Lecture Notes in Networks and Systems 36

Igor Kabashkin
Irina Yatskiv
Olegas Prentkovskis *Editors*

Reliability and Statistics in Transportation and Communication

Selected Papers from the 17th International Conference on Reliability and Statistics in Transportation and Communication, RelStat'17, 18–21 October, 2017, Riga, Latvia

Springer

Lecture Notes in Networks and Systems

Volume 36

Series editor

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

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

More information about this series at http://www.springer.com/series/15179

Igor Kabashkin · Irina Yatskiv
Olegas Prentkovskis
Editors

Reliability and Statistics in Transportation and Communication

Selected Papers from the 17th International Conference on Reliability and Statistics in Transportation and Communication, RelStat'17, 18–21 October, 2017, Riga, Latvia

Springer

Editors
Igor Kabashkin
Transport and Telecommunication Institute
Riga
Latvia

Olegas Prentkovskis
Vilnius Gediminas Technical University
Vilnius
Lithuania

Irina Yatskiv
Transport and Telecommunication Institute
Riga
Latvia

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-319-74453-7 ISBN 978-3-319-74454-4 (eBook)
https://doi.org/10.1007/978-3-319-74454-4

Library of Congress Control Number: 2017964223

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

In this volume of “Lecture Notes in Networks and Systems,” we are pleased to present the proceedings of the *17th International Multidisciplinary Conference on Reliability and Statistics in Transportation and Communication (RelStat-2017)*, which took place in Riga in Latvia from October 18 to October 21, 2017. This event belongs to a conference series started in 2001 and organized annually by the Transport and Telecommunication Institute (TTI) in Riga, Latvia. The mission of RelStat is to promote a more comprehensive approach supporting new ideas, theories, technologies, systems, tools, applications, as well as work in progress and activities on all theoretical and practical issues arising in transport, information, and communication technologies. Results of previous editions Relstat were published by the “Procedia Engineering” by Elsevier (Relstat 2016) and by the Transport and Telecommunication Institute (TTI) Publishing House (Relstat 2001–2015) in the journal “Transport and Telecommunication” (ISSN 1407-6160).

Design, implementation, operation, and maintenance of contemporary complex systems have brought many new challenges to “classic” reliability theory. We define complex systems as integrated unities of assets: technical, information, organization, economical, software, and human (users, administrators, and management) ones. Their complexity comes not only from their technical and organizational internal structure, which is built upon diverse hardware and software resources, but also from the complexity of information processes (data processing, monitoring, management, etc.) that must be executed in their specific environment. During the operations of such wide-ranging (and often also geographically distributed) systems, their resources are dynamically allocated to ongoing tasks, and the rhythm of system events (incoming and/or ongoing tasks, decisions of a management subsystem, system faults, defensive system reactions and adaptations, etc.) may be considered as deterministic and/or probabilistic stream of events. Security and confidentiality issues enforced by social context of information processing introduce further complications into the modeling and evaluation methods based on statistical and mathematical approach. Diversity of the processes being realized, their concurrency, and their reliance on in-system intelligence often make construction of strict mathematical models impossible and lead to application of intelligent and soft computing methods.

A system approach to the evaluation of the efficiency of complex systems at all phases of their life cycle is the contemporary answer to new challenges in the use of such systems. The dependability approach in theory and engineering of complex systems (not only computer systems and networks) is based on a multi-disciplinary approach to system theory, technology, and maintenance of the systems working in real, very often unfriendly, environment. Usability and dependability concentrate on efficient realization of tasks, services, and jobs by a system considered as a unity of all technical, information, and human assets, in contrast to "classical" reliability, which is more restrained to analysis of technical resources. This difference has caused a natural evolution in the topical range of subsequent RelStat conferences, with an increased focus on dependability approaches over the classical reliability approach. Efficiency of different modes of transport; transport for smart city; reliability, safety, and risk management for transport applications; statistics, modeling, and multi-criteria decision making in transportation and logistics; smart solutions, telematics, intelligent transport systems, innovative economics, and education and training in engineering are the main topics of RelStat.

This year the RelStat conference was supported by the HORIZON2020 funded project ALLIANCE (Enhancing Excellence and Innovation Capacity in Sustainable Transport Interchanges) and led by the TTI, in collaboration with University of Thessaly (Greece) and the Fraunhofer Institute for Factory Operation and Automation (Germany). The special session entitled "Sustainable Transport Interchanges" was organized to allow collaborative research teams from Latvia, Greece, and Germany to present and discuss their findings in the areas of governance and policy development, smart solutions, and decision making.

The Program Committee of the 17th International RelStat Conference, the organizers, and the editors of these proceedings would like to acknowledge the participation of all reviewers who helped to refine contents of this volume and evaluated conference submissions. Our thanks go to all members of Program Committee:

- Prof. Igor Kabashkin, Transport and Telecommunication Institute, Latvia – Chairman
- Prof. Irina Yatskiv (Jackiva), Transport and Telecommunication Institute, Latvia – Co-Chairman
- Prof. Irina Kuzmina-Merlino, Transport and Telecommunication Institute, Latvia – Co-Chairman
- Prof. Lutfihak Alpkan, Gebze Institute of Technology, Turkey

- Prof. Liudmyla Batenko, Kyiv National Economic University named after Vadym Hetman, Ukraine
- Prof. Maurizio Bielli, Institute of System Analysis and Informatics, Italy
- Dr. Brent D. Bowen, Purdue University, USA
- Prof. Inta Bruna, University of Latvia, Latvia
- Dr. Vadim Donchenko, Scientific and Research Institute of Motor Transport, Russia
- Prof. Ernst Frankel, Massachusetts Institute of Technology, USA
- Dr. Ilia B. Frenkel, Industrial Engineering and Management Department, Sami Shamoon College of Engineering, Israel
- Prof. Alexander Grakovski, Transport and Telecommunication Institute, Latvia
- Prof. Stefan Hittmar, University of Zilina, Slovakia
- As. Prof. Ishgaly Ishmuhametov, Transport and Telecommunication Institute, Latvia
- Prof. Dr. Nicos Komninos, Aristotle University of Thessaloniki, Greece
- Prof. Vulfs Kozlinskis, Riga International School of Economics and Business Administration, Latvia
- Dr. Gatis Krumins, Vidzemes Augstskola, University of Applied Sciences, Latvia
- Prof. Zohar Laslo, Sami Shamoon College of Engineering, Israel
- Prof. Agita Livina, Vidzemes Augstskola, University of Applied Sciences, Latvia
- As. Prof. Jacek Mazurkiewicz, Wroclaw University of Technology, Poland
- Prof. Massimo Merlino, University of Bergamo, Italy
- Prof. Boriss Misnevs, Transport and Telecommunication Institute, Latvia
- Prof. Dr. Andres Monzon de Caceres, Universidad Politécnica de Madrid, Spain
- As. Prof. Eftihia Nathanail, University of Thessaly, Greece
- Prof. Andrzej Niewczas, Lublin University of Technology, Poland
- Prof. Lauri Ojala, Turku School of Economics, Finland
- Prof. Ramunas Palšaitis, Vilnius Gediminas Technical University, Lithuania
- Asist. Prof. Dmitry Pavlyuk, Transport and Telecommunication Institute, Latvia
- Prof. Gunnar Prause, Tallinn Technical University, Estonia
- Prof. Olegas Prentkovskis, Vilnius Gediminas Technical University, Lithuania
- Prof. Svetlana Saksonova, University of Latvia, Latvia
- Prof. Natalia Salienko, Bauman Moscow State Technical University, Russia
- As. Prof. Mihails Savrasovs, Transport and Telecommunication Institute, Latvia
- Dr. Ilze Sproge, Transport and Telecommunication Institute, Latvia
- As. Prof. Julia Stukalina, Transport and Telecommunication Institute, Latvia
- Prof. Juri Toluyew, Transport and Telecommunication Institute, Latvia
- Prof. Tatjana Volkova, BA School of Business and Finance, Latvia
- Prof. Edmundas Zavadskas, Vilnius Gediminas Technical University, Lithuania

Thanking all the authors who have chosen RelStat as the publication platform for their research, we would like to express our hope that their papers will help in further developments in design and analysis of complex systems, offering a valuable and timely resource for scientists, researchers, practitioners, and students who work in these areas.

Igor Kabashkin
Irina Yatskiv (Jackiva)
Olegas Prentkovskis

Contents

Smart Solutions for Supply Chain Management

Intelligent Transport Systems

Telematics

Plenary Session

Smart Specialