} p^k q^{N_n - k}}{\sum_{k=0}^{N_n} \left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} p^k q^{N_n - k}} = \frac{p_{n,k} \cdot p^k q^{N_n - k}}{\sum_{k=0}^{N_n} p_{n,k} \cdot p^k q^{N_n - k}}$$

Proof The ordinary generating functions method will be applied in this proof. It is enough to show the equality:

$$P_{I_{n+1}}(s) = P_{I_n}(s)P_{n+1}(s) \tag{3}$$

where $P_{I_{n+1}}(s)$ (or $P_{I_n}(s)$) is the generating function for the number of inversions for n or $n+1$ (respectively). $P_{n+1}(s)$ is the generating function of the additional number of inversions. In the considered Eq. (3) note that the denominators on the both sides are:

$$w = \prod_{i=1}^{n+1} w_i,$$

hence considering the numerators is sufficient. We have:

$$P_{I_n}(s)P_{n+1}(s) = \left(\frac{1}{w}\right)\left(\sum\nolimits_{k=0}^{N_n} \left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} p^k q^{N_n-k} s^k\right)\left(q^n + q^{n-1}ps + q^{n-2}p^2s^2 + \ldots + p^n s^n\right) \tag{4}$$

and:

$$P_{I_{n+1}}(s) = \left(\frac{1}{w}\right)\left(\sum\nolimits_{k=0}^{N_{n+1}} \left\{ \begin{matrix} N_n+1 \\ k \end{matrix} \right\} p^k q^{N_{n+1}-k} s^k\right).$$

Let us consider the coefficient before s^i. Firstly for $i = 0$, after multiplication we have $\left\{ \begin{matrix} N_n \\ 0 \end{matrix} \right\} q^{N_n} q^n = \left\{ \begin{matrix} N_n+1 \\ 0 \end{matrix} \right\} q^{N_{n+1}}$, because $\left\{ \begin{matrix} N_n \\ 0 \end{matrix} \right\} = \left\{ \begin{matrix} N_{n+1} \\ 0 \end{matrix} \right\}$, and $N_{n+1} = N_n + n$. So, this coefficient is equal to the first coefficient of the numerator of $P_{I_{n+1}}(s)$.

For $i = 1$ we obtain:

$$\begin{aligned} \left\{ \begin{matrix} N_n \\ 0 \end{matrix} \right\} q^{N_n} q^{n-1} ps + \left\{ \begin{matrix} N_n \\ 1 \end{matrix} \right\} pq^{N_n-1} q^n s &= \left(\left\{ \begin{matrix} N_n \\ 0 \end{matrix} \right\} + \left\{ \begin{matrix} N_n \\ 1 \end{matrix} \right\}\right) q^{N_n+n-1} ps \\ &= \left\{ \begin{matrix} N_{n+1} \\ 1 \end{matrix} \right\} q^{N_{n+1}-1} ps, \end{aligned}$$

Hence the coefficient on the right hand side of the Eq. (4) is consistent with the corresponding coefficient on the left hand side.

To establish this part of theorem for the general case (coefficient before s^k) we consider the following expression as the result of multiplication:

$$\begin{aligned} &\left\{ \begin{matrix} N_n \\ 0 \end{matrix} \right\} q^{N_n} p^k s^k + \left\{ \begin{matrix} N_n \\ 1 \end{matrix} \right\} q^{N_n-1} psqp^{k-1} s^{k-1} + \\ &+ \left\{ \begin{matrix} N_n \\ 2 \end{matrix} \right\} q^{N_n-2} p^2 s^2 q^2 p^{k-2} s^{k-2} + \ldots + \left\{ \begin{matrix} N_n \\ k \end{matrix} \right\} q^{N_n-k} p^k s^k q^k \\ &= \left(\left\{ \begin{matrix} N_n \\ 0 \end{matrix} \right\} + \left\{ \begin{matrix} N_n \\ 1 \end{matrix} \right\} + \left\{ \begin{matrix} N_n \\ 2 \end{matrix} \right\} + \ldots + \left\{ \begin{matrix} N_n \\ k \end{matrix} \right\}\right) = \left\{ \begin{matrix} N_{n+1} \\ k \end{matrix} \right\} q^{N_n-k} p^k s^k \end{aligned}$$

Therefore, we have the corresponding coefficient on the left hand side of Eq. (3). The proof by induction is completed.

4 The Power of the Test

As it was mentioned above, there were many attempts to find the power of the rank-based test against some alternatives. The most popular alternative was serial correlation. Hallin and Melard [6] considered serial dependence testing hypotheses about randomness and independence (lack of correlation) in the series of random variables. Some results concerning this problem were presented by Ferguson et al. [4, 5].

They presented the results of Monte Carlo simulation for the AR(1)

$$X_i - \theta X_{i-1} = \varepsilon_i \tag{5}$$

model under some additional assumption concerning the distribution of random disturbances. The frequency of rejecting hypothesis (significance level = 0.05) of randomness against serial correlation (sample $n = 20$) varies for $\theta = 0.25$, from 23.5 % to 27.7 % under assumption about normal, logistic or Laplace distribution of random disturbances. Only for the case of Cauchy distribution this probability (frequency) equals 51,9 %.

For one-sided test at the 5 % level applied to the series length $n = 50$ the percentage of rejection varies from the first three cases from 48.5 % to 58.3 %. Only under assumption about distribution of Cauchy the analyzed frequency equals 88.8 %.

These results do not give a reliable tool to test the considered hypothesis (in spite of the opinion of the above mentioned authors). The problem is in the formulation of an alternative (alternatives). Serial correlation is one of the ways, but the expected power could not t be high because of the nature of the autoregressive process. Let us consider the probability of inversion at point i (disorder at i and $i - 1$). Only to simplify an example let us assume $0 < \theta < 1$.

After simple algebra we get:

$$p = P(X_i < X_{i-1}) = P(X_i - X_{i-1} < 0) = P(\theta X_{i-1} + \varepsilon_i - X_{i-1} < 0)$$

thus:

$$p = P(X_{i-1} < \frac{\varepsilon_i}{1-\theta}).$$

Hence, this probability does not depend on X_i. On the average this probability equals 0. Yet, the results mentioned above cannot be totally disqualified? It is because of transitivity of dependence. Considering random variables X_i and X_{i-1} in Eq. (5), we should note the correlation of these random variables with time (for positive value of θ). Therefore (transitivity) we can observe artificial correlation of random variables X_i and X_{i-1}.

In this paper the tested hypothesis is presented in Table 1, but the alternative one in Table 5. Unfortunately the alternative hypothesis is not any important kind of

Table 9 Critical values at the 5 % significance level

n	N_n	cv($\alpha = 0.05$)	P($I_n \geq$ cv)
5	10	8	0.049
6	15	12	0.035
7	21	16	0.046
8	28	21	0.043
9	36	26	0.038
10	45	32	0.036
11	55	40	0.043

Table 10 The probabilities $\beta = P(\{I_n < cv|p\})$ at the 5 % significance level

n	N_n	p = 0.55	p = 0.6	p = 0.7	p = 0.8
5	10	0.785	0.653	0.350	0.117
6	15	0.835	0.682	0.315	0.075
7	21	0.801	0.589	0.174	0.018
8	28	0.796	0.535	0.101	0.005
9	36	0.736	0.405	0.034	0.0004
10	45	0.632	0.319	0.012	0.00004
11	55	0.745	0.323	0.007	0.00001

serial correlation. It could be such an example after choosing some special values of parameters.

Instead of that, we have a clear mechanism of generating numbers, associated with the numbers of inversions according to Tables 1 and 5.

In Table 9 we show the critical values for one-sided test of randomness versus dependence at the 5 % significance level (series of length from 5 to 11).

The probabilities of committing a type two error - β are presented in Table 10, for chosen values of p in the alternative hypothesis.

In Table 11 we compute two most important parameters from the distributions presented in Tables 1 and 5. We can compute the expected value and variance in each particular case. In Table 11 we have these values for the case of n = 30 (Table 12 and Fig. 3).

Apart from this tables it is easy to observe (analyzing theorem 1) because of symmetry that:

$$E(I_n|p) = N_n - E(I_n|1-p)$$

Table 11 Expected values and variances for the case of $n = 30$ and chosen probabilities of inversion

p	0.5	0.51	0.52	0.53	0,54	0.55	0.56	0.57	0.58	0.59
$E(I_{30})$	217.5	249.5	278.0	302.5	322.7	339.0	352.2	362.8	371.5	378.8
$V(I_{30})$	785.4	751.7	665.0	556.4	451.3	361.9	290.6	235.3	192.6	159.6

Table 12 Expected values and variances for the case of $n = 30$ and chosen probabilities of inversion ($p \geq 0.6$)

p	0.6	0.7	0.8	0.9
$E(I_{30})$	383.8	414.2	425.5	431.4
$V(I_{30})$	133.7	34.7	12.4	4.0

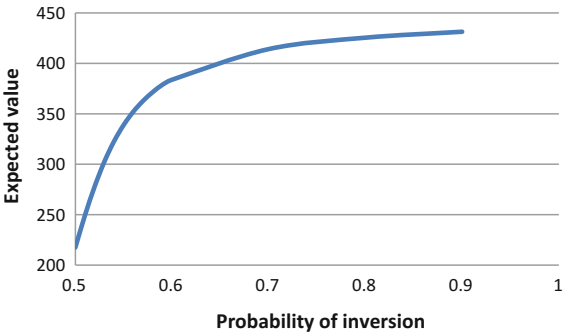


Fig. 3 Probabilities of inversion and expected value

and:

$$V(I_n|p) = V(I_n|1-p)$$

We can notice two tendencies. Growth in p expected values and decrease in p variances. These tendencies are very clear. Therefore, we can use these values to calculate the power of the test assuming (approximately) normality and taking distinguished values of parameters (Fig. 4).

The parameters in the asymptotic case seem to be dependent on the sum $\sum_n \frac{z^n}{1-z^n}$, where $z = p/q$ or $z = q/p$. This series is well known in the theory of analytical

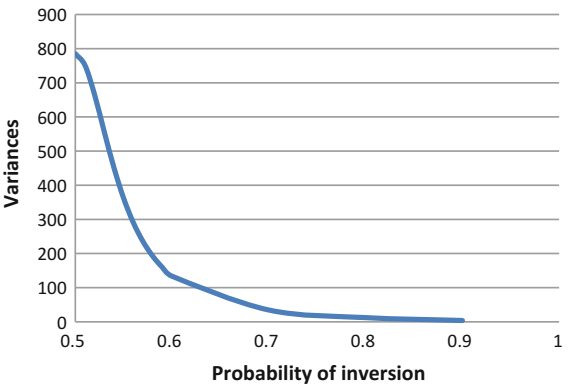


Fig. 4 Probabilities of inversion and variances

functions. One should consider the ordinary generating functions and their first and second derivative. The considered sums depends on the number of divisors of n. So we have the classical problem of the number theory.

References

1. Bukietyńska A: Kombinatoryczny test inwersji, Metody Ilosciowe w ekonomii, WSB Poznan, pp. 152162, (2008)
2. David, F.N., Kendall, M.G., Barton, D.E.: Symmetric Function and Allied Tables, p. 241. Cambridge (1966)
3. Feller, W.: An Introduction to Probability Theory and its Application. Wiley, New York, London (1961)
4. Ferguson, S., Genest, Ch., Hallin, M.: Kendall's tau for autocorrelation. Can. J. Stat. **28**, 587–604 (2000)
5. Ferguson, S., Genest, Ch., Hallin, M.: Kendall's tau for autocorrelation. Dep. Stat. Pap., UCLA (2011)
6. Hallin, M., Metard, G.: Rank–based test for randomness against first order dependence. J. Am. Stat. Assoc. **83**, 1117–1128 (1988)
7. Janjic, M.: A generating function for numbers of insets. J. Integer Seq. **17** #14.9.7 (2014)
8. Kendall, M.G., Buckland, W.R.: A Dictionary of Statistical Terms. OLIVER AND BOYD, Edinburgh, London (1960)
9. Netto E.: Lehrbuch der Combinatorik. 2nd ed., Teubner, Leipzig, p. 96. (1927)
10. The On-Line Encyclopedia of Integer Sequences, sequence A008302

Part V
Network and Transport Systems

Reduction of Congestion in Transport Networks with a Fractal Structure

Grzegorz Bocewicz, Zbigniew Banaszak and Izabela Nielsen

Abstract Transport Systems (TS) and the processes of movement of goods from the point of origin to destination which take place in those systems determine the competitiveness of companies and businesses using them. Transport networks which make up a TS encompass various modes of transport, e.g., road vehicles, trains, freight cars, containers, material packages, etc. Together, these modes form streams of traffic in the system. Assuming that the structure of a TS network determines its behavior, in this study, we attempt to develop a declarative model which would enable analysis of the relationships between the structure of a TS and its potential behavior. The problem in question boils down to determining sufficient conditions ensuring smooth traffic flow in a transport network with a fractal structure. The proposed approach, which assumes a recursive, fractal network structure, enables rapid prototyping, in polynomial time, of alternative transport routes and associated schedules. An example is used to illustrate the quantitative and qualitative relationships between the morphological characteristics of the investigated TS structures and the functional parameters of the transport processes carried out in them.

Keywords Congestion · Transport network · Fractal structure · Declarative modeling · Multimodal process · Constraint satisfaction problem

G. Bocewicz (✉)
Faculty of Electronics and Computer Science, Department of Computer Science and Management, Koszalin University of Technology, Koszalin, Poland
e-mail: bocewicz@ie.tu.koszalin.pl

Z. Banaszak
Faculty of Management, Department of Business Informatics, Warsaw University of Technology, Warsaw, Poland
e-mail: Z.Banaszak@wz.pw.edu.pl

I. Nielsen
Department of Mechanical and Manufacturing Engineering, Aalborg University, Aalborg East, Aalborg, Denmark
e-mail: izabela@m-tech.aau.dk

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_15

1 Introduction

Transport Systems (TS) and the associated movement of cargo, people, goods, energy, financial capital, data, etc. from a point of origin to a destination determine the competitiveness of companies and institutions which use those systems. A commonly accepted definition of a system describes it as being characterized, regardless of its specific nature and character, by a structure (the component subsystems and relations between them) and a behavior, which determines the responses of the system to changes in (expectations of) the environment. This means that the categories of structure and behavior can also be distinguished in the class of TS considered in the present paper. They encompass the transport network and the modes of transport moving in that network (road vehicles, freight cars, containers, packages, money transfers, etc.) which form traffic streams, and transport processes involved in the movement of objects (goods and/or passengers) from their points of origin to their destinations. The goal of a TS understood in this way is to move people and/or cargo [1, 3, 6].

Under this approach, the following TS analysis and synthesis problems are usually considered: Does an arbitrarily given transport network structure of a TS make it possible to carry out transport processes that meet user expectations? Is there a TS network structure which guarantees execution of transport processes that match given expectations of its users? The distinction made between the above two classes of problems assumes that just as any structure of a system determines its admissible behavior, so too the behavior of a system can be determined by its different structures. The elements that condition the solution to these problems are the relationships between selected structural and functional parameters of the system. This means that declarative models of analysis and synthesis problems should incorporate decision variables specifying the topology of transport networks, the fleet of vehicles that use it, and the stations and loading/unloading stops across the network as well as the transport routes of the objects being moved and transport route schedules. The constraints which connect decision variables found in this type of models allow formulation of suitably dedicated constraint satisfaction problems [2] which are easily implemented in constraint programming languages such as OzMozart, ILOG, ECLiPSe, [2, 8, 10, 12].

In the above context, the class of TS analyzed here is limited to network structures with a regular, recursive morphology typical of tree or mesh (grid) topologies (Fig. 1). This category of topologies, which include urban transport systems, are the subject of intensive research [1, 6, 7, 11].

In practice, however, most of this research is limited to either identifying the fractal pattern of the analyzed transport network, or to estimating qualitative and quantitative parameters of TS operation depended on the manner fractal patterns are propagated in order to satisfy the needs of urbanized communication infrastructures. For the purposes of further discussion, it is assumed that a TS encompasses all possible branches of transport and transport technologies, including road and rail (surface and underground) transport, e.g. buses, streetcars, subway lines.

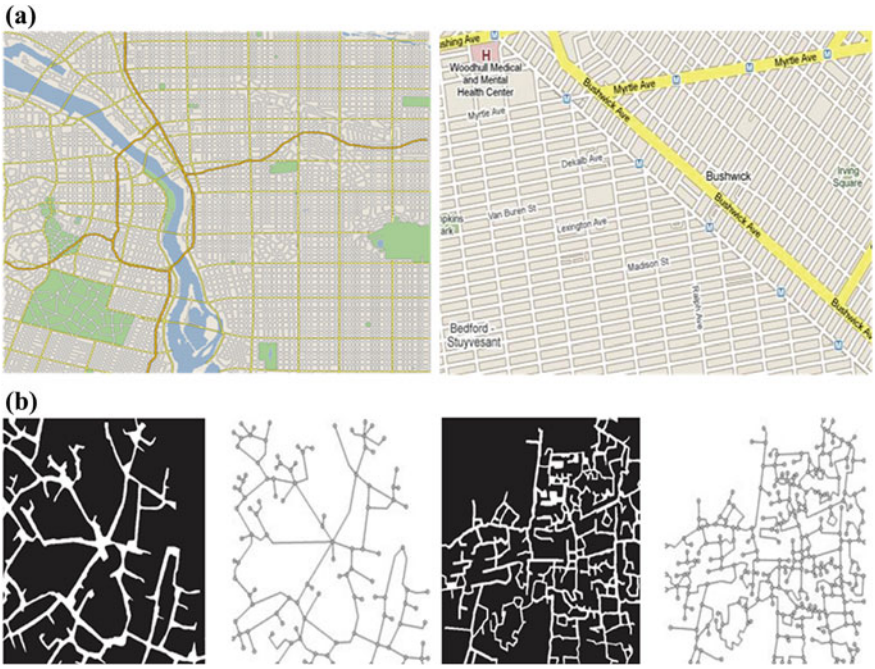


Fig. 1 Communication structures with a mesh topology (*source* [16]) (**a**), and a tree topology (**b**)—the *white spaces* on the *black* background are communication route maps and the black edges and nodes on the *white* background are graph representations of route maps (*source* [6])

This means that the transport processes associated with the movement of objects are multimodal. It is worth noting that the various modes of transport (buses, streetcars, commuter rail, subway lines) which form traffic flows in a TS run to scheduled timetables, moving along a fixed, closed-loop route. This observation implies that simultaneous access of different means of transport to interchange (transshipment) stations on the same transport route is limited. This limitation entails that access of alternative means of transport to shared stations and/or stops must be regulated by dispatching rules (e.g. fixed timetables) which constitute suitably dedicated implementations of a mutual-exclusion protocol.

Referring to the terminology according to which multimodal transport is the transportation of freight performed with different, alternative modes of transport along the same transport route, during which goods can be transshipped between different transport modes [14], authors of [2, 4] have introduced the concept of a multimodal transport process (MTP). According to this definition, an MTP involves the movement of objects using different modes of transport in a single, integrated transport chain on a given route. Examples of MTPs include the processes associated with daily commuting (bus—streetcar—subway), courier services (e.g., DHL), etc. A characteristic feature of such multi-modal processes is that their transport routes are made up of local segments operated by one mode of transport or

involving one type of transport processes, and the objects are moved by suitable local means of transport. A good illustration of this feature is a passenger travelling by subway who, in the course of his journey, changes from one line to another in accordance with an itinerary.

The present study adopts these assumptions to investigate the problem of TS analysis, more specifically the reachability problem (looking for specific behaviors in a system with a given structure), and the synthesis problem (looking for a system structure which can guarantee a desired behavior of the system). The former problem boils down to the problems of routing and scheduling of MTPs carried out in transport networks with a fixed fractal structure. Formulated in this way, it assumes that network topology, routes, parameters of local means of transport, dispatching rules governing access to shared stations or stops, and initial and terminal points of the alternative routes of MTP are the givens. What is sought are the variants of routes which guarantee delivery times not exceeding a pre-set deadline.

In the TS synthesis problem, it is assumed that the topology of the transport network, the routes, and parameters of the means of local transport are given, and the unknown is the dispatching rules governing access to shared stations or stops, which guarantee timely completion of an MTP, as scheduled.

The above mentioned problems belong to the class of computationally hard problems. The adopted simplifying assumptions enable rapid prototyping of admissible solutions in polynomial time. The investigations presented in this paper and the results obtained in the course of this study are a continuation of previous studies collected in [2–5].

Section 2 introduces the problems of modeling of transport networks with a fractal structure and management of local transport processes. Section 3 presents a declarative model of a reference TS and formulates related constraint satisfaction problems. Section 4 presents the main results of the study, including those related to sufficient conditions for a smooth (cyclic) execution of local processes in transport networks, and the conditions ensuring timely execution of a MTP. The prospects and limitations of the line of research undertaken in this study as well as the scope of future work in this area are reviewed in Sect. 5.

2 A Reference Model of a Transport Network with a Fractal Structure

The structures of integrated, road and rail (streetcar lines, subway lines, commuter rail lines and inter-city rail lines) transport networks which comprise the infrastructure of urban regions, respond to the needs of those regions, at the same time determining their future development. Much of recent research [7, 13, 15] draws attention to the fact that the development of urban agglomerations, and in particular the morphology of urban regions, is subject to the laws of recursion, which are best modeled by fractal structures. The consequences of this fact can be used both in

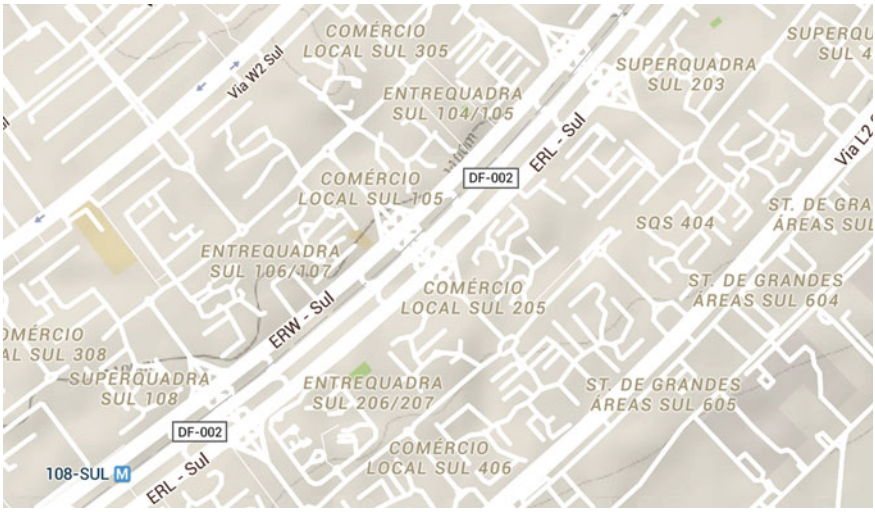


Fig. 2 Examples of routes with a tree-like structure in the city of Brasilia (*source* [17])

predicting the needs related to the expansion of the existing transport infrastructure, as well as planning new industrial and/or urban agglomerations. It is easy to notice that just as the development of technology leaves its mark on the structure of transport networks (e.g. in the past, the development of the automotive industry fostered construction of regular grid-like mesh structures such as the street layout of Manhattan, Fig. 1a, and tree structures such as the street layout of Brasilia, the capital of Brazil, Fig. 2), so too the choice of the topology of transport structures

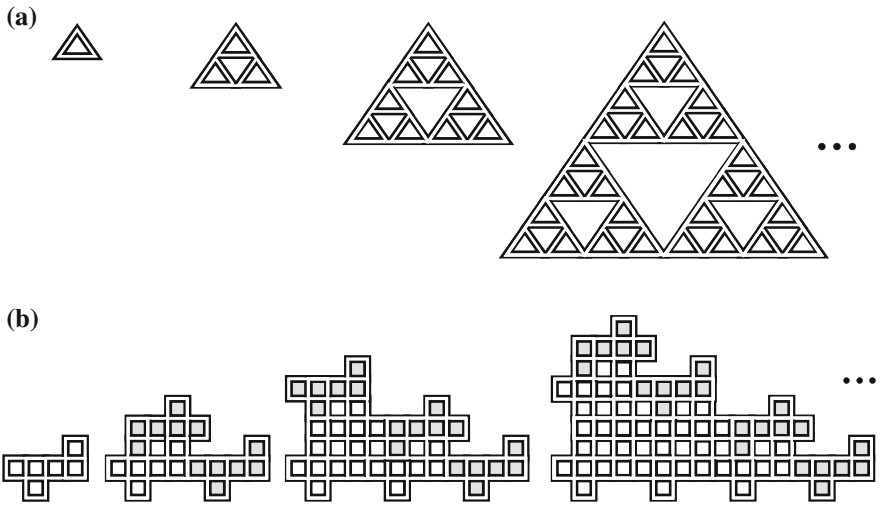


Fig. 3 Examples of routes in transport networks with a fractal structure: generated by shape ▲ (**a**), generated by shape ⌐ (**b**)

(which is dependent, among others, on the morphology of the terrain they occupy) is reflected in the organization and quality of public transport and the efficiency of the transport of goods.

In order to determine the relationships between the structure of transport networks along with the means of transport moving in these network and MTPs which determine the routes for the transport of objects, let us consider a reference TS model which integrates the models of a TS network (routes), local transportation processes and MTP.

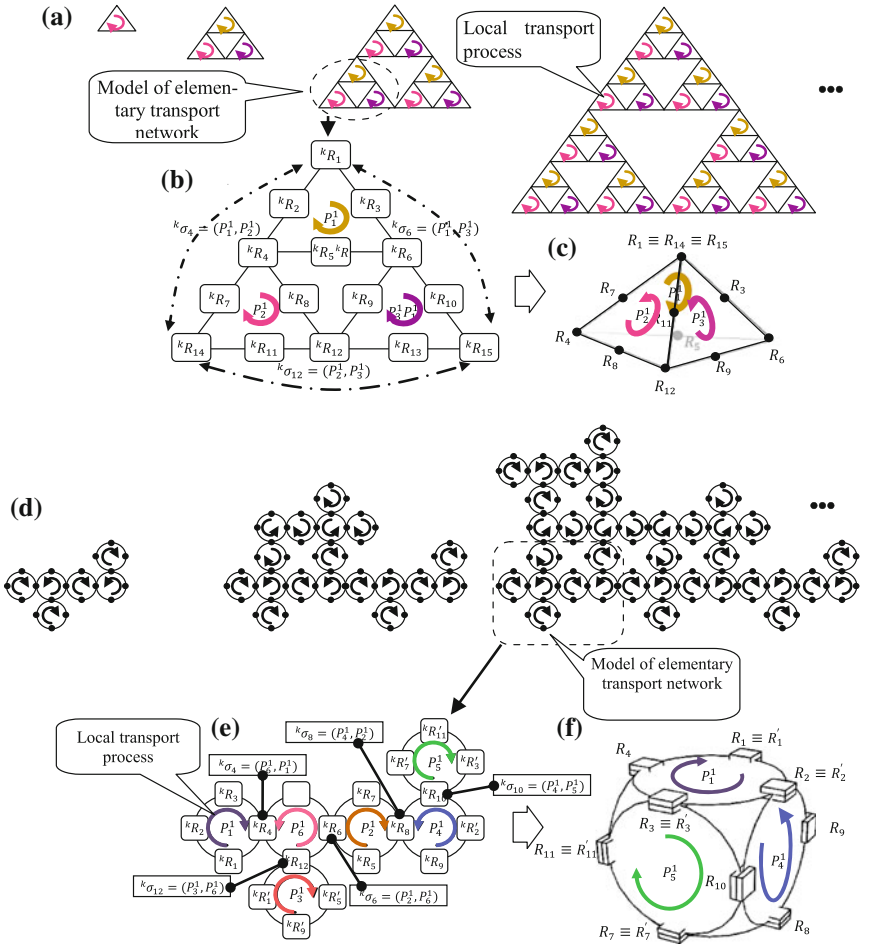


Fig. 4 Graph theoretical models of fractal structures corresponding to the route patterns in Fig. 3a, b are shown in (**a**) and (**d**), respectively; uncovered forms of elementary structures (**b**) and (**e**), respectively; covered forms of elementary structures (**c**) and (**f**), respectively

For definiteness, let us consider two types of transport networks with fractal structures as those shown in Fig. 3. Graph theoretical models of these structures with vertices (which represent network resources, i.e. stations and stops and shared route sections) and edges are shown in Fig. 4a. Network resources are denoted by ${}^{k}R_{r}$ that means the r-th resource in the k-th elementary transport network. Local transport processes are marked with labeled arcs whose orientation indicates the direction of flow of local traffic (transport modes); for example, the arc labeled P_i^j means that the considered traffic flow, in the local transport process, is comprised of "j" units of an i-th transport mode (Fig. 4b, e). Models of MTP routes, showing the sequences of the resources between which objects are moved are represented graphically with bold symbols of nodes and arcs; see Fig. 5a. Other elements that connect the component models of the reference model include a set of dispatching

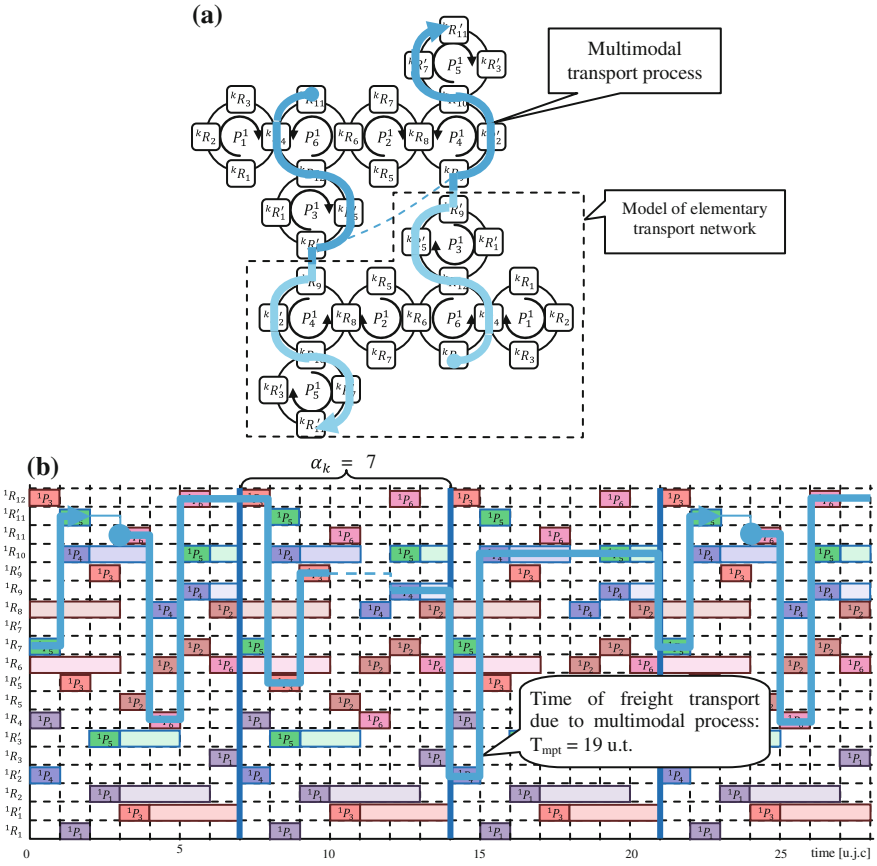


Fig. 5 A representation of a sample MTP performed in a transport network having the structure given in Fig. 3 (**b**); a graphic model of a transportation structure, incorporating an elementary structure (**a**), Gantt chart of local transport processes and the sample MTP

rules, travel/dwelling times of the individual modes of transport, and Gantt charts illustrating the dynamics of movement of vehicles and goods.

Dispatching rules for synchronization of access of local transport processes to shared network resources are marked with labels ${}^{k}\sigma_r = (A_j, B_q, \ldots, A_j, \ldots, Z_q)$ describing the order of access of means of transport $A_j, B_q, \ldots, A_j, \ldots, Z_q$ to a shared resource ${}^{k}R_r$; see Fig. 4b, c. Travel and/or dwelling times of means of transport for various network resources are designated with the symbol $t_{i,j}$ which denotes the time of execution of the j-th operation of the i-th transport process. A graphic model which can be used to represent both travel/dwelling times and the wait times of the means of transport used and the goods moved using those means is a Gantt chart (Fig. 5b).

By illustrating the behavior trajectory of a TS in the admissible-state space (i.e. by providing a graphical representation of timetables and/or supply schedules), a Gantt chart allows one to assess waiting times associated with the fact that a means of transport has to wait for access to a requested but currently occupied resource, as well as waiting times of objects transported in MTP chains resulting from unavailability of the scheduled means of transport.

3 Problem Formulation

The reference model of a transport network with a fractal structure presented above makes it possible to develop an appropriate dedicated declarative model allowing formulation of the aforementioned TS analysis and synthesis problems as constraint satisfaction problems. A constraint satisfaction problem: $PSO = ((X, D), C)$ is usually given by [12] a finite set of decision variables $X = \{x_1, x_2, \ldots, x_n\}$, a finite family of finite domains of discrete decision variables $D = \{D_i | D_i = \{d_{i,1}, d_{i,2}, \ldots, d_{i,j} \ldots, d_{i,m}\}, i = 1 \ldots .n\}$, and a finite set of constraints limiting the values of the decision variables $C = \{C_i | i = 1 \ldots .L\}$, where: C_i is a predicate $P[x_k, x_l, \ldots, x_h]$ defined on a subset of set X. What is sought is an admissible solution, i.e. a solution in which the values of all decision variables X satisfy all constraints C.

Accordingly, the declarative TS model comprises:

- sets of decision variables describing the structures of

– local transport processes, i.e. the type and number of resources and modes of transport they use, as well as the associated travel/dwelling times,
– the MTP, i.e. the type and number of resources in a chain and the type and number of transport modes used, as well as the associated travel/dwelling times,

- Domains of decision variables,
- Sets of determining constraints:

– sets of dispatching rules assigned to shared network resources
– transport schedules determining the periods (takts) and dates of delivery of transported goods.

These assumptions, on the one hand, explicitly constrain the topology of TS routes to transport networks with fractal structures and, on the other hand, implicitly make the efficiency of potential MTPs (e.g. regarding the possible delivery dates) conditional on the admissible flow of traffic (e.g. congestion-free traffic) operating under local transport processes. This observation implies that research can be limited to certain elementary structures that make up the whole transport network. Examples of such structures are marked with a dotted line in Fig. 4a, b. A further assumption is that the repetitive, cyclic behavior of an elementary structure marked out in this way implies cyclic, i.e. congestion-free [9, 17] flow of traffic (behavior) across the network. This allows searching for alternative admissible congestion-free variants of solutions to problems of MTP routing and scheduling.

Let us consider "covered" forms (Fig. 4c, h) of elementary structures (Fig. 4b, d). As it can be easily noted, "covered" forms arise as a result of "gluing together" of selected vertices of elementary structures. Just which vertices are "glued" together in the "covered" form is determined by the choice of those resources of the elementary transport structure which are shared with the resources of neighboring structures of the transport network. For example, a vertex corresponding to resource ${}^{k}R_{12}$ is glued with a vertex corresponding to resource ${}^{k}R'_{12}$, because resource ${}^{k}R'_{12}$ is shared with resource ${}^{l}R_{12}$ which is a counterpart of ${}^{k}R_{12}$, see Fig. 5a.

It can be shown that if the traffic flow in a given covered form of an elementary transport structure is free of congestion, i.e. the schedule which specifies it is a cyclic schedule, then the flow of traffic in the entire transport network consisting of uncovered forms of elementary structures also has a cyclic nature. This observation allows one to focus on formulating the following constraint satisfaction problem, the solution to which is a structure (a set of dispatching rules) that guarantees a congestion-free flow of traffic. In other words, assuming that the behavior of each i-th elementary structure is represented by a cyclic schedule ${}^{(i)}X' = \left({}^{(i)}X_k | k = 1, \ldots, h, \ldots, L_i\right)$, where: ${}^{(i)}X_h$ is a set of beginning moments of operation of the h-th local process of the i-th elementary structure and L_i denotes the cardinality of the set of local processes comprising the i-th elementary structure, the constraint satisfaction problem in question has the following form:

$$PS_i = (\left(\{{}^{(i)}X', {}^{(i)}\Theta, {}^{(i)}\alpha\}, \{D_X, D_\Theta, D_\alpha\}\right), \{C_L, C_M, C_D\}) \tag{1}$$

where: ${}^{(i)}X'$, ${}^{(i)}\Theta$, ${}^{(i)}\alpha$—decision variables,

- ${}^{(i)}X'$—cyclic schedule of the i-th elementary structure,
- ${}^{(i)}\Theta$—set of dispatching rules determining the order of operations competing for access to the common resources of the i-th elementary structure,
- ${}^{(i)}\alpha$—set of values of periods of local processes occurring in the i-th elementary structure,

D_X, D_Θ, D_α—domains of admissible values of discrete decision variables
C_L, C_M, C_D—finite sets of constraints limiting the values of decision variables

- C_L, C_M—sets of conditions constraining the set of potential behaviors of the i-th elementary structure [5],
- C_D—a set of sufficient conditions the satisfaction of which guarantees congestion-free (i.e. deadlock-free and collision-free) flow of traffic in a transport network modeled by the i-th elementary structure and, execution of transport operations and loading/unloading operations (i.e. operations competing for access to common resources).

The sought solution to problem (1) is schedule ${}^{(i)}X'$ which satisfies all the constraints of the family of sets $\{C_L, C_M, C_D\}$. Constraints C_L, C_M [5] ensure that local processes in the "uncovered" form of an elementary structure are executed in a cyclic manner, i.e. the execution of operations is specified by an appropriate cyclic schedule; they do not guarantee, however, the same for the "covered" form of this structure. The additional constraints C_D given below, which follow from the match-up rule [3] that conditions the fit between cyclic schedules, guarantee that the local processes occurring in the structures which satisfy them are executed in a cyclic manner.

4 Conditions for Preventing Congestion

It is quite obvious that efficient execution of a set of operations which are part of transport processes competing for access to shared network resources is possible when there is no overlap between execution times. An illustration of a condition determining collision-free execution of operations belonging to two processes competing for access to a shared resource is supplied by the description of the following situation. Periods (time intervals) of execution of two operations $o_{i,j}$ and $o_{q,r}$ performed on shared resource $R_{k,i}$ do not overlap if operation $o_{i,j}$ performed on the shared resource starts at moment $x_{i,j}$, i.e. after the resource has been released (with a possible delay Δt) by operation $o_{q,r}$ (at moment x_{q,r^*} scheduled for starting the next operation), and releases the resource (at moment x_{i,j^*} scheduled for starting the next operation) before the next execution of operation $o_{q,r}$ is initiated (at moment $x_{q,r} + \alpha_q$). This means that conflict-free execution of local process operations, i.e. one that does not cause deadlocks, is possible if the following condition holds:

$$\begin{aligned}&[(x_{i,j} \geq x_{q,r^*} + k'' \cdot \alpha_q + \Delta t) \wedge (x_{i,j^*} + k' \cdot \alpha_i + \Delta t \leq x_{q,r} + \alpha_q)] \vee \\ &\vee [(x_{q,r} \geq x_{i,j^*} + k' \cdot \alpha_i + \Delta t) \wedge (x_{q,r^*} + k'' \cdot \alpha_q + \Delta t \leq x_{i,j} + \alpha_i)]\end{aligned} \tag{2}$$

where: $j^* = (j+1)\, MOD\, lr(i)$, $r^* = (r+1)\, MOD\, lr(q)$, $lr(i)$—number of operations in process P_i, αb—cycle of b-th local process

$$k' = \begin{cases} 0 \, for \, j+1 \leq lr(i) \\ 1 \, for \, j+1 > lr(i) \end{cases} \quad k'' = \begin{cases} 0 \, for \, r+1 \leq lr(q) \\ 1 \, for \, r+1 > lr(q) \end{cases}$$

Satisfaction of condition (2) means that for each shared resource of the "covered" representation of an elementary structure, the local processes in that structure are executed alternately. Condition (2) includes instances in which two processes compete for access to a shared resource. In the case when there are three or more processes competing for access to a shared resource, then the condition must be satisfied for each pair of these processes, which means that in the general case one should expect $n(n-1)/2$ trials, where n is the number of processes competing for access to a shared resource. Of course, in the general case, other conditions can also be considered, the satisfaction of which would involve non-alternating order of access to shared resources, for example, conditions imposing an order in which the first two executions of an operation of a first process are followed by three successive executions of an operation from a second process, and then after two consecutive operations of the first process, there follow three successive executions of the operation from the second process, and so on.

When applying condition (2), it is easy to notice that in the case when two elementary structures in which local processes are executed in a cyclic manner are combined, as a result of which some of their resources become shared resources, the local processes performed in this newly created structure are also executed in a cyclic fashion. This means that when cyclic schedules $X_a^{'}$ and $X_b^{'}$ of the elementary structures "a" and "b" that are being combined in this way are characterized, respectively, by cycles α_a and α_b, such that $\alpha_a \, MOD \, \alpha_b = 0$ or $\alpha_b \, MOD \, \alpha_c = 0$, then cycle α_c of schedule $X_c^{'}$ of the newly created structure "c" is equal to the lowest common multiple of periods α_a and α_b, i.e. $\alpha_c = LCM(\alpha_a, \alpha_b)$.

Introduction of condition (2) which allows the formulation of constraint satisfaction problem (1), also has another consequence, which is the theorem on the cyclic character of transport processes performed in networks with fractal structures. The theorem is presented in a verbal form below.

Theorem *Given is a transport network with a fractal structure composed of copies of an elementary structure (i.e. multiple elementary structures of the same shape). If local transport processes performed in the "covered" form of the elementary structure are executed in a cyclic manner, then the local transport processes carried out across the transport network under consideration, which is composed of "uncovered" forms of the elementary structure, are also executed in a cyclic manner.*

The proof of this theorem follows from the assumption of a cyclic character of both the "covered" and the "uncovered" forms of elementary transport structures and the way elementary structures are combined in a transport network with a fractal structure. As mentioned previously, elementary structures overlap at nodes corresponding to elementary structure resources which are shared with neighboring structures of the whole transport network. This case is illustrated by the situation shown in Fig. 5a, in which a vertex of a local network corresponding to resource ${}^{k}R_{12}^{'}$ is glued together with a vertex corresponding to shared resource ${}^{l}R_{12}$.

As an illustration of the approach, consider the transport network model shown in Fig. 5a. Problem (1) considered for the selected k-th elementary network structure was implemented and solved in the constraint programming environment OzMozart (CPU Intel Core 2 Duo 3 GHz RAM 4 GB). When the assumption was made that all operation times in local processes are the same and equal to $t_{i,j} = 1$ u. t. (unit of time), the first acceptable solution was obtained in less than 1 s. An analysis of cyclic schedule ${}^{k}X'$ allows an easy deduction of cycle length $\alpha_k = 7$ u.t., time of freight transport in the MTP considered $T_{mpt} = 19$ u.t., and dispatching rules, for example: ${}^{k}\sigma_{12} = (P_3^1, P_6^1)$, ${}^{k}\sigma_{12} = (P_2^1, P_4^1)$, etc.

5 Conclusion

The declarative reference model of a transport system presented in this study enables an analysis of the relationships between the structure of the system and its potential behavior, thus allowing formulation and solving of analysis and synthesis problems corresponding to questions such as: Is it possible to make supplies which meet customer demands in a transport network with a preset structure? Is there a transport network structure that ensures deliveries which meet user expectations? Because this model focuses on transport networks with a fractal structure, it allows one to formulate a constraint satisfaction problem and, in particular, determine the constraints of this problem in the form of sufficient conditions, the satisfaction of which guarantees smooth execution of traffic flows in this type of networks. These conditions, when implemented in commercially available constraint programming platforms, allow rapid prototyping of alternative transport routes and associated schedules in polynomial time.

The issues of planning and/or prototyping of alternative structures and/or behavior of transport networks with fractal structures presented in this work are part of the broader topic of cyclic scheduling which includes problems occurring in tasks associated with determining timetables, telecommunications transmissions, production planning, etc. In future, while continuing along the line of inquiry related to preventing traffic flow congestion in transport networks, we plan to broaden the scope of our research to include the problems of robust scheduling and the related problem of preventing re-scheduling of timetables in urban transport networks.

References

1. Bahrehdar, S.A., Moghaddam, H.R.G.: A decision support system for urban journey planning in multimodal public transit network. Int. J. Adv. Railway Eng. **2**(1), 58–71 (2014)
2. Banaszak, Z., Bocewicz, G.: Declarative modeling for production orders portfolio scheduling. Found. Manage. **6**(3), 7–24 (2014)

3. Bocewicz, G., Banaszak, Z., Pawlewski, P.: Multimodal cyclic processes scheduling in fractal structure networks environment. In: Proceedings of the 19th World Congress: The International Federation of Automatic Control, pp. 8939–8946. Cape Town (2014)
4. Bocewicz, G.: Robustness of multimodal transportation networks. Eksploatacja i Niezawodność – Maintenance Reliab. **16**(2), 259–269 (2014)
5. Bocewicz, G., Muszyński, W., Banaszak, Z.: Models of multimodal networks and transport processes. Bull. Pol. Acad. Sci. Tech. Sci. **63**(3), 636–650 (2015)
6. Buhl, J., Gautrais, J., Reeves, N., Sol´e, R.V., Valverde, S., Kuntz, P., Theraulaz, G.: Topological patterns in street networks of self-organized urban settlements. Eur. Phys. J. B **49**, 513–522 (2006)
7. Courtat, T.: Walk on city maps—mathematical and physical phenomenology of the city, a geometrical approach. In: Modeling and Simulation. Université Paris-Diderot, Paris VII (2012)
8. Dang, Q.-V., Nielsen, I., Steger-Jensen, K., Madsen, O.: Scheduling a single mobile robot for part-feeding tasks of production lines. J. Intell. Manuf. **25**(6), 1271–1287 (2014)
9. Duy, N.P., Currie, G., Young, W.: New method for evaluating public transport congestion relief. In: Proceedings of the Conference of Australian Institutes of Transport Research (CAITR), 33rd, 2015. Melbourne, Victoria, Australia (2015)
10. Relich, M.: A computational intelligence approach to predicting new product success. In: Proceedings of the 11th International Conference on Strategic Management and its Support by Information Systems, pp. 142–150 (2015)
11. Sandkuhl, K., Kirikova, M.: Analysing enterprise models from a fractal organisation perspective—potentials and limitations. In: Lecture Notes in Business Information Processing, vol. 92, pp. 193–207 (2011)
12. Sitek, P., Wikarek, J.: A hybrid framework for the modelling and optimisation of decision problems in sustainable supply chain management. Int. J. Prod. Res. 1–18 (2015)
13. Sun, Y., Maoxiang Lang, M., Wang, D.: Optimization models and solution algorithms for freight routing planning problem in the multi-modal transportation networks: a review of the state-of-the-art. Open Civ. Eng. J. **9**, 714–723 (2015)
14. Susan, J.P.: Vehicle re-routing strategies for congestion avoidance. New Jersey Institute of Technology, 139 pages (2014)
15. Zhang, J., Liao, F., Arentze, T., Timmermans, H.: A multimodal transport network model for advanced traveler information systems. Procedia Soc. Behav. Sci. **20**, 313–322 (2011)
16. http://www.gostreetmaps.com/new-york.html
17. http://www.mapsguides.com/m/brazil_detailed_street_map_brasilia_en.php

Modelling of Switching Networks with Multiservice Traffic by the IPGBMT Method

Mariusz Głąbowski and Michał Dominik Stasiak

Abstract This paper presents a new analytical method for approximation of blocking probabilities in multiservice switching networks with point-to-group selection. The method is based on the classic PGBMT method which has been adequately modified to improve the accuracy of the final results. In order to estimate and verify the correctness of all theoretical assumptions adopted in the study, the results of the analytical modelling are compared with the results of the digital simulation and the results obtained by the classic PGBMT method in a number of selected three-stage multiservice Clos networks.

Keywords Switching networks · Blocking probability · Multiservice traffic

1 Introduction

Effectiveness and parameters of telecommunications and computer networks largely depend on the operational quality of switching devices, such as switches, routers, etc. Switching devices are based on switching networks, while the latter can be divided into blocking and non-blocking networks [1, 2]. The internal blocking phenomenon is non-existent in non-blocking networks, but, due to a large number of switches required, these networks are virtually not used in practice. Blocking networks require fewer switches, but are susceptible to the internal blocking phenomenon. In practice, any construction of blocking networks that are currently used has to ensure the minimum level of the internal blocking probability [2–4].

M. Głąbowski (✉)
Chair of Communication and Computer Networks, Poznan University of Technology, ul. Polanka 3, 60-965 Poznań, Poland
e-mail: mariusz.glabowski@put.poznan.pl

M.D. Stasiak
Department of Operations Research, Poznan University of Economic and Business, Al Niepodległości 10, 60-875 Poznań, Poland
e-mail: michal.stasiak@ue.poznan.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_16

In order to determine the internal blocking probability, this article proposes a new method for point-to-group blocking probability calculation in multiservice switching networks, the so-called Iterative Point-to-Group Blocking with Multi-rate Traffic (IPGBMT) method. The method is based on the classic Point-to-Group Blocking with Multi-rate Traffic (PGBMT) method [5] using the concept of effective availability [6, 7] and traffic distribution associated with the phenomena of internal and external blocking. Such an approach was applied earlier to evaluate blocking probabilities in switching networks with point-to-point selection and resource reservation at output links of switching networks [8]. As compared to the PGBMT method and other computational methods for a determination of the point-to-group blocking probability, the IPGBMT method makes it possible to evaluate the blocking probability in multiservice switching networks more accurately.

The further part of the paper is organizes as follows. Section 2 describes switching networks with point-to-group selection. The concept of the effective availability methods is presented in Sect. 3. In Sect. 4 the new IPGBMT method is proposed. Section 5 concludes the paper.

2 Multiservice Switching Network with Point-to-Group Selection

Figure 1 shows a multiservice three-stage Clos network that will provide a basis for the analysis proposed in this paper. In each stage, the network consists of k symmetrical switches of k x k links, each with the capacity of f Basic Bandwidth Units (BBU) [9]. The network services M classes of multiservice traffic with the intensity $A(1),\ldots, A(M)$, that require respectively $t(1),\ldots, t(M)$ BBUs to set up a connection. The outputs of the switching network are grouped into directions in such a way that each i-th output of each of the switches of the last stage belongs to the i-th direction of the output direction.

Our assumption is that the network operates in the point-to-group selection mode and that a new call of class i appears at its input. The control algorithm first checks whether there are free output links in a given direction that can service the call of class i. (i.e. links that have at least $t(i)$ free BBUs). If all output links in a given direction are occupied, the control algorithm will reject this call due to external blocking. Otherwise, the control algorithm attempts to set up a connection to a selected switch of the last stage, having a free output link. If the connection fails to be set up, the algorithm will select another switch of the last stage that has a free output link in the demanded direction and will retry to set up a connection. When setting up a connection with all switches of the last stage (having free outgoing links) is not possible, the call of class i will be lost due to the internal blocking.

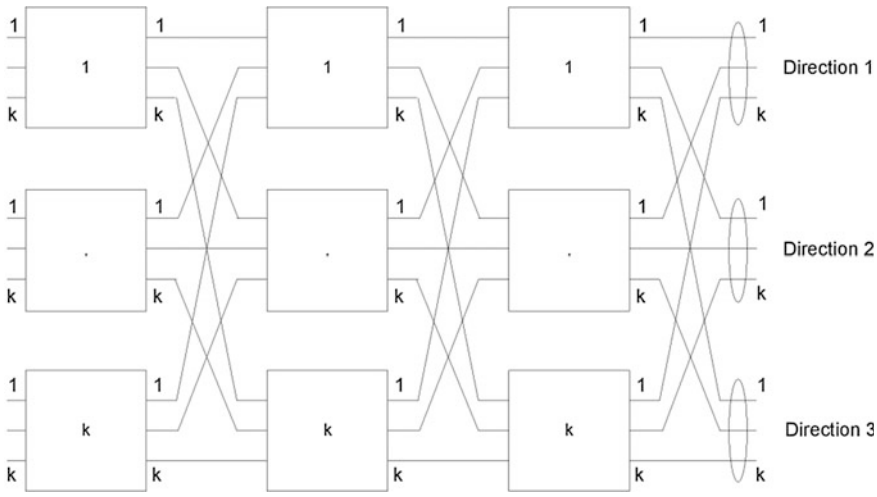


Fig. 1 Three-stage multiservice Clos network

3 The PGBMT Method

The PGBMT method [5] allows the total, internal and external blocking probability in a multiservice switching network to be approximated. The internal blocking probability is determined on the basis of the following formula:

$$E_{int}(i) = \sum_{s=1}^{v-d(i)} \frac{P(i,s)}{1-P(i,0)} \left[\frac{\binom{v-s}{d(I)}}{\binom{v}{d(i)}} \right], \tag{1}$$

where:

v—capacity of the output group (direction), expressed in the number of links. For the network in Fig. 1 we have then:

$$v = k \tag{2}$$

$d(i)$—effective availability for calls of class i in multiservice switching network. Effective availability determines the average number of switches of the last stage (regardless of the occupancy state of output links in these switches in a given direction) with which a connection of class i can be set up. The method for a calculation of this parameter will be given in a further section of this article.
$P(i, s)$—distribution of free links in an output link (direction). This distribution defines the probability that s output links in a given direction can service a call of class i. In this article, the distribution is determined on the basis of a model of the

limited-availability group (LAG). The model will be presented further on in the article.

Formula (1) shows that internal blocking in the network, for a call of class *i,* occurs when $d(i)$ available output links in a given direction have no sufficient number of free BBUs to service a call of class *i.*

In the PGBMT method, the external blocking probability can be approximated by the distribution $P(i, s)$:

$$E_{ext}(i) = P(i, 0). \tag{3}$$

The total blocking probability in the PGBMT method is a sum of the external blocking probability and the internal blocking probability, with the fact taken into consideration that events of internal and external blocking cannot occur simultaneously:

$$E_{tot}(i) = E_{ext}(i) + E_{int}(i)[1 - E_{ext}(i)]. \tag{4}$$

To calculate the blocking probability $E_{\text{tot}}(i)$ for calls of class i, it is necessary to know the effective availability $d(i)$ and the distribution of available links of the output group $P(i, s)$.

3.1 Effective Availability in Switching Networks

Effective availability for calls of class i determines the average number of switches of the last stage with which a connection of class i can be set up (irrespectively of the occupancy state of output links in these switches in the considered direction). Effective availability for calls of class i in a three-stage network can be expressed by the following formula [5]:

$$d(i) = [1 - \pi(i)]k + \pi(i)b(i) + \pi(i)[k - b(i)][1 - b(i)]b(i) \tag{5}$$

where:

k—capacity of output group (direction), expressed in the number of links (Fig. 1),
$b(i)$—fictitious load of the inter-stage link in the switching network for calls of class i,
$\pi(i)$—probability of direct non-availability for calls of class i.

Fictitious load of the inter-stage link $b(i)$ for calls of class i in the switching network corresponds to the blocking probability for calls of class i in an inter-stage link of the switching network. This parameter can be determined on the basis of a model of the multiservice full-availability group (FAG) [10, 11]:

$$nP[n]_f = \sum_{i=1}^{M} A_{FAG}(i)t(i)P[n - t(i)]_f \tag{6}$$

$$b(i) = \sum_{n=f-t(i)+1}^{f} P[n]_f, \tag{7}$$

where:

$A_{\mathrm{FAG}}(i)$—intensity of traffic of class i offered to one inter-stage link,
$P[n]_f$—occupancy distribution in FAG, determines the occupancy probability n BBUs in the full-availability group with the capacity f BBUs.

The probability of direct non-availability $\pi(i)$ for calls of class i in the switching network can be determined on the basis an analysis of a probability graph. This graph shows all possible connection paths in the switching network between a required pair of switches of the first and last stage. Each edge of the graph is assigned the value of fictitious load $b(i)$, determined by Formula (7). The probability $\pi(i)$ in the three-stage switching Clos network can be determined using Lee's method [12], and it is equal [5]:

$$\pi(i) = \{1 - [1 - b(i)]^2\}^k \tag{8}$$

3.2 Limited-Availability Group (LAG)

The limited-availability group (LAG) [13] is a model of a system that consists of k separate component links with identical capacity, equal to f BBU. The capacity of system V is determined by Formula (2). A call of a given class can be serviced only when it can be serviced by free BBUs of one (any randomly selected) link that is included in the group. The distribution $P(i, s)$ in Formulas (1), (3) determines the probability that s output links in LAG can service a call of class i:

$$P(i, s) = \sum_{n=0}^{V} P[n]_V P(i, s|V - n), \tag{9}$$

where $[P_n]_V$ is the occupancy distribution in LAG with the total capacity V BBUs:

$$nP[n]_V = \sum_{i=1}^{M} A_{LAG}(i)t(i)\varsigma_n(i)P[n - t(i)]_V, \tag{10}$$

where $A_{\mathrm{LAG}}(i)$ is the intensity of traffic of class i offered to a given direction. The parameter $\varsigma_n(i)$ in (10) is the conditional transitional probability for a traffic stream of class i. This parameter determines the probability of such a distribution of free BBUs in the occupancy state n BBUs that allows a call of class i in LAG to be serviced:

$$\varsigma_n(i) = \frac{F(V-n,k,f,0) - F(V-n,k,t(i)-1,0)}{F(V-n,k,f,0)}. \tag{11}$$

The combinatorial function $F(x, k, f, t)$ expresses the number of arrangements of x free BBUs in k links, each with the capacity f BBUs, with the assumption that initially t free PJP were arranged in each link:

$$F(x,k,f,t) = \sum_{j=0}^{\left\lfloor \frac{x-kt}{f-t+1} \right\rfloor} (-1)^j \binom{k}{j} \binom{x-k(t-1)-1-j(f-t+1)}{k-1}. \tag{12}$$

Function $F(x, k, f, t)$ allows the distribution $P(i, s|V-n)$, i.e. the conditional distribution of free links, in Formula (9) to be determined. This distribution determines the probability that s output links in LAG can service a call of class i, conditioned by the fact that V-n BBUs in the group are unoccupied:

$$P(i,s|x) = \frac{\binom{k}{s} \sum_{w=st(i)}^{\Psi} F(w,s,f,t(i)) F(x-w,k-s,t(i)-1,0)}{F(x,k,f,0)}. \tag{13}$$

where $\Psi = sf$ if $x \geq sf$ and $\Psi = x$ if $x < sf$.

3.3 Algorithm for the Operation of the Classic PGBMT Method

The sequence of calculations in the PGBMT method in a determination of the blocking probability in a multiservice switching network with point-to-group selection can be written in a simplified form as the following algorithm:

1. Introduction of structural parameters and parameters of traffic offered to multiservice switching network $A_{\mathrm{FAG}}(i)$ i $A_{\mathrm{LAG}}(i)$.
2. Determination of occupancy distribution and blocking probabilities $b(i)$, where $1 \leq i \leq M$, in FAG that approximates one link of multiservice switching network (Formulas (6), (7)).
3. Calculation of the values of effective availability parameters $d(i)$, where $1 \leq i \leq M$, in the multiservice switching network (Formulas (5), (8)).
4. Determination of the distribution of free links $P(i, s)$, where $1 \leq i \leq M$, in LAG that approximates the output direction of the switching network (Formulas (9)–(13)).
5. Determination of the internal $E_{\mathrm{int}}(i)$ (Formula (1)), external $E_{\mathrm{ext}}(i)$ (Formula (3)) and total blocking probability $E_{\mathrm{tot}}(i)$ (Formula (4)) for each class of traffic i, where $1 \leq i \leq M$.

4 The IPGBMT Method

In analytical models of multiservice switching networks there is a possibility that the phenomenon of internal and external blocking can occur simultaneously. This problem is solved in the PGBMT method by the application of Formula (4) in which a possibility of mutual occurrence of external $E_{\text{ext}}(i)$ and internal $E_{\text{int}}(i)$ blocking can be reduced by truncation of the product $E_{\text{int}}(i)E_{\text{ext}}(i)$ from the sum $E_{\text{int}}(i) + E_{\text{ext}}(i)$.

In reality, a call that is rejected due to the external blocking phenomenon is not offered to the switching network. In the IPGBMT method proposed in the article this problem is solved by the use of the approach presented in [8] in which, to model multiservice switching networks, the fixed point method FPM (Fixed Point Method) is used [14].

The assumption in the proposed IPGBMT method is that output links in a given direction can be offered only this part of the total traffic that is not lost at inter-stage links. In a similar way, inter-stage links can be offered only this part the total traffic that is not lost at the output links of the switching network. Such an approach means that, when determining the internal blocking probability, we take into account only this part of traffic that is not lost due to external blocking. When calculating the external blocking probability, in turn, we use this part of traffic that is not lost due to internal blocking. With these assumptions, traffic of a given class offered to a single inter-stage link can be determined by the following formula:

$$A_{FAG}(i) = A_{FAG}(i)[1 - E_{ext}(i)]. \tag{14}$$

With the adopted assumptions, the intensity of traffic offered to a single direction, the knowledge of which is necessary to determine the external blocking probability, on the basis of Formulas (9)–(13), will be expressed by the formula:

$$A_{LAG}(i) = A_{LAG}(i)[1 - E_{int}(i)]. \tag{15}$$

Since the exclusion of the concurrency of the internal and external blocking events in the proposed model is effected at the level of offered traffic, then the total blocking probability can be written directly in the form of a sum of external and internal blocking probabilities:

$$E_{tot}(i) = E_{ext}(i) + E_{int}(i). \tag{16}$$

4.1 Algorithm for the Operation of the IPGBMT Method

As it follows immediately from the form of Formulas (14), (15), to determine the external blocking probability $E_{\text{ext}}(i)$ and internal blocking probability $E_{\text{int}}(i)$ for

calls of class i in the switching network it is necessary to construct an iterative process that can be written, in its simplified form, as the following algorithm:

1. Introduction of structural parameters and parameters of traffic offered to multiservice switching network $A_{\mathrm{FAG}}(i)$ and $A_{\mathrm{LAG}}(i)$.
2. Setting the iteration step $j = 0$.
3. Determination of initial approximations for external and internal blocking probabilities for all traffic classes i, where $1 \leq i \leq M$:

$$E_{int}^{[0]}(i) = 0, E_{ext}^{[0]}(i) = 0. \tag{17}$$

4. Increase in the iteration step:

$$j = j + 1, \tag{18}$$

5. Determination of the value of traffic offered to a single inter-stage link $A_{\mathrm{FAG}}^{[j]}(i)$ (Formula (14)) and to a single direction of multiservice switching network $A_{\mathrm{LAG}}^{[j]}(i)$ (Formula (15)) for each class i,, where $1 \leq i \leq M$:

$$A_{FAG}^{[j]}(i) = A_{FAG}(i)[1 - E_{ext}^{[j-1]}(i)] \tag{19}$$

$$A_{LAG}^{[j]}(i) = A_{LAG}(i)[1 - E_{int}^{[j-1]}(i)] \tag{20}$$

6. Determination of the occupancy distribution and blocking probabilities $b^{[j]}(i)$, where $1 \leq i \leq M$, in FAG that approximates one inter-stage link of switching network (Formulas (6), (7)).
7. Determination of values of effective availability parameters $d^{[j]}(i)$, where $1 \leq i \leq M$, in multiservice switching network (Formulas (5), (8)).
8. Determination of the distribution of free links $P^{[j]}(i, s)$, where $1 \leq i \leq M$, in LAG approximating the output direction of switching network (Formulas (9)–(13)).
9. Determination of the internal blocking probability $E_{\mathrm{int}}^{[j]}(i)$ (Formula (1)), external blocking probability $E_{\mathrm{ext}}^{[j]}(i)$ (Formula (3)) and total blocking probability $E_{\mathrm{tot}}^{[j]}(i)$ (Formula (4)) for each class of traffic i, where $1 \leq i \leq M$.
10. Validation of the accuracy of calculations for each class of traffic i offered to switching network ($1 \leq i \leq M$):

$$\left| \frac{E_{tot}^{[j]}(i) - E_{tot}^{[j-1]}(i)}{E_{tot}^{[j]}(i)} \right| \leq \varepsilon. \tag{21}$$

(a) If Condition (21) is not satisfied, go to Step 4,
(b) If Condition (21) is satisfied, termination of calculations.

The IPGBMT algorithm presented above assumes that $X^{[j]}$ denotes the value of parameter X in j-th iteration. The parameter ε is the demanded relative error of calculations that determines the accuracy of the iteration process. In this algorithm, the values of the internal, external and total blocking probabilities for individual call classes are determined in each iteration on the basis of offered traffic $A_{\text{FAG}}^{[j]}(i)$ and $A_{\text{LAG}}^{[j]}(i)$. This traffic is determined, in turn, on the basis of the external $E_{\text{ext}}^{[j-1]}(i)$ and internal $E_{\text{int}}^{[j-1]}(i)$ blocking probability, determined in the preceding iteration. The IPGBMT algorithm proposed above is not complicated. It is based on performing a certain number of standardized calculations, which makes the algorithm easily programmable.

4.2 A Comparison of the PGBMT and IPGBMT Methods

To determine the accuracy of the IPGBMT method, the results of the analytical modelling were compared with the results obtained in the digital simulation. Figure 2 shows a comparison of the internal blocking probability, determined by the IPGBMT and PGBMT method as well as the digital simulation for an exemplary network with the following parameters (during the research, IPGBMT method was evaluated for many different switching networks):

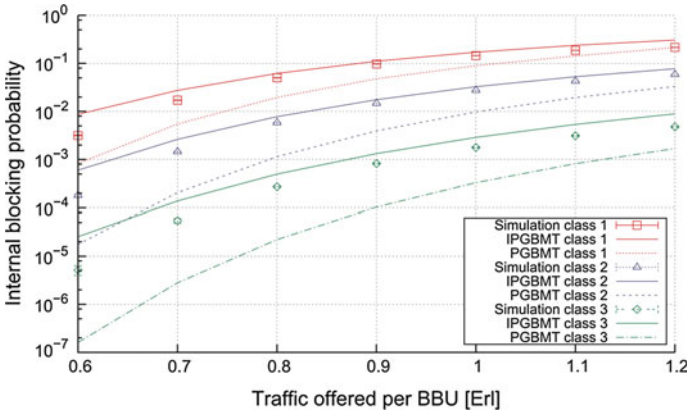


Fig. 2 The internal blocking probability in the switching network

– size of switches in network—4 × 4 links,
– capacity of input, output and inter-stage links—f = 30 BBU,
– selection mode: point-to-group

The network was offered multi-service traffic with the parameters:

– the number of traffic classes M = 3,
– demanded number of BBU for calls of individual classes:

$t(1)$ = 7 BBU, $t(2)$ = 4 BBU, $t(3)$ = 2 BBU,

– proportions of offered traffic: $A(1)t(1) : A(2)t(2) : A(3)t(3) = 1 : 1 : 1$.

In the proposed iterative IPGBMT method, the increase in accuracy of the results of the external blocking probability is not as high as in the case of the internal blocking probability. Consequently, taking into account the limited number of pages of the paper, we have limited ourselves to present the results of the internal blocking probability only.

In Fig. 2 the simulation results are shown in the charts in the form of marks with 95 % confidence interval calculated after the t-Student distribution for five series of minimum 100,000 calls of each class. The results of the modelling are presented in comparison to traffic value offered to one BBU at input links to the switching network.

$$a = \sum_{i=1}^{M} \left(\frac{A_{LAG}(i)t_i}{kf} \right). \tag{22}$$

5 Conclusions

This article presents the IPGBMT method for modelling multiservice switching networks with point-to-group selection. The results of the simulation confirm high accuracy of the method, far better than that of the PGBMT method, especially for the internal blocking probability. This better accuracy is caused by the exclusion of the concurrence of internal and external blocking events in the proposed model at the level of offered traffic. This means that this method can be applied to model, dimension and optimize multiservice switching networks with the point-to-group selection mode.

References

1. Clos, C.: A study of non-blocking switching networks. Bell Syst. Tech. J. **32**(2), 406–424 (1953)
2. Kabaciński, W.: Nonblocking Electronic and Photonic Switching Fabrics. Springer (2005)

3. Lotze, A., Roder, A., Thierer, G.: PCM-charts. University of Stuttgart, Institute of switching and data technics (1979)
4. Głąbowski, M., Stasiak, M.D.: Modelling of multiservice switching networks with overflow links for any traffic class. IET Circuits Devices Syst. **8**(5), 358–366 (2014)
5. Stasiak, M.: Combinatorial considerations for switching systems carrying multi-channel traffic streams. Ann. Telecommun. **51**(11–12), 611–625 (1996)
6. Bininda, N., Wendt, W.: Die effektive Erreichbarkeit fur Abnehmerbundel hinter Zwischenleitungsanungen, Nachrichtentechnische Zeitschricht, Heft 11, No. 12, 1959, s. 579–585
7. Charkiewicz, A.D.: An approximate method for calculating the number of junctions in a crossbar system exchange, Elektrosvyaz, No. 2, 1959, s. 55–63
8. Głąbowski, M., Stasiak, M.D.: Multiservice switching networks with overflow links and resource reservation, mathematical problems in engineering (2016). Article Number: 4090656
9. Roberts, J., Mocci, V., Virtamo, J. (eds.): Broadband Network Teletraffic, Final Report of Action COST 242. Commission of the European Communities, Springer (1996)
10. Kaufman, J.S.: Blocking in a shared resource environment. IEEE Trans. Commun. **COM-29** (10), 1474–1481 (1981)
11. Roberts, J.: A service system with heterogeneous user requirements. In: Pujolle, G. (ed.) Performance of data communications systems and their applications, pp. 423–431. North Holland Pub Co, Amsterdam (1981)
12. Lee, C.: Analysis of switching networks. Bell Syst. Tech. J. **34**(6), 1287–1315 (1955)
13. Stasiak, M.: Blocking probability in a limited-availability group carrying mixture of different multi-channel traffic streams. Ann. Telecommun. **48**(1–2), 71–76 (1993)
14. Kelly, F.P: Fixed point models of loss networks. J. Aust. Math. Soc. **B31**, s. 319–378 (1989)

Time to Buffer Overflow in a Finite-Capacity Queueing Model with Setup and Closedown Times

Wojciech M. Kempa and Iwona Paprocka

Abstract A single-channel queueing model with finite buffer capacity, Poisson arrivals and generally distributed processing times is investigated. According to frequent energy saving requirements, after each busy period the service station is being switched off during a randomly distributed closedown time. Similarly, the first processing in each busy period is preceded by a random setup time, during which the service process is suspended and the machine is being switched on and achieves full readiness for the processing. A system of Volterra-type integral equations for the distribution of the time to the first buffer overflow, conditioned by the initial level of buffer saturation, is built, by applying the idea of embedded Markov chain and continuous version of total probability law. Using the linear algebraic approach, the solution of the corresponding system written for Laplace transforms is obtained explicitly.

Keywords Buffer overflow · Closedown time · Finite-capacity queue · Setup time · Transient state

1 Introduction and Preliminaries

As it is well known, queueing models, in particular with finite buffer capacities, can be successfully used in modelling different-type real issues typical for telecommunication and computer networks, manufacturing processes, transport organiza-

W.M. Kempa (✉)
Faculty of Applied Mathematics, Institute of Mathematics,
Silesian University of Technology, 23 Kaszubska Str.,
44-100 Gliwice, Poland
e-mail: wojciech.kempa@polsl.pl

I. Paprocka
Faculty of Mechanical Engineering, Institute of Engineering Processes Automation
and Integrated Manufacturing Systems, Silesian University of Technology,
18A Konarskiego Str., 44-100 Gliwice, Poland
e-mail: iwona.paprocka@polsl.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_17

tion and logistics. A special role in modelling play systems with less or more complex mechanisms limiting access to the service station. It is associated with the dynamically developing market of different-type solutions for energy consumption reducing and minimizing the costs of system operation. One of such mechanisms are setup and closedown times, occurring at the start and completion epoch of each busy period of the system, respectively. Due to frequent energy saving requirements, the service station is being switched off when there are no jobs waiting for processing, and is being switched on when a job arrives at the empty system. Safe deactivation of the server requires a random time, called a closedown time. Similarly, a randomly distributed setup time is needed for the server to initialize its work and achieve full readiness for processing. In the paper we consider a a single-channel queueing model with finite buffer capacity and the processing organized according to the FIFO service discipline, in which each busy period starts with a setup time and completes with a closedown time. One of the most important stochastic characteristics of each finite-capacity queueing model is the CDF (=cumulative distribution function) of the time to buffer overflow, i.e. the random time from the start epoch of the system to the first moment at which the buffer becomes saturated (the system contains maximal number of jobs). The knowledge of that time is of key importance, e.g. in QoS (=Quality of Service) in telecommunications, since during the buffer overflow period all incoming jobs or packets are lost. In the article, using the approach based on the conception of embedded Markov chain, the continuous version of total probability law and linear algebra, we derive a closed-form representation for the LT (=Laplace transform) of the CDF of the time to buffer overflow, conditioned by the number of jobs accumulated in the buffer at the opening of the system.

As one can observe, in the literature most results for different-type stochastic characteristics of queueing models are found only for the stable systems (stationary state). However, as it seems, non-stationary (transient) analysis is often recommended or directly necessary, e.g. due to high changeability of the traffic (e.g., packet streams in nodes of TCP/IP-based networks), enormous traffic load or unreliable server being subject to breakdowns. In these situations the stationary state of the system is difficult to achieve.

In [1] the queueing model of the M/G/1-type with server breakdowns, setup and closedown times, and with the controlled vacation periods is considered. The stationary results for the batch-arrival queue with multiple vacation policy and server setup and closedown times can be found in [2]. One can find some other results for models with group arrivals in [3, 4]. In [5] the case of discrete time is investigated: a combined control mechanism based on multiple vacation policy and server setup and closedown times is analyzed. The model of a multi-server queue is studied in [6]. Transient results for infinite-capacity systems with server setup times can be found in [7, 8]. In [9] analytical solution for non-stationary queue-size distribution in the queueing system with finite buffer and setup and closedown times is obtained (see also [10]). Similar technique is applied in [11] for the model with server breakdowns. Distributions of the time to buffer overflow in finite-capacity models are studied e.g. in [12, 13]. In [12] the case of MMPP-type (Markov-Modulated

Poisson Process) arrival stream is investigated, while in [13] the general independent arrival process is assumed.

The remaining part of the paper is organized as follows. In the next Sect. 2 we give a detailed mathematical description of the studied queueing model. In Sect. 3 we find a system of integral equations for conditional CDF of the time to buffer overflow. In Sect. 4 we obtain the corresponding system for LTs and write it in a specific form. Section 5 contains main result: the compact-form representation for the LT of conditional distribution of the time to buffer overflow. In the last Sect. 6 a short conclusion can be found.

2 Description of Queueing Model

In the article we deal with a single-channel queueing model with finite capacity of the buffer accumulating jobs (customers, packets, calls, etc.) waiting for processing. We assume that the arrival stream is governed by simple Poisson process with rate λ, while the service time of each job is generally distributed random variable with a CDF $F(\cdot)$. A number of jobs simultaneously present in the system is bounded by a non-random value N, i.e. we have a buffer with $N-1$ places and one place in service station. At the opening of the system, at time $t=0$, a buffer may contain a number of jobs waiting for service in the buffer. Every time when the system empties, i.e. if at the completion epoch of the job service there is no job waiting in the buffer, the service station starts a randomly distributed closedown time with a CDF $C(\cdot)$. If at the completion epoch of a closedown time a buffer contains at least one job waiting for service, the server is being switched on at this moment, otherwise the server waits in the standby mode for the first arrival. The first processing after the idle period is always preceded by a setup time, with random duration with a CDF $S(\cdot)$ (the server begins the setup time simultaneously with the moment of its switching on). Besides, if a job arrives at the system during the closedown time, after its completion epoch the server immediately begins the setup time and next starts the service process. The service station needs the closedown time to be closed safely. Similarly, the setup time is needed for a server to achieve full readiness for processing after the idle period. We assume the well-known FIFO service discipline. Moreover, if the entering job finds the buffer already being overflowed, it is lost (it leaves the system without service).

3 Equations for Conditional CDF of Time to Buffer Overflow

Let Δ_n be the time to the buffer overflow, conditioned by the number of jobs accumulated in the buffer before the opening of the system, i.e.

$$\Delta_n \stackrel{\text{def}}{=} \inf\{t > 0 : X(t) = N | X(0) = n\}, \quad 0 \le n \le N-1, \tag{1}$$

where $X(t)$ denotes the number of jobs present in the system at time t, including the one being processed at this time (if any).

Assume, firstly, that the system is empty at the start moment $t = 0$. In this case we consider $t = 0$ as the moment at which the last service completes during a busy period. Thus, in consequence, a closedown time begins at this time. Let us note that we can distinguish six different mutually excluding random events:

1. the first job arrives during the closedown time, and both the closedown and setup times finishes before time t (let us denote this random event by $E_1(t)$);
2. the first job arrives during the closedown time and before time t, but the following setup time finishes after t ($E_2(t)$);
3. the first job occurs before t and during the closedown time but that period completes after t ($E_3(t)$);
4. the first job enters after the closedown time but before t, and the setup time finishes after t ($E_4(t)$);
5. the first job enters after the closedown time but before t, and the setup time ends before t ($E_5(t)$);
6. the first job occurs after time t ($E_6(t)$).

Observe that the following representations hold true:

$$\boldsymbol{P}\{(\Delta_0 > t) \cap E_1(t)\} = \int_{u=0}^{t} dC(u) \int_{x=0}^{u} \lambda e^{-\lambda x} dx \int_{v=0}^{t-u} \sum_{k=0}^{N-2} \frac{[\lambda(u+v-x)]^k}{k!} e^{-\lambda(u+v-x)} P\{\Delta_{k+1} > t-u-v\} dS(v), \tag{2}$$

$$\boldsymbol{P}\{(\Delta_0 > t) \cap E_2(t)\} = \int_{u=0}^{t} \bar{S}(t-u) dC(u) \int_{x=0}^{u} \lambda e^{-\lambda x} \sum_{k=0}^{N-2} \frac{[\lambda(t-x)]^k}{k!} e^{-\lambda(t-x)} dx, \tag{3}$$

$$\boldsymbol{P}\{(\Delta_0 > t) \cap E_3(t)\} = \bar{C}(t) \sum_{k=1}^{N-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{4}$$

$$\boldsymbol{P}\{(\Delta_0 > t) \cap E_4(t)\} = \int_{u=0}^{t} dC(u) \int_{x=u}^{t} \lambda e^{-\lambda x} \bar{S}(t-x) \sum_{k=0}^{N-2} \frac{[\lambda(t-x)]^k}{k!} e^{-\lambda(t-x)} dx, \tag{5}$$

$$\boldsymbol{P}\{(\Delta_0 > t) \cap E_5(t)\} = \int_{u=0}^{t} dC(u) \int_{x=u}^{t} \lambda e^{-\lambda x} dx \int_{v=0}^{t-x} \sum_{k=0}^{N-2} \frac{(\lambda v)^k}{k!} e^{-\lambda v} P\{\Delta_{k+1} > t - \mathrm{x} - v\} dS(v) \tag{6}$$

and

$$\boldsymbol{P}\{(\Delta_0 > t) \cap E_6(t)\} = e^{-\lambda t}. \tag{7}$$

In the formulae above we use the nomenclature $\bar{G}(t) = 1 - G(t)$, where $G(\cdot)$ denotes arbitrary CDF. Comment shortly the formulae (2)–(7). In (2) and (6) the setup time completes before the moment t, hence the service station starts the processing with the number of jobs which have occurred during the service suspension period. In (3)–(5) and (7), if the condition determined in the definition of the appropriate $E_i(t)$ is satisfied, the time to buffer overflow exceeds t with probability 1.

Now, let us investigate the situation in which at $t = 0$ the buffer contains at least one job, i.e. the level of buffer saturation equals n, where $1 \leq n \leq N - 1$. Due to the fact that successive service completion moments are Markov epochs in the operation of the system, then, by virtue of the continuous version of the total probability law applied in relation to the first departure moment after the opening of the system at time $t = 0$, the following system of Volterra-type integral equations can be written:

$$\boldsymbol{P}\{\Delta_n > t\} = \sum_{k=0}^{N-n-1} \int_0^t \boldsymbol{P}\{\Delta_{n+k-1} > t - x\} \frac{(\lambda x)^k}{k!} e^{-\lambda x} dF(x) + \bar{F}(t) \sum_{k=0}^{N-n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{8}$$

where $1 \leq n \leq N - 1$. Indeed, the first summand on the right side of (8) refers to the case in which the first service finishes at time $x < t$. The second summand presents the situation where the first departure occurs after t, hence the time to buffer overflow exceeds t with probability 1 if and only if at most $N - n - 1$ jobs arrive up to t.

4 Corresponding System of Equations for LTs

In this section we obtain the corresponding system of equations for LTs of conditional CDFs of the time to buffer overflow (more precisely: for the tail of CDF) and write it in a specific form. Let us start with introducing the following notation:

$$\delta_n(s) \stackrel{\text{def}}{=} \int_0^\infty e^{-st} \boldsymbol{P}\{\Delta_n > t\} dt, \quad \text{Re}(s) > 0,\ 0 \leq n \leq N - 1. \tag{9}$$

Now, since $\delta_0(s) = \sum_{i=1}^{6} \int_0^\infty e^{-st} \boldsymbol{P}\{(\Delta_0 > t) \cap E_i(t)\} dt$, then just from the representations (2)–(7) we get

$$\delta_0(s) = \sum_{k=0}^{N-2} \gamma_k(s)\delta_{k+1}(s) + \eta(s), \tag{10}$$

where

$$\begin{aligned}\gamma_k(s) =^{\text{def}} & \frac{\lambda^{k+1}}{(k+1)!} \int_{u=0}^{\infty} e^{-(\lambda+s)u} dC(u) \int_{v=0}^{\infty} e^{-(\lambda+s)v} \left[(u+v)^{k+1} - v^{k+1}\right] dS(v) \\ & + \frac{\lambda}{\lambda+s} \tilde{c}(\lambda+s) \int_{0}^{\infty} e^{-(\lambda+s)v} \frac{(\lambda v)^k}{k!} dS(v),\end{aligned} \tag{11}$$

$$\begin{aligned}\eta(s) =^{\text{def}} & \int_{u=0}^{\infty} dC(u) \int_{t=u}^{\infty} e^{-(\lambda+s)t} \bar{S}(t-u) \sum_{k=0}^{N-2} \frac{\lambda^{k+1}}{(k+1)!} \left[t^{k+1} - (t-u)^{k+1}\right] dt + \sum_{k=1}^{N-1} \int_{0}^{\infty} e^{-(\lambda+s)t} \frac{(\lambda t)^k}{k!} \bar{C}(t) dt \\ & + \int_{t=0}^{\infty} e^{-(\lambda+s)t} dt \int_{u=0}^{t} dC(u) \int_{x=u}^{t} \sum_{k=0}^{N-2} \frac{\lambda^{k+1}}{k!} (t-x)^k \bar{S}(t-x) dx + \frac{1}{\lambda+s}\end{aligned} \tag{12}$$

and

$$\tilde{c}(s) \stackrel{\text{def}}{=} \int_{0}^{\infty} e^{-st} dC(t). \tag{13}$$

Similarly, the system of Eq. (8) will be transformed in the following way:

$$\delta_n(s) = \sum_{k=0}^{N-n-1} a_k(s)\delta_{n+k-1}(s) + \theta_{N-n-1}(s), \tag{14}$$

where $1 \leq n \leq N-1$ and

$$a_k(s) \stackrel{\text{def}}{=} \int_{0}^{\infty} e^{-(\lambda+s)t} \frac{(\lambda t)^k}{k!} dF(t), \tag{15}$$

$$\theta_k(s) \stackrel{\text{def}}{=} \int_{0}^{\infty} e^{-(\lambda+s)t} \bar{F}(t) \sum_{i=0}^{k} \frac{(\lambda t)^i}{i!} dt. \tag{16}$$

Now, let us apply to (10) and (14) the following substitution:

$$\delta_n(s) \stackrel{\text{def}}{=} D_{N-n}(s), \tag{17}$$

where $0 \leq n \leq N-1$. After this transformation (10) and (14) can be reformulated as follows:

$$D_N(s) = \sum_{k=0}^{N-2} \gamma_k(s) D_{n-k-1}(s) + \eta(s), \tag{18}$$

and

$$\sum_{k=-1}^{n-1} a_{k+1}(s) D_{n-k}(s) - D_n(s) = \phi_n(s), 1 \le n \le N-1, \tag{19}$$

where

$$\phi_n(s) \stackrel{\text{def}}{=} a_n(s) D_1(s) - \theta_{n-1}(s). \tag{20}$$

5 Compact Solution for LT of CDF of Time to Buffer Overflow

In [14] the system of equations of type (19) was considered but with infinitely many equations (infinite-sized), namely for $n \ge 1$. Moreover, originally, the system had coefficients being defined by usual number sequences, not by functional sequences as in (19). As it was proved in [14], each solution of (19) can be represented in the following form (we adjust here the original representation from [14] to the case of functional sequences $(a_k(s))$ and $(\phi_k(s))$):

$$D_n(s) = A(s) R_n(s) + \sum_{k=1}^{n} \phi_k(s) R_{n-k}(s), \quad n \ge 1, \tag{21}$$

where $A(s)$ is a function of variable s which is independent on n, and successive terms of the functional sequence $R_k(s), k \ge 0$, can be found from the following recursion, by using coefficients $a_k(s), k \ge 0$:

$$\begin{aligned} &R_0(s) = 0, \\ &R_1(s) = a_0^{-1}(s), \quad R_{k+1}(s) = R_1(s) \left(R_k(s) - \sum_{i=0}^{k} a_{i+1}(s) R_{k-i}(s) \right), \quad k \ge 1. \end{aligned} \tag{22}$$

Let us note that, due the fact that the number of equations in (19) is finite, we can use the Eq. (18) written for $n = N$ as a boundary condition which allows for expressing the function $A(s)$ explicitly. Hence we can find the representation for the unknown function $D_n(s)$ from (21) in a compact form.

Starting with substituting $n = 1$ into (21), we obtain

$$D_1(s) = A(s)R_1(s). \tag{23}$$

Next, taking in (21) $n = N$ and referring to (23), we get

$$\begin{aligned} D_N(s) &= A(s)R_N(s) + \sum_{k=1}^{N} \phi_k(s)R_{N-k}(s) \\ &= A(s)R_N(s) + \sum_{k=1}^{N} [a_k(s)D_1(s) - \theta_{k-1}(s)]R_{N-k}(s) \\ &= A(s)\left(R_N(s) + R_1(s)\sum_{k=1}^{N} a_k(s)R_{N-k}(s)\right) - \sum_{k=1}^{N} \theta_{k-1}(s)R_{N-k}(s). \end{aligned} \tag{24}$$

Introducing (21) into (18), we can rewrite (18) as follows:

$$\begin{aligned} D_N(s) = &\sum_{k=0}^{N-2} \gamma_k(s)\left\{A(s)R_{N-k-1}(s) + \sum_{i=1}^{N-k-1} [A(s)a_i(s)R_1(s) - \theta_{i-1}(s)]R_{N-k-1-i}(s)\right\} \\ &+ \eta(s) = A(s)\sum_{k=0}^{N-2} \gamma_k(s)\left[R_{N-k-1}(s) + R_1(s)\sum_{i=1}^{N-k-1} a_i(s)R_{N-k-1-i}(s)\right] \\ &+ \eta(s) - \sum_{k=0}^{N-2} \gamma_k(s)\sum_{i=1}^{N-k-1} \theta_{i-1}(s)R_{N-k-1-i}(s). \end{aligned} \tag{25}$$

Comparing the right sides of (24) and (25), we eliminate $A(s)$ in the following way:

$$A(s) = \Psi_1(s)\Psi_2^{-1}, \tag{26}$$

where

$$\Psi_1(s) \stackrel{\text{def}}{=} \eta(s) - \sum_{k=0}^{N-2} \gamma_k(s)\sum_{i=1}^{N-k-1} \theta_{i-1}(s)R_{N-k-1-i}(s) + \sum_{k=1}^{N} \theta_{k-1}(s)R_{N-k}(s) \tag{27}$$

and

$$\begin{aligned} \Psi_2(s) \stackrel{\text{def}}{=} &R_N(s) + R_1(s)\sum_{k=1}^{N} a_k(s)R_{N-k}(s) \\ &- \sum_{k=0}^{N-2} \gamma_k(s)\left[R_{N-k-1}(s) + R_1(s)\sum_{i=1}^{N-k-1} a_i(s)R_{N-k-1-i}(s)\right]. \end{aligned} \tag{28}$$

Collecting the formulae (17), (20), (21) and (26), we can state the following main theorem:

Theorem 1 *The representation for the LT $\delta_n(s)$ of the conditional tail CDF of the time to the buffer overflow in the M/G/1/N-type finite-capacity queueing system with generally distributed setup and closedown times is following*:

$$\begin{aligned}\delta_n(s) &= \int_0^\infty e^{-st}\boldsymbol{P}\{\Delta_n > t\}dt \\ &= \Psi_1(s)\Psi_2^{-1}\left(R_{N-n}(s) + R_1(s)\sum_{k=1}^{N-n} a_k(s)\right) - \sum_{k=1}^{N-n}\theta_{k-1}(s)R_{N-n-k}(s), \quad (29)\end{aligned}$$

where the formulae for $\Psi_1(s), \Psi_2(s)$, $R_k(s), a_k(s)$ and $\theta_k(s)$ are given in (27), (28), (22), (15) and (16), respectively.

6 Conclusion

In the article a single-channel queueing model with finite buffer capacity and a mechanism of setup-closedown times of the service station is investigated. The arrival stream is described by a single Poisson process while the processing, setup and closedown times are generally distributed random variables. By using the analytical approach based on the concept of embedded Markov chain, continuous version of the total probability law and linear algebra a compact-form representation for the LT of the conditional tail CDF of the time to the buffer overflow is obtained. The final formulae are written in terms of "input" system characteristics and a functional sequence, defined recursively, connected with them.

References

1. Ke, J-Ch.: On M/G/1 system under NT policies with breakdowns, startup and closedown. Appl. Math. Model. **30**, 49–66 (2006)
2. Arumuganathan, R., Jeyakumar, S.: Steady state analysis of a bulk queue with multiple vacations, setup times with N-policy policy and closedown times. Appl. Math. Model. **29**, 972–986 (2005)
3. Ke, J-Ch.: Batch arrival queues under vacation policies with server breakdowns and startup/close-down times. Appl. Math. Model. **31**, 1282–1292 (2007)
4. Krishna, G.V., Reddy, R., Nadarajan, R.: Arumuganathan: Analysis of a bulk queue with N-policy multiple vacations and setup times. Comput. Oper. Res. **25**, 957–967 (1998)
5. Moreno, P.: A discrete-time single-server queueing system under multiple vacations and setup-closedown times. Stochast. Anal. Appl. **27**, 221–239 (2009)
6. Artalejo, J.R., Economou, A., Lopez-Herrero, M.J.: Analysis of a multiserver queue with setup times. Queueing Syst. **52**, 53–76 (2005)

7. Kempa, W.M.: The transient analysis of the queue-length distribution in the batch arrival system with N-policy, multiple vacations and setup times. In: Venkov, G., Kovacheva, R., Pasheva, V. (eds.) Applications of Mathematics in Engineering and Economics (AMEE-10), 36th International Conference, 5–10 June 2010, Sozopol, Bulgaria, Melville, American Institute of Physics, (AIP Conference Proceedings, vol. 1293), pp. 235–242 (2010)
8. Kempa, W.M.: On transient queue-size distribution in the batch arrival system with the N-policy and setup times. Math. Commun. **17**, 285–302 (2012)
9. Kempa, W.M., Paprocka, I.: Analytical solution for time-dependent queue-size behavior in the manufacturing line with finite buffer capacity and machine setup and closedown times. In: Slatineanu, L. et al. (ed.) Selected, Peer Reviewed Papers from the 19th Innovative Manufacturing Engineering 2015 (IManE 2015), May 21–22, 2015, pp. 1360–1365. Iasi, Romania. Zurich, Trans Tech Publications (2015) (Applied Mechanics and Materials, vol. 809/810)
10. Kempa, W.M., Paprocka, I., Grabowik, C., Kalinowski, K.: Time-dependent solution for the manufacturing line with unreliable machine and batched arrivals. In: Modern Technologies in Industrial Engineering (ModTech2015), 17–20 June 2015. Mamaia, Romania, Bristol, Institute of Physics Publishing (2015), 1–6 (IOP Conference Series, Materials Science and Engineering, vol. 95)
11. Kempa, W.M.: On transient virtual delay in a finite-buffer queueing model with server breakdowns. In: Information systems architecture and technology. In: Grzech, A. et al. (ed.) Selected Aspects of Communication and Computational Systems, pp. 77–86. Wrocław University of Technology, Wrocław (2014)
12. Chydziński, A.: Time to buffer overflow in an MMPP queue. In: International Conference NETWORKING 2007: Ad Hoc and Sensor Networks, Wireless Networks Next Generation Internet (Lecture Notes in Computer Science, vol. 4479), pp. 879–889
13. Kempa, W.M.: On the distribution of the time buffer overflow in a queueing system with a general-type input stream. In: Proceedings of the 35th International Conference on Telecommunications and Signal Processing (TSP-2012), 3–4 July 2012, pp. 207–211. Prague, Czech Republic, IEEE (2012). DOI:10.1109/TSP.2012.6256283)
14. Korolyuk, V.S.: Boundary-Value Problems for Compound Poisson Processes. Naukova Dumka, Kiev (1975)

The Multiple Criteria Optimization Problem of Joint Matching Carpoolers and Common Route Planning

Modeling and the Concept of Solution Procedure

Grzegorz Filcek and Jacek Żak

Abstract The paper concerns a joint problem of matching carpoolers and planning common routes as a multiple criteria optimization problem. The model of the optimization problem, as well as the concept of solution procedure, is presented. The solution algorithm uses dynamic programming method and Dijkstra algorithm to obtain matches of the carpoolers and routes. The solution is evaluated by aggregated quality index taking into account preferences of carpoolers equally, concerning all the criteria. Preferences are obtained by AHP method.

Keywords Carpooling · Modelling · Multiple criteria decision making · Optimization · Dynamic programming · Route planning

1 Introduction

In recent years, much attention of researchers has been focused on specific traveling mode which is carpooling (also car-sharing, ride-sharing, lift-sharing) (see. [1–3] as a survey). Carpooling is the sharing of car journeys so that more than one person travels in a car. The main aims for using this mode of transport is cost sharing, as well as reduction of air pollution, traffic congestion and the need for parking spaces.

The idea of carpooling is not new (it was already promoted during II world war [4]), and is commonly practiced for instance in the USA, where even road infrastructure promotes this kind of transportation. It is done by providing so-called HOV lanes (High-occupancy vehicle lane) which have progressed slowly since the 1970s

G. Filcek (✉)
Faculty of Computer Science and Management, Wroclaw University of Science and Technology, Wroclaw, Poland
e-mail: Grzegorz.Filcek@pwr.edu.pl

J. Żak
Faculty of Machines and Transportation Logistics Division, Poznan University of Technology, Poznan, Poland
e-mail: Jacek.Zak@put.poznan.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_18

(oil and energy crisis) [5]. HOV lanes are also implemented in Canada, Europe, Australia and New Zealand, Indonesia and China, however carpooling is much less popular in these regions. In Europe, carpooling idea is being developed and promoted mainly by Internet means of communication like social media (e.g. Facebook) and web services like BlaBlaCar [6]. The Internet also helps to evolve the idea of real-time ridesharing by the use of mobile devices like smartphones and tablets [7]. However, those systems still do not support a process of automatic connecting people and finding the routes for them, which would enable possibly best association between passengers and drivers. Consequently, some dedicated applications are developed. However, their capabilities are limited, because there is still a lack of the algorithms, which enable taking into account all fundamental real life carpooling problem requirements [8].

The task of common route planning for carpoolers consists in finding an acceptable, compromise path that would satisfy the interests of a driver (car owner) and different passengers at a rational, balanced cost. In many cases, the interests mentioned above of carpoolers may have a contradictory character, and their objectives may differ. Thus, the proposed (generated) solution must be a compromise one and guarantee a certain level of satisfaction for all of them. In these circumstances, the "carpooling problem" may be seen as multiple criteria multiple-path optimization problem. Similar problems were considered by many authors [9–21].

However, despite results presented in the work of Knapen et al. [22], the algorithms, given in the mentioned papers, concern finding one path (i.e. the problem is solved only from one person's point of view). Moreover, neither constraints concerning the times of travel beginning and completion nor the dependencies between paths of travelers (drivers and passengers) are considered.

In this paper we focus on the problem of making associations between carpoolers, taking into account their travel preferences to suggest the best routes for the drivers. The internet services allow carpoolers to search journeys defined by the drivers willing to travel with other people mainly to share travel costs. We propose a mechanism that can help carpoolers to find fellow travelers and a route which will suit best to all of them. The idea of AHP (Analytical Hierarchy Process) [23] is used to collect preferences of the carpoolers and use it to obtain the best routes presented to drivers. The AHP preference vector is used to compute aggregated cost function for common routes of the driver and passengers assigned to him/her. The route for the driver is constructed from sub-routes defined as paths between the consecutive pickup and drop off points of the passengers where origin and destination point of the driver is also treated as pick up and drop off point, respectively. The sub-routes are obtained by Dijkstra algorithm [24]. The whole route is built with the use of dynamic programming approach (see e.g. [25]).

The paper is organized as follows. After the introduction, a mathematical model with problem formulation is given in Sect. 2. Next section presents the idea of the solution algorithm Sect. 4 concludes the paper.

2 Mathematical Model

Let us denote the graph $G = <V, E>$ (where V corresponds to a set of nodes and E reflects a set of links) that defines a map of the area that carpoolers operate. It is assumed that there are K drivers who travel from the origin node $\underline{v}_k \in V$ to their destination $\bar{v}_k \in V$, $(k \in \overline{1,K})$.[1] They are willing to reduce their travel cost by offering a carpool service, not compromising at the same time much of the comfort of travel. To achieve this goal, passengers are allowed to join their journey. The driver may take no more passengers than the capacity of the vehicle (number of available seats in his/her car given by $(S_k \in \mathbf{N}_+)$. There are also P passengers who eager to go with drivers offering a carpool service. Drivers and passengers are called carpoolers and indexed by c (for drivers $c \in \overline{1,K}$, for passengers $c \in \overline{K+1, K+P}$). Any carpooler c may start his/her trip only in the node $\underline{v}_c \in V$ and must finish his/her trip in the node $\bar{v}_c \in V$. We assume without loss of generality that every node associated with every passenger is disjunctive with any other node associated with any other passenger. Furthermore, it is supposed that any driver's starting or final node is disjunctive with nodes (either starting or finals) associated with any passenger. Carpoolers define constraints concerning their M journey criteria by providing values of $\tilde{w}_{m,c}$, $(c \in \overline{1,K+P}\ m \in \overline{1,M})$ by default, $\tilde{w}_{m,c} = \bar{M}$ ($\tilde{w}_{m,c}$ is the biggest acceptable value of the mth criterion for carpooler c, $\bar{M}$ is a very large number). The example criteria may have an interpretation of the travel time, distance, travel cost, comfort, punctuality, or picturesqueness of the view. Additionally, carpoolers define their time windows, composed of the earliest $(\underline{t}_c^E(\underline{v}_c))$ and latest $(\underline{t}_c^L(\underline{v}_c))$ departure time from the origin node as well as earliest $(\bar{t}_c^E(\bar{v}_c))$ and latest $(\bar{t}_c^L(\bar{v}_c))$ arrival time to the destination node.

For the sake of simplicity, we introduce the variable $\bar{e}_{i,j}$ describing the existence of an arc in G, $\bar{e}_{i,j} = 1(0)$ if there exists an arc in G connecting ith node with the jth node (otherwise).

Moreover, there are weights related to every arc in G, reflecting properties such as traveling time, distance, and cost. The variable $w_{i,j,m}$, element of the matrix $\mathbf{W} = [w_{i,j,m}]$ $i = 1, 2, \ldots, |V|$, $j = 1, 2, \ldots, |V|$, $m = 1, 2, \ldots, M$. denotes the value of property type m on the arc connecting nodes i and j in G.

The first decision to be made is to assign passengers to the drivers by indicating the nodes that form the set of activity nodes. This assignment should be featured by a good match between passengers and drivers and their objectives. The set of activity nodes (AC) includes nodes of the graph G, in which the driver begins or ends the route or has to stop because of picking up or dropping off a passenger. Let us denote a binary decision variable $y_{k,d} \in \{0, 1\}$ describing such assignment of the

[1] We use $\overline{1,Z} \triangleq \{1, 2, 3, \ldots, Z\}$ notation for a set of natural numbers from 1 to Z.

dth node of the graph G to the set of activity nodes (AC) in the path of the kth driver, an element of the matrix $\mathbf{y} = [y_{k,d}]_{\substack{k=1,2,\ldots,K \\ d=1,2,\ldots,|V|}}$. Let us introduce the following constraints for decision matrix $\mathbf{y}$

$$\forall_{k\in\overline{1,K}}(\forall_{d:v_d\in\{\underline{v}_k,\bar{v}_k\}}(y_{k,d}=1)\wedge\forall_{p\in\overline{K+1,K+P}}\forall_{(a,b):v_a=\bar{v}_p\wedge v_b=\underline{v}_p}(y_{k,a}=y_{k,b}), \tag{1}$$

$$\forall_{k\in\overline{1,K}}\sum\nolimits_{d:v_d\in\bigcup\limits_{c\in\overline{1,K}\setminus\{k\}}(\{\underline{v}_c,\bar{v}_c\})}y_{k,d}=0, \tag{2}$$

$$\forall_{p\in\overline{K+1,K+P}}\forall_{d:v_d=\underline{v}_p}\sum_{k=1}^{K}y_{k,d}\leq 1, \tag{3}$$

where (1) assures that for every passenger who travels with the kth driver, his origin node, and his destination node are included in the kth driver's path as well as the origin and destination nodes of this driver. Constraint (2) excludes from AC of the kth driver's path all origin nor destination nodes that belong to other drivers. The assignment of each passenger to no more than exactly one driver is reached when (3) is satisfied.

Before a travel path for each driver can be obtained, there is another decision to be made, which consists of defining the appropriate order of nodes in AC that the kth driver has to visit. This order builds the overriding path for the driver, which cannot have any cycle. To model this decision let us introduce a binary decision variable $z_{k,a,b}\in\{0,1\}$ describing the existence of a path from the node $v_a\in V$ to $v_b\in V$ which belong to the kth driver's AC, an element of the matrix $\mathbf{z}=[z_{k,a,b}]_{\substack{k=1,2,\ldots,K \\ a=1,2,\ldots,|V| \\ b=1,2,\ldots,|V|}}$. The decision matrix $\mathbf{z}$ must, in this case, satisfy the following constraints:

$$\forall_{k\in\overline{1,K}}\forall_{a\in\overline{1,|V|}}\sum_{b=1}^{|V|}(y_{k,a}y_{k,b}(z_{k,a,b}-z_{k,b,a}))=\begin{cases}1 & \text{for} \quad v_a=\underline{v}_k\\ -1 & \text{for} \quad v_a=\bar{v}_k\\ 0 & \text{otherwise}\end{cases}, \tag{4}$$

$$\forall_{k\in\overline{1,K}}\forall_{a\in\overline{1,|V|}}\sum_{b=1}^{|V|}(1-y_{k,a})(z_{k,a,b}+z_{k,b,a})=0, \tag{5}$$

$$\forall_{k\in\overline{1,K}}\forall_{a\in\overline{1,|V|}\setminus\{d:v_d=\underline{v}_k\}}\sum_{b=1}^{|V|}y_{k,a}z_{k,b,a}=1, \tag{6}$$

where (4) describes the law of flow preservation for each driver from the origin of his/her travel to the final destination. Constraint (5) assures that there are no routes

from or to the nodes that do not belong to the AC of any driver. To ensure, that $\mathbf{z}$ describes only a simple path between nodes in each driver's AC, (6) is introduced.

When these conditions are satisfied, the final path for each driver can be obtained. Let us denote by $x_{k,a,b} \in \{0,1\}$ a binary decision variable describing the existence of the edge (v_a, v_b) in the kth driver's path between nodes v_a and v_b belonging to the kth driver's AC, an element of the matrix $\mathbf{x} = [x_{k,a,b}]_{\substack{k=1,2,\ldots,K \\ a=1,2,\ldots,|V| \\ b=1,2,\ldots,|V|}}$. Let us formulate appropriate constraints for matrix $\mathbf{x}$:

$$\forall_{k\in\overline{1,K}}\forall_{a,b\in\overline{1,|V|}} \sum_{d=1}^{|V|}(x_{k,a,b} - x_{k,b,a}) = \begin{cases} 1 & \text{for } \sum_{c=1}^{|V|} z_{k,a,c} = 1 \\ -1 & \text{for } \sum_{c=1}^{|V|} z_{k,c,a} = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

$$\forall_{k\in\overline{1,K}}\forall_{a,b\in\overline{1,|V|}} \sum_{c=1}^{|V|} x_{k,a,c} = \sum_{c=1}^{|V|} x_{k,c,b} = z_{k,a,b}, \tag{8}$$

$$\forall_{k\in\overline{1,K}}\forall_{a,b\in\overline{1,|V|}}\ x_{k,a,b} \le \bar{e}_{a,b}, \tag{9}$$

where (7) assures the law of flow preservation between nodes v_a and v_b. Constraint (8) ensures that each driver's path is aggregated from simple sub-paths only (with no cycles) connecting consecutive nodes in the overriding path described by $\mathbf{z}$. By inequality (9) we assure, that the path belongs to G. The variables $\mathbf{y}$ and $\mathbf{z}$ are auxiliary and strictly depend on $\mathbf{x}$, but have their interpretation and help to follow the idea of the model. The dependence between them and main decision variable $\mathbf{x}$ is as follows: $y_{k,\bar{a}} = \max_b\{z_{k,\bar{a},b}; z_{k,b,\bar{a}}\} = \max_b\{x_{k,\bar{a},b}; x_{k,b,\bar{a}}\}$, $z_{k,a,b} = \sum_{d=1}^{|V|} x_{k,a,d} \cdot \sum_{d=1}^{|V|} x_{k,d,b}$.

Let us introduce another auxiliary variable $\alpha_{k,p,d} \in \{0,1\}$, that assumes value 1 when the pth passenger is in the kth driver's car before reaching by this car node v_d, and 0 otherwise. The index $\alpha_{k,p,d}$ for each driver k can be calculated using the following recurrent formulas:

$$\forall_{k\in\overline{1,K}}\forall_{p\in\overline{K+1,K+P}}\ (v_a = \underline{v}_k \wedge \alpha_{k,p,a} = 0), \tag{10}$$

$$\forall_{k\in\overline{1,K}}\forall_{p\in\overline{K+1,K+P}}\ \alpha_{k,p,d} = \sum_{a=1}^{|V|}(\alpha_{k,p,a} \cdot x_{k,a,d} + \sum_{b:v_b=\underline{v}_p} x_{k,a,b} - \sum_{b:v_b=\bar{v}_p} x_{k,a,b}), \tag{11}$$

To assure that the number of passengers who travel with one driver will never exceed the capacity of the driver's car the following constraint is given:

$$\forall_{k\in\overline{1,K}}\forall_{d\in\overline{1,|V|}}\sum_{p=K+1}^{K+P}\alpha_{k,p,d}\le S_k. \tag{12}$$

The decision must also satisfy carpoolers' constraints concerning pickup and drop-off times, as well as the beginning and end of drivers' travel times. Let us assume, that one of the weights on the arc is travel time, and let it be noted as $w_{a,b,1}$ (travel time between nodes a and b), then

$$\forall_{k\in\overline{1,K}}\ \overline{t}_k^E(\bar{v}_k)-\underline{t}_k^E(\underline{v}_k)\le\sum\nolimits_{a,b}w_{a,b,1}x_{k,a,b}\le\overline{t}_k^L(\bar{v}_k)-\underline{t}_k^L(\underline{v}_k), \tag{13}$$

$$\forall_{p\in\overline{K+1,K+P}}\forall_{k\in\overline{1,K}}\ \overline{t}_p^E(\bar{v}_p)-\underline{t}_p^E(\underline{v}_p)\le\sum\nolimits_{a,b}w_{a,b,1}x_{k,a,b}\alpha_{k,p,b}\le\overline{t}_p^L(\bar{v}_p)-\underline{t}_p^L(\underline{v}_p), \tag{14}$$

Of course, the path common for some carpoolers must satisfy their constraints, what supports the following formulas

$$\forall_{m\in\overline{1,M}}\forall_{k\in\overline{1,K}}\sum\nolimits_{(a,b)}\left(w_{a,b,m}x_{k,a,b}\right)\le\tilde{w}_{m,k}, \tag{15}$$

$$\forall_{m\in\overline{1,M}}\forall_{k\in\overline{1,K}}\forall_{p\in\overline{K+1,K+P}}\sum\nolimits_{(a,b)}\left(w_{a,b,m}x_{k,a,b}\alpha_{k,p,b}\right)\le\tilde{w}_{m,p}. \tag{16}$$

We evaluate the decision $\mathbf{x}$ by applying the following vector of criteria

$$\begin{aligned}\bar{Q}(\mathbf{x})\overset{\Delta}{=}[&\sum_{a,b\in\overline{1,|V|}}\sum_{k\in\overline{1,K}}w_{a,b,1}x_{k,a,b};\ \sum_{a,b\in\overline{1,|V|}}\sum_{k\in\overline{1,K}}w_{a,b,2}x_{k,a,b};\ \ldots\\ &\ldots;\ \sum_{a,b\in\overline{1,|V|}}\sum_{k\in\overline{1,K}}w_{a,b,M}x_{k,a,b};\ \sum_{k\in\overline{1,K}}\sum_{p\in\overline{K+1,K+P}}\sum_{d:v_d=\underline{v}_p}(1-x_{k,p,d})],\end{aligned} \tag{17}$$

where each objective is to be minimized.

Obtaining the solution of this multi-criteria optimization problem as a Pareto front is a very hard task (computational complexity is NP-hard even for one carpooler). We propose to build aggregated quality index based on the carpoolers preferences acquired by AHP method. We collect from each carpooler its preferences according to the importance of every of all M criteria using pairwise comparison. Each of the stakeholders may be treated equally or not. We assume that none of the carpoolers is more preferred than other. The preferences are used, by applying scalarization of (17), to obtain the global cost function in a form:

$$Q(\mathbf{x}) \stackrel{\Delta}{=} \frac{\sum\limits_{k\in\overline{1,K}} \left(\sum_{(a,b)} \sum\limits_{m\in\overline{1,M}} \left(\sum\limits_{c\in\overline{K+1,K+P}} (\bar{w}_{m,c}\alpha_{k,c,b}) + \sum\limits_{c\in\overline{1,K}} \bar{w}_{m,c}\right) f_m(w_{a,b,m}) x_{k,a,b}\right.}{\left(\sum\limits_{p\in\overline{K+1,K+P}} (\alpha_{k,p,b}) + 1\right)} \\ \left. + \sum\limits_{p\in\overline{K+1,K+P}} \sum\nolimits_{d:v_d=\underline{v}_p} (1-\alpha_{k,p,d})\bar{M}\right) \tag{18}$$

where $\bar{w}_{m,c}$ are the preference weights of the criterion obtained from carpooler c, using the AHP. f_m is a normalization function used to score each factor (weight of the arc) under criterion m, and $\bar{M}$ is a very large value that corresponds to the situation when some of the passengers are not assigned to the driver.

We propose the following form of function f_m based on min and max values of the given factor from all arcs in the network, i.e.

$$f_m(w_{a,b,m}) \stackrel{\Delta}{=} \frac{w_{a,b,m} - \min\limits_{\tilde{a},\tilde{b}}(w_{\tilde{a},\tilde{b},m})}{\max\limits_{\tilde{a},\tilde{b}}(w_{\tilde{a},\tilde{b},m}) - \min\limits_{\tilde{a},\tilde{b}}(w_{\tilde{a},\tilde{b},m})}. \tag{19}$$

Using (18) a single criteria optimization problem may be formulated.

Problem formulation.

For the given data: G, $\bar{e}_{i,j}$ $(i = 1,2,\ldots,|V|, j = 1,2,\ldots,|V|)$, $\mathbf{W}$, M, K, P, $\underline{v}_c$, $\bar{v}_c$, $\underline{t}^L_c(\underline{v}_c)$ $\underline{t}^E_c(\underline{v}_c)$, $\bar{t}^E_c(\bar{v}_c)$, $\bar{t}^L_c(\bar{v}_c)$, $\bar{w}_{m,c}$, $\tilde{w}_{m,c}$, $(c = 1,2,\ldots,K+P,\ m = 1,2,\ldots,M)$, S_k, $(k = 1,2,\ldots,K)$, determine $\mathbf{x}^*$ feasible with respect to constraints (1)-(16) to minimize (18), i.e.

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} Q(\mathbf{x}). \tag{20}$$

3 Solution

To obtain a solution, a few steps must be undertaken. First, the preferences of the drivers and passengers are collected mainly via AHP. The preferences represented by weights of the criteria are aggregated and used to build the quality index (18). Then a dynamic programming is used to obtain the passenger list with pick up and drop off points for every driver. During execution of the algorithm, we use Dijkstra's algorithm to obtain best sub-routes between consecutive pickup or drop off points. It is worth noticing that the problem solved by dynamic programming, in this case, can be reduced to a multiple knapsack problem, which by definition is NP-hard. The heuristic or approximation algorithm should be developed to deal with the problem for big instances, e.g. for a large number of carpoolers operating in a small area. For a little number of carpoolers on a large area, a decomposition

may be applied. It may divide carpoolers into groups with similar objectives and locations in order to solve smaller instances of the problem. For example, passengers that can be assigned to a driver must have similar travel direction, and their travel origins and destinations should not be too far from the route of the driver that he would take if traveling alone. These ideas will be developed in further works.

3.1 Preferences of the Carpoolers

We collect the preferences of the drivers and passengers that concern criteria by asking them to make a pairwise comparison of all criteria considered in the problem. It is worth emphasizing, that the number of compared criteria should not be in this case greater than 5 (max 7), while with the grow of the number of criteria n by one, the number of comparisons increase by n – 1. So, for a larger number of comparisons, the carpooler may not provide the correct (consistent) data or even resign from using the system due to the fact that he/she may perceive it as being too complicated and exhaustive in terms of obtaining/generating the carpoolers' preferences. We assume that every carpooler is treated equally, so the opinion of any of them has the same weight. This assumption, in general, does not have to be true. The weights obtained from the pairwise comparison are used further as weights of the criteria in aggregated evaluation function applied for each arc of the network by which carpoolers travel. We collect not only preferences concerning criteria, but also preferences that define constraints ($\bar{w}_{m,c}$, ($c \in \overline{1, K+P}$)), e.g. maximal travel time or maximal route length. This knowledge is used to define the main optimization problem (see Sect. 2).

3.2 Dynamic Programming

One of the possible ways to solve the defined problem is to divide the decision into a string of consecutive N decisions and then apply dynamic programming. To use this method we define a state, decisions, and local objective function for a decision-making moment n in the following way:

Let $\mathbf{s}(n) = ([(s1_k(n), s2_k(n))]_{k=0,1,2,\ldots,K}, s3(n))$ represent the state, where $s1_k(n) \subseteq \overline{K+1, K+P}$ is the set of passengers traveling with driver k. The driver $k = 0$ is a virtual one and has unlimited vehicle capacity, $\underline{v}_0$ and $\bar{v}_0$ are additional artificial nodes), $s2_k(n) \in \cup_{c=0}^{K+P}\{\underline{v}_c, \bar{v}_c\}$ is the last node that driver k has visited, and $s3(n) \subseteq (\cup_{k=0}^{K} \bar{v}_k) \cup (\cup_{p=K+1}^{K+P}(\{\underline{v}_p, \bar{v}_p\}))$ is the set of origin and destination nodes of the passengers and destination nodes of the drivers to be considered in the nth decision-making moment.

Let $\mathbf{u}(n) = (u1(n), u2(n))$ be the decision made in moment n, where $u1(n) \in (\cup_{k=0}^{K} \bar{v}_k) \cup (\cup_{p=K+1}^{K+P}(\{\underline{v}_p, \bar{v}_p\}))$ is the next node to assign to a driver, and $u2(n) \in \overline{0,K}$ is the number of the driver that will serve this node.

The next state $\mathbf{s}(n+1)$ depends on the current state and decision in the following way: $s1_k(n+1) = \begin{cases} s1_k(n+1)\backslash\{p\} & \text{if} \quad k = u2(n) \wedge u1(n) = \bar{v}_p \\ s1_k(n+1) & \text{if} \quad k \neq u2(n) \\ s1_k(n+1) \cup \{p\} & \text{if} \quad k = u2(n) \wedge u1(n) = \underline{v}_p \end{cases}$, $s2_k(n+1) = u1(n)$, $s3(n+1) = s3(n)\backslash\{u1(n)\}$. To evaluate the decision, we use the following quality index $g(\mathbf{s}(n), u(n)) \stackrel{\Delta}{=} \begin{cases} \bar{M}/2 & \text{if} \quad u2(n) = 0 \\ \bar{g}(s2_{u2(n)}(n), u1(n)) & \text{in other case} \end{cases}$, where $\bar{g}(s2_{u2(n)}(n), u1(n))$ is achieved from the solution of the shortest path problem between nodes $s2_{u2(n)}(n)$ and $u1(n)$. It is interpreted as the path length for the shortest path problem. In fact, it has the interpretation of cost function just as (18). The objective function of the dynamic programming problem is the sum of $g(\mathbf{s}(n), u(n))$ for a string of decisions made from state $\mathbf{s}(0) = ([(\emptyset, \underline{v}_k)]_{k=0,1,\ldots,K}, (\cup_{k=0}^{K} \bar{v}_k) \cup (\cup_{p=K+1}^{K+P}(\{\bar{v}_p, \underline{v}_p\})))$ to state $\mathbf{s}(N) = ([(\emptyset, \bar{v}_k)]_{k=0,1,\ldots,K}, \emptyset)$, i.e.

$$Q((\mathbf{s}(0), \mathbf{s}(N), u(0), \ldots, u(N-1)) = \sum_{n \in 0, N-1} g(\mathbf{s}(n), u(n)), \tag{21}$$

where $N = K + 2P + 1$. The decisions in the nth moment are constrained as follows. The node to be considered must belong to the set defined by $s3(n)(u1(n) \in s3(n))$ and cannot be the destination node of passenger, whose origin node has not been assigned to any driver yet $(\forall_{p \in \overline{K+1,K+P}: \{\underline{v}_p, \bar{v}_p\} \subseteq s3(n)} (u1(n) \neq \bar{v}_p))$. The drivers destination node cannot be selected if the driver is still travelling with passengers $(u1(n) \notin \{v \in V : k = 1, 2, \ldots, K \wedge v = \bar{v}_k \wedge s1_k(n) \neq \emptyset\})$. The driver may serve any origin node or the destination node of the passenger he/she is transporting or his/her $(u2(n) \in \{k \in \overline{1,K} : u1(n) = \bar{v}_p \wedge p \in s1_k(n) \vee u1(n) \in (\bigcup_{p=K+1}^{K+P} \underline{v}_p) \cup \bar{v}_k\})$. We also take user preferences concerning constraints as follows. For the given state $\mathbf{s}(n)$ and decision $u(n)$ for every driver k, using Dijkstra algorithm, we obtain the shortest paths given by $\mathbf{x}^*(n) = [x_{k+1,a,b}(n)]$ $k = 0, 1, , \ldots, K$, $a = 1, 2, \ldots, |V|$, $b = 1, 2, \ldots, |V|$. ($x_{d,a,b}(n)$ is the value of $x_{d-1,a,b}$ in nth moment, for $d > 1$) connecting origin node ($\underline{v}_k$) of the driver with the last node he visits, given by $u1(n)$.

$$\forall_{m\in\overline{1,M}}\forall_{k\in\overline{1,K}}\sum\nolimits_{(a,b)}\left(w_{a,b,m}x_{k+1,a,b}(n)\right)\leq\tilde{w}_{m,k}, \tag{22}$$

$$\forall_{m\in\overline{1,M}}\forall_{k\in\overline{1,K}}\forall_{p\in s1_k(n)}\sum\nolimits_{(a,b)}\left(w_{a,b,m}x_{k+1,a,b}(n)\alpha_{k+1,p,b}(n)\right)\leq\tilde{w}_{m,p}, \tag{23}$$

where $\alpha_{d,p,b}(n)$ is the value of $\alpha_{d-1,p,b}$ in moment n for $d>1$, and calculated analogically as in (10)–(11). For the optimal solution obtained by this method, the values of quality indices (21) and (18) are equal.

3.3 Shortest Path

We apply the Dijkstra algorithm for the search of the shortest way from node v_1 to v_2 in moment n.

To evaluate each link between node a and b, we use local cost function given by the following equation:

$$\bar{g}_L(a,b,\mathbf{x}(n))\stackrel{\Delta}{=}\frac{\sum_{k\in\overline{1,K}}\sum_{m\in\overline{1,M}}\left(\sum_{c\in\overline{K+1,K+P}}(\bar{w}_{m,c}\alpha_{k+1,c,b}(n)+\sum_{K+1,K+P}\bar{w}_{m,c})f_m(w_{a,b,m})x_{k+1,a,b}(n)\right)}{\left(\sum_{p\in\overline{K+1,K+P}}(\alpha_{k+1,p,b}(n))+1\right)}. \tag{24}$$

So, the whole path can be evaluated by

$$\bar{g}(\mathbf{x}(n))=\sum\nolimits_{(a,b)\in\{(i,j):\{v_i,v_j\}\subseteq V\}}\bar{g}_L(a,b,\mathbf{x}(n)), \tag{25}$$

The objective is to minimize the value of $\bar{g}(\mathbf{x}(n))$

$$\bar{g}(v_1,v_2)\stackrel{\Delta}{=}\min_{\mathbf{x}}\bar{g}(\mathbf{x}(n)), \tag{26}$$

with respect to constraints (7), (9), $\sum_{v\in V}x_{k+1,v_1,v}(n)=\sum_{v\in V}x_{k+1,v,v_2}(n)=1$, and $\forall_{\tilde{v}\in\cup_{c=1}^{K+P}(\{\underline{v}_c,\bar{v}_c\})\setminus\{v_1,v_2\}}\sum_{v\in V}x_{k+1,\tilde{v},v}(n)=\sum_{v\in V}x_{k+1,v,\tilde{v}}(n)=0$.

4 Conclusions

The main contribution of this paper is the model of the joint problem of matching carpoolers and routes planning as a multiple criteria optimization problem, and the solution procedure. The further research concerning this problem will include

exhaustive computational experiments, development of decomposition techniques, approximation, and computationally efficient heuristic algorithms.

References

1. Agatz, N., Erera, A., Savelsbergh, M., Wang, X.: Optimization for dynamic ride-sharing: a review. Eur. J. Oper. Res. **223**(2), 295–303 (2012)
2. Jorge, D., Correia, G.: Carsharing systems demand estimation and defined operations: a literature review. Eur. J. Transp. Infrastruct. Res. **13**(3) (2013)
3. Hwang, K., Giuliano, G.: The Determinants of Ridesharing: Literature Review. Berkeley, California, UCTC Working Paper no 38 (1990)
4. Oliphant, M., Amey, A.: Dynamic Ridesharing: Carpooling Meets the Information Age (2010)
5. Turnbull, K.F.: HOV project case studies: history and institutional arrangements. Washington, D.C.: U.S. Department of Transportation ; Distributed in Cooperation with Technology Sharing Program (1992). Accessed 26 April 2012
6. Wauters, R.: Ride-sharing has arrived in Europe, and the race is on between BlaBlaCar and Carpooling.com (2014). http://tech.eu/features/481/ride-sharing-europe-carpooling-blablacar/. Accessed 11 June 2016
7. Amey, A., Attanucci, J., Mishalani, R.: Real-time ridesharing: opportunities and challenges in using mobile phone technology to improve rideshare services. Transp. Res. Rec.: J. Transp. Res. Board **2217**, 103–110 (2011)
8. Shakshuki, E., Younas, M., Cho, S., Yasar, A.-U.-H., Knapen, L., Bellemans, T., Wets, G.: A conceptual design of an agent-based interaction model for the carpooling application. Procedia Comput. Sci. **10**, 801–807 (2012)
9. Qian, Z., Zhang, M.H.: Modeling multi-modal morning commute in a one-to-one corridor network. Transp. Res. Part C: Emerg. Technol. **19**(2), 254–269 (2011)
10. Androutsopoulos, K.N., Zografos, K.G.: Solving the multi-criteria time-dependent routing and scheduling problem in a multimodal fixed scheduled network. Eur. J. Oper. Res. **192**(1), 18–28 (2009)
11. Bozkurt, S., Yazici, A., Keskin, K.: A multicriteria route planning approach considering driver preferences. In: 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES), pp. 324–328 (2012)
12. Granat, J., Guerriero, F.: The interactive analysis of the multicriteria shortest path problem by the reference point method. Eur. J. Oper. Res. **151**(1), 103–118 (2003)
13. Guerriero, F., Musmanno, R.: Label correcting methods to solve multicriteria shortest path problems. J. Optim. Theory Appl. **111**(3), 589–613 (2001)
14. Horn, M.E.T.: Multi-modal and demand-responsive passenger transport systems: a modelling framework with embedded control systems. Transp. Res. Part A: Policy Pract. **36**(2), 167–188 (2002)
15. Jozefowicz, N., Semet, F., Talbi, E.-G.: Multi-objective vehicle routing problems. Eur. J. Oper. Res. **189**(2), 293–309 (2008)
16. Nadi, S., Delavar, M.R.: Multi-criteria, personalized route planning using quantifier-guided ordered weighted averaging operators. Int. J. Appl. Earth Obs. Geoinf. **13**(3), 322–335 (2011)
17. Niaraki, A.S., Kim, K.: Ontology based personalized route planning system using a multi-criteria decision making approach. Expert Syst. Appl. **36** (2, Part 1), 2250–2259 (2009)
18. Opasanon, S., Miller-Hooks, E.: Multicriteria adaptive paths in stochastic, time-varying networks. Eur. J. Oper. Res. **173**(1), 72–91 (2006)
19. Pajor, T.: Multi-modal route planning. Master thesis, Univ. Karlsruhe (TH) (2009)

20. Yu, H.C., Lu, F.: A multi-modal route planning approach with an improved genetic algorithm. In: Advances in Geo-Spatial Information Science, vol. 1–0, pp. 193–202. CRC Press (2012)
21. Filcek, G., Gąsior, D.: Common Route Planning for Carpoolers – Model and Exact Algorithm. In J. Swiątek, A. Grzech, P. Swiątek, J. M. Tomczak (Eds.), Advances in Systems Science, pp. 543–551. Springer International Publishing (2014)
22. Shakshuki, E., Younas, M., Knapen, L., Keren, D., Yasar, A.-U.-H., Cho, S., Wets, G.: ANT 2012 and MobiWIS 2012 analysis of the co-routing problem in agent-based carpooling simulation. Procedia Comput. Sci. **10**, 821–826 (2012)
23. Saaty, T.L.: Decision making with the analytic hierarchy process. Int. J. Serv. Sci. **1**(1), 83 (2008)
24. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**(1), 269–271 (1959)
25. Sniedovich, M.: Dynamic Programming: Foundations and Principles, Second Edition vol. 20105564. CRC Press (2010)

The Impact of Structure Network Parameters on Consumers Behavior: A Cellular Automata Model

Agnieszka Kowalska-Styczeń

Abstract In this paper, the dynamics of the consumer preference change, affected by a word-of-mouth communication (w-o-m), is investigated. Is seems to be interesting, how the network structure parameters (number of contacts, diversity of groups, possibility to change the source of information) influence the consumer behavior on the duopoly market. In the model presented in this article, a two-dimensional cellular automaton (CA) is used which enables to simulate the number of informal contacts, by adoption of variety of different neighborhood radii and also different neighborhood shapes. Although, one type of networks is examined in this article (CA type model), different network structure parameters are obtained by changing the above neighborhood parameters together with varied population densities and an agent movement possibility. The results indicate that the network structure parameters, are of particular importance in the processes of preferences' change.

Keywords Word of mouth · Consumer behavior · Network structure · Cellular automata · Agent models

1 Introduction

Many works that have appeared in the last decade, have shown the usefulness of specific simulation tools to study social systems. A special role is played by the complex systems paradigm of studying the dynamics of large systems based on local interactions of their elementary components [1]. Various directions of model studies, defined by Gotts et al. [1] as 'based on simulation of social agents', known also as an agent based social simulation—(ABSS), are used to describe different social systems. Agent models provide a bridge between macro and micro level and

A. Kowalska-Styczeń (✉)
Faculty of Organization and Management, Silesian University of Technology, Zabrze, Poland
e-mail: agnieszka.kowalska@polsl.pl

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0_19

use simulations to search for causal mechanisms of social behavior [2]. The agents approach is also frequently used because of the limitations posed by traditional research techniques, especially for the prediction the market shares of new products or the effects of marketing strategies [3].

A typical agent-based model is cellular automata (CA). Proponents of the CA usage in social and economic studies emphasize in particular, that it is possible to test the paradigm of macro-collective behavior based on local relationships, in an easy way [4]. The example of CA model application in marketing may be the study of investment behavior in the stock market [5], the consumer decision model on the network market [6], the model of knowledge diffusion dynamics [7], the study of saddle effect on electronics industry market [8], the study of network externalities effect on the growth rate of a new product [9] and the Bass innovation diffusion model presented by Goldberg et al. [10] and Goldenberg and Efroni [11].

It was shown by Kiesling et al. [12] that agent-based models (ABM) allow not only for modelling of interactions that are responsible for the social impact, but also take into account the structure of social interactions. This analysis, based on the paradigm of complex systems, can provide managers with tools to assist decision-making process in the context of competitive market. Therefore, in this article, consumer behavior in the competing brands market is studied, and a two-dimensional CA is used to analyze some structural parameters of the network of social contacts.

One, of the most important factors influencing consumer behavior is interpersonal impact. The product a person decides to buy, depends on the experience of other purchasers [13]. These are informal networks of contacts, which have long been seen as a key source of social influence in social systems [14]. The basic mechanism of this informal contacts is the word-of-mouth communication (w-o-m). Many studies show that w-o-m communication has a significant impact on consumer choices [15, 16, 17]. In certain circumstances, as shown in Goldsmith and Horowitz [18], this type of communication is more effective than any other form of advertising. The number of informal contacts can be simulated (to some extent), using a variety of different radii of neighborhood and also different neighborhood shapes. The diversity of informal groups sizes can be modelled by both agents placed randomly on the lattice, and also population density changes. Dynamics of changes in informal sources of information can also be entered by the appropriate rules of agents movement.

Research questions can be summarized as follows:

- How does the network structure parameters, determined by the number and diversity of word-of-mouth contacts, influence the tendency of a society as a whole to change their shopping preferences?
- How do the structural network parameters (number of contacts, diversity of groups, possibility to change the source of information) shape the process of a preferences change dynamics?

Although, not different types of networks (e.g., small world, and others) are examined in this article, changes of network structures are obtained by assuming different parameters of the proposed model. The general hypothesis is that, the dynamics of preferences change in a competitive market, depend on network structures. This belief is intuitively fairly obvious and also has been confirmed in a number of empirical studies realized in areas related to our interests, for example, Burt [19], Fleming et al. [20], Granovetter [21]. Moreover Alkemade and Castaldi [22] conclude in their paper that the companies are not fully aware of the structure of channels used for communication with consumers. Also Bohlmann et al. [23] postulated that managers should pay particular attention to the structure of the network, because their understanding and understanding of the influence of this structure on communication processes will enable the company to make marketing decisions more rationally. Therefore, the social networks structure and the dynamics of the process of competition between consumers of one of the two competing firms is examined.

It seems that, the proposed model can facilitate quantitative and qualitative analysis which supports making marketing decision.

2 The Model

The environment of consumers is presented as a square lattice of size L x L with periodic boundary conditions. Each i-th cell (i = 1, 2,…, L2) can be in one of three states: empty, occupied by an agent of type A, or occupied by an agent of type B. In this work, there will be analysis of type A and B as consumers of the two competing firms. For example in work Bouzdine-Chameeva and Galam [24], A and B were interpreted as two competing wine brands.

Initially agents are distributed randomly on the lattice, i.e. with probability $p \in [0, 1]$ a cell is occupied and with probability $1 - p$ it is empty. Probability p is simply an average concentration of agents.

The total number of agents will be donated by N, $N = pL^2 \cdot N_A$ denotes the number of A-agents and $N_B = N - N_A$ number of B-agents. The concentration of A-agents in introduces as $c = N_A/N$ and thus the concentration of B-agents is equal $1 - c$.

At each time step t, all cells i = 1, …, L^2 are updated simultaneously (typical cellular automata approach as described in Wolfram [25–27] and Sarkar [28].

As in other works on marketing [5, 29], the presented model also uses the majority rule. This seems to be a reasonable choice because studies show that people tend to herd behavior, as seen in the stock market, the online auction [5] or behavior of the crowd [30].

Therefore, following local rules are proposed:

- Agent checks the preferences of its neighborhood, and changes its preference for the dominant in the neighborhood (i.e., if more than 50 % of its neighbours have a different opinion than it, then it changes its opinion on the dominant one).
- If more than 50 % have the same opinion as the test agent, of course, nothing changes.
- If there is no dominant preference in the agent neighborhood, then the agent moves to the nearest empty place in a randomly determined direction (north, east, south or west).

Defining the cellular automaton in this way, effects of interpersonal influence are taken into account, and at the same time, mobility in the 'opinion space' is introduced. This way, the model can be interpreted as a variant of a 'mobility game', as in the paper of Yu and Hebling [31]. The general idea of the model is illustrated in Fig. 1.

Possibility of consumer's movement through the lattice distinguishes the proposed model of the known ones proposed in the field of marketing such as Goldenberg et al. [8, 10], Garber et al. [32], Goldenberg et al. [9], Ma and Chao [33]. Consumer mobility is an important assumption, because by changing place in the lattice the sort of an actual consumer behavior can be mapped. In situations of uncertainty consumer decisions are often discussed with different groups of people (this may mean the first step consultation in the family, then in a group of friends or use of the Internet for example).

In the model presented in this paper, family or group of friends are represented by the von Neumann's neighborhood (Fig. 2).

The main modification of the classical approaches in the proposed model is also the density of the lattice filling (population density). The lattice is populated by consumers who are represented by individual cells. Simulations can be carried out for different densities. This makes the fact that the selection of von Neumann's

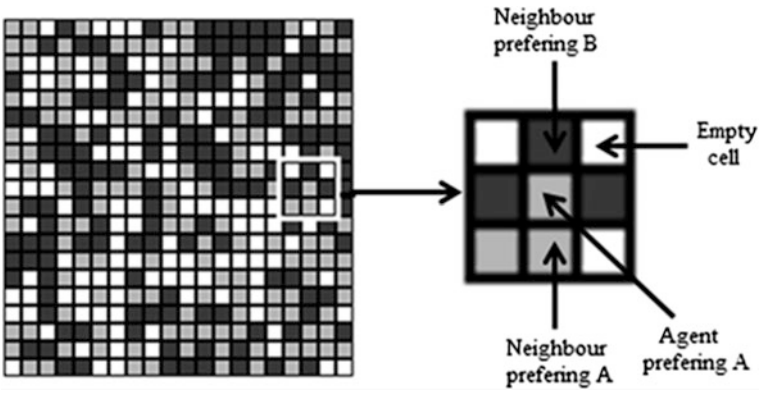


Fig. 1 The general idea of a proposed model

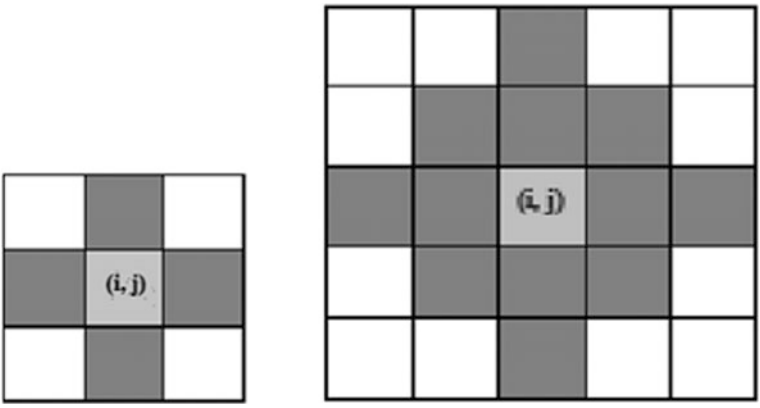


Fig. 2 Von Neumann's neighborhood of a two-dimensional automaton cell (i, j) with r = 1, 4-element neighborhood (*left panel*) and r = 2, 12-element neighborhood (*right panel*)

neighborhood for the agent/consumer may consist of from 0 to 4 when r = 1 and from 0 to 12 people when r = 2. This assumption is a step in the direction of the approximation of the CA model to social reality, where the nearest neighborhood, influencing decisions of individuals, is composed of varying numbers of people (family or friends containing a different number of members).

3 Simulation Experiments

3.1 Experimental Design

The proposed CA model has been verified by tracking variables at intermediate steps in simulation and running the simulation with extreme values of parameters in the work Kowalska-Styczeń and Sznajd-Weron [34]. This is one way to verify the model [35].

As mentioned earlier, experiments that have been conducted here are designed to investigate the effect of selected parameters of the network structure on the preferences change dynamics, for the two competing companies. The parameters, forming the structure of social networks, were thus changed by defining:

- P—population density
- C—the initial concentration of agents opting for a particular firm
- L—size of the lattice
- neighbourhood size
- agent movement possibility

As dependent variables describing an agents/consumers behaviour following parameters were adopted:

- the preferences change risk (standard deviation from the mean value of the final concentration of A agents),
- the number (fraction) of people choosing a particular product,
- In the study, two levels of factors changes were assumed:
- p (population density): 0,3 and 0,7;
- r (von Neumann's neighbourhood radius): 1, 2;
- possibility of agent with or without movement.

Factors c (fraction size, which prefers product A) and L (lattice size) were changed with the step of 5 %.

In the next section there is a discussion concerning the results of simulation. the behavior unpredictability.

3.2 Simulation Results

In simulations designed, consumers had a choice between the two of competing firms products (A and B). Moreover, it was assumed that the number of consumers preferring A- firm products is equal to those preferring B-firm products, except for the experiments in which the change of c (the initial concentration of A consumers) has been analyzed.

Initially, an analysis of the impact of the change of L—lattice size (which reflects the size of the population) on a consumer behavior was performed. The standard deviation of the number of consumers preferring A option from the initial state (c = 0,5) for 1000 samples was determined (Monte Carlo simulations). This measure may be identified with the risk, put on the process of maintaining a stable relationship on the market, shared by two competing firms. It may be also associated with a preferences change tendency in the analyzed community.

Dynamics of preferences change depending on the lattice size, for the two population densities: p = 0,3 (sparse lattice), p = 0,7 (dense lattice) is present in Fig. 3. Results are averaged over 1000 simulation steps.

As can be seen in Fig. 3, a responsiveness of a single agent to w-o-m communication decreases along with the L (size of lattice). In the presented model, with the increase of the lattice size, the social network density (the proportion of the possible connections among these agents that are actually present) decreases. The average number of neighbors for a given population density is constant, so if the lattice size increases, the density of the social network decreases. This is consistent with one of the oldest arguments in social psychology, namely: norms—shared ideas about the proper way to behave are clearer, easier to enforce in a more dense social network [21].

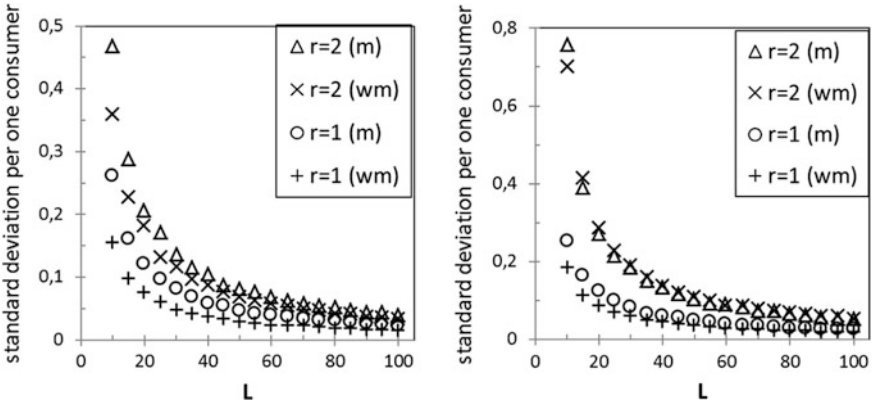


Fig. 3 Standard deviation of the A-agents number from the final concentration A-agents per one consumer, as a function of L, for p = 0,3 (*left panel*) and p = 0,7 (*right panel*). Model with movement—(m), without movement—(wm)

It may also be noted, that the higher population density (p = 0,7) facilitates the change, and therefore, responsiveness sensitivity to w-o-m communication (standard deviation), is greater for dense lattices than for sparse ones (p = 0,3).

In the above experiments, the standard deviation of the A-agents number from the initial state (for c = 0,5; or 50 % for A and 50 % for B) was analyzed. As mentioned, this index is one of the basic measures of risk, the risk of preferences instability, on the market divided between two competing firms.

A major impact on the dynamics of respondents preferences has the size of groups, within which, opinions are confronted and formulated. These groups are represented by the neighborhoods, respectively, with a radius: r = 1 and r = 2. Larger groups of friends, generally cause increase of the preferences change risk. Smaller groups have a better cohesion-performance relationship than larger groups [36], therefore the preferences change risk is greater for larger groups. This risk is also modified by the possibility to consult opinion in other information sources. Possibility to consult opinion with more neighbors, influences the behavior of agents, though not very significantly (see Fig. 3). The availability of additional information sources (the agents movement) has a greater impact in case of smaller population density (p = 0,3).

Subsequently, the impact of the population density (the determinant of a network consistency) on the consumers behavior was explored. Figure 4 depicts the results of preferences' change study for a lattice size L = 50 and for the parameters described above. The lattice size does not affect the shape of the obtained plots curves. The same results are obtained for larger lattices.

As one can see in Fig. 4, the risk of preferences change is very much dependent on the neighborhood size. Larger number of contacts, in the w-o-m mechanism,

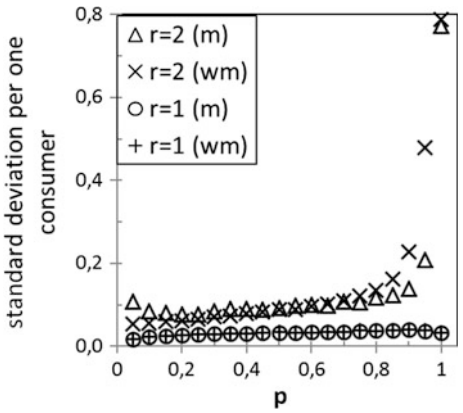


Fig. 4 Standard deviation of the A-agents number from the final concentration A-agents per one consumer, as a function of L, for p = 0,3 (*left panel*) and p = 0,7 (*right panel*). Model with movement —(m), without movement—(wm)

causes much greater risk of changes for the average representative of consumer group. This risk increases along with the population density. The greater the degree of population density (the greater consistency of the social network), the greater the vulnerability to preferences change of consumers. This confirms, among others, the empirical study of Fleming et al. [20], who analyzed data on utility patents in the careers of 35,400 collaborative inventors. This study showed, that high levels of network closure facilitate the change adoption.

The possibility of diversifying sources of information (movement in the opinion space) has not a very significant impact, but slightly increases the level of average deviations.

4 Discussion and Conclusion

The proposed model has been designed to investigate the properties of the market processes connected with a consumer behavior in the two competing products market. In particular, there has been a search for answers to the questions related to the analysis of the impact of the structural network parameters on a consumer behavior. Therefore, there has been an analysis done on how the change of the lattice size (i.e., the size of the studied community), the population density, and the number of agents preferring the product at the beginning of the simulation affect the agents behavior representing the consumers community.

In the study, as a measure of risk the standard deviation of the number of votes for A from the initial state was assumed (one of the key risk measures in the field of finance and economics). Simulation experiments has been performed for the

products of two firms, changing the nature of mobility agents and the type and number of information w-o-m sources. The results, considerably, are the answers to the questions posed in the paper. The most important of these responses can be summarized as follows:

1. In larger communities, the preferences change (modified by w-o-m mechanism) are more stable (larger lattice—less deviation per one consumer). As mentioned earlier, it is connected with the fact that the larger the social group, the lower the density of the social network, because people have cognitive, emotional, spatial and temporal limits on how many social ties they can sustain. Thus, the larger the group, the lower its ability to crystallize and enforce norms [21]. Therefore, it is more difficult to prompt larger communities to change their preferences.
2. Structure of the network, is of particular importance in the processes of preferences changes. The dynamics of these processes is strongly affected by the neighborhood size—that is, the number of w-o-m contacts. Larger quantities of these contacts, cause an increased risk of preferences change (but also, on the other hand, greater tendency to change). The possibility to consult with the other information sources (agents movement) influences the risk, but to a much lesser extent. In general, the tendency to preferences change (risk), decreases with the size of the lattice, and increases with the population density—that is, with the potential number of possible contacts.

It seems that the results generally confirm the potential of simple tool such as the cellular automata. As demonstrated in this article, CA can be used in the complex processes of the market study, especially those concerning the network structure between consumers and the dynamics of decision-making process on the market. Dependences shown above can be useful for a better understanding of these processes, and thereby assist in the development of more efficient and effective marketing campaigns.

The simulation results can also be an incentive to conduct empirical research on the structure of relations between consumers.

Moreover, according to Mason et al. [37], when many individuals interact over time, their behaviors are interdependent, creating a complex, dynamic system that may have unpredictable, emergent outcomes and such results can be obtained also by means of simple models. Such is the case of this article. A simple model with simple local rules was used, but interesting emergent phenomena at the macro level appeared. The local rule is based on the majority rule, which is a typical herd behavior [13, 30] and the results of simulation are consistent with studies of Muller and Copper [36], Granovetter [21] and Fleming et al. [20]. Thus, the results are compatible with the psychological knowledge, as well as empirical research.

Acknowledgments The work was partially financially supported by the Polish National Science Center grant no. UMO-2014/15/B/HS4/04433.

References

1. Gotts, N.M., Polhill, J.G., Law, A.N.R.: Agent-based simulation in the study of social dilemmas. Artif. Intell. Rev. **19**, 3–92 (2003)
2. Macy, M.W., Willer, R.: From factors to actors: computational sociology and agentbased modelling. Annu. Rev. Sociol. **28**, 143–166 (2002)
3. Gilbert, N., Jager, W., Deffuant, G., Adjali, I.: Complexities in markets: introduction to the special issue. J. Bus. Res. **60**, 813–815 (2007)
4. Moldovan, S., Goldenberg, J.: Cellular automata modeling of resistance to innovations: effects and solutions. Technol. Forecast. Soc. Change **71**, 425–442 (2004)
5. Wei, Y., Ying, S., Fana, Y., Wang, B.: The cellular automaton model of investment behavior in the stock market. Phys. A **325**, 507–516 (2003)
6. Frels, J., Heisler, D., Reggia, J., Schuetze, H.: A cellular automata model of competition in technology markets with network externalities. In: Sunderam, V.S., et al. (eds.) ICCS 2005, LNCS 3515, 378–385. Springer-Verlag, Berlin, Heidelberg (2005)
7. Morone, P., Taylor, R.: Knowledge diffusion dynamics and network properties of face-to-face interactions. J. Evol. Econ. **14**, 327–351 (2004)
8. Goldenberg, J., Libai, B., Muller, E.: Riding the Saddie: how cross-market communications can create a major slump in sales. J. Mark. **66**, 1–16 (2002)
9. Goldenberg, J., Libai, B., Muller, E.: The chilling effects of network externalities. Int. J. Res. Mark. **27**, 4–15 (2010)
10. Goldenberg, J., Libai, B., Muller, E.: Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata. Acad. Market. Sci. Rev. 9, 20 pages (2001)
11. Goldenberg, J., Efroni, S.: Using cellular automata modeling of the emergence of innovation. Technol. Forecast. Soc. Change **68**(3), 293–308 (2001)
12. Kiesling, E., Günther, M., Stummer, C., Wakolbinger, L.M.: Agent-based simulation of innovation diffusion: a review. Central Eur. J. Oper. Res. **20**(2), 183–230 (2012)
13. Teraji, S.: Herd behavior and the quality of opinions. J. Socio-Econ. **32**, 661–673 (2003)
14. Brass, D.J.: Being in the right place—a structural analysis of individual influence in an organization. Adm. Sci. Q. **29**, 518–539 (1984)
15. Brown, J.J., Reingen, P.H.: Social ties and word-of-mouth referral behavior. J. Consum. Res. **14**, 350–362 (1987)
16. Hennig-Thurau, T., Walsh, G.: Electronic word-of-mouth: motives for and consequences of reading consumer articulations on the Internet. Int. J. Electron. Commer. **8**(2), 51–74 (2004)
17. Steffes, E.M., Burgee, L.E.: Social ties and online word of mouth. Internet Res. **19**, 42–59 (2009)
18. Goldsmith, R.E., Horowitz, D.: Measuring motivations for online opinion seeking. J. Interact. Advertising **6**(2), 1–16 (2006)
19. Burt, R.S.: Secondhand brokerage: evidence on the importance of local structure for managers, bankers, and analysts. Acad. Manag. J. **50**(1), 119–148 (2007)
20. Fleming, L., Mingo, S., Chen, D.: Collaborative brokerage, generative creativity, and creative success. Adm. Sci. Q. **52**, 443–475 (2007)
21. Granovetter, M.: The impact of social structure on economic outcomes. J. Econ. Perspect. **19** (1), 33–50 (2005)
22. Alkemade, F., Castaldi, C.: Strategies for the diffusion of innovations on social networks. Comput. Econ. **25**(1), 3–23 (2005)
23. Bohlmann, J.D., Calantone, R.J., Zhao, M.: The effects of market network heterogeneity on innovation diffusion: an agent-based modeling approach. J. Prod. Innov. Manage. **27**, 741–760 (2010)
24. Bouzdine-Chameeva, T., Galam, S.: Word-of-mouth versus experts and reputation in the individual dynamics of wine purchasing. Adv. Complex Syst. **14**, 871–885 (2011)
25. Wolfram, S.: Statistical mechanics of cellular automata. Rev. Mod. Phys. **55**, 601–644 (1983)

26. Wolfram, S.: Universality and complexity in cellular automata. Phys. D **10**, 1–35 (1984)
27. Wolfram, S.: A New Kind of Science. Wolfram Media, Inc. (2002)
28. Sarkar, P.: A brief history of cellular automata. ACM Comput. Surv. **32**(1), 80–107 (2000)
29. Zimbres, R.A., Oliveira, P.P.B.: Dynamics of quality perception in a social network: a cellular automaton based model in aesthetics services. Electron. Notes Theor. Comput. Sci. **252**, 157–180 (2009)
30. Becker, G.S.: A note on restaurant pricing and other examples of social influences on price. J. Polit. Econ. **99**, 1109–1116 (1991)
31. Yu, W., Helbing, D.: Game theoretical interactions of moving agents. In: Hoekstra, A., Kroc, J., Sloot, P. (eds.) Simulating Complex Systems by Cellular Automata. Springer-Verlag, Berlin, Heidelberg (2010)
32. Garber, T., Goldenberg, J., Libai, B., Muller, E.: From density to destiny: using spatial dimension of sales data for early prediction of new product success. Mark. Sci. **23**(3), 419–428 (2004)
33. Ma, F., Chao, G.: Research on communication products diffusion in china using cellular automata. Int. Bus. Res. **4**(2), 147–152 (2011)
34. Kowalska-Styczeń, A., Sznajd-Weron, K.: Access to information in word of mouth marketing within a cellular automata model. Adv. Complex Syst. **15**(1250080) 17 pages (2012). doi:10.1142/S0219525912500804
35. Davis, J.P., Bingham, C.B.: Developing theory through simulation methods. Acad. Manage. Rev. **32**(2), 480–499 (2007)
36. Mullen, B., Copper, C.: The relation between group cohesiveness and performance: an integration. Psychol. Bull. **115**(2), 210–227 (1994)
37. Mason, W.A., Conrey, F.R., Smith, E.R.: Situating social influence processes: dynamic, multidirectional flows of influence within social networks. Pers. Soc. Psychol. Rev. **11**(3), 279–300 (2007)

Author Index

A
Augustynek, Krzysztof, 41

B
Banaszak, Zbigniew, 189
Bocewicz, Grzegorz, 189
Brzeski, Adam, 55
Bukietyńska, Agnieszka, 175

C
Cieciński, Piotr, 29
Czekala, Mariusz, 175

D
Dobritoiu, Maria, 15

E
Elefterie, Liana, 15

F
Filcek, Grzegorz, 225

G
Gąbka, Joanna, 165
Giel, Robert, 127
Głąbowski, Mariusz, 203
Grobelny, Jerzy, 3

H
Hołodnik-Janczura, Grażyna, 137

I
Ilias, Nicolae, 15

K
Kecs, Wilhelm, 15
Kempa, Wojciech M., 215
Kierzkowski, Artur, 109
Kisiel, Tomasz, 109
Kowalska-Styczeń, Agnieszka, 237

L
Lubicz, Marek, 91

M
Michalski, Rafał, 3

N
Nielsen, Izabela, 189

P
Pałka, Dariusz, 79
Paprocka, Iwona, 215
Pieniążek, Jacek, 29
Plewa, Marcin, 127

R
Radosiński, Edward, 153

S
Stasiak, Michał Dominik, 203
Sudol, Adam, 65

T
Tudoroiu, Elena-Roxana, 15
Tudoroiu, Nicolae, 15
Turek, Michał, 79

W
Warwas, Kornel, 41

Z
Zabawa, Jacek, 153
Żak, Jacek, 225

J. Świątek et al. (eds.), *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology—ISAT 2016—Part III*, Advances in Intelligent Systems and Computing 523, DOI 10.1007/978-3-319-46589-0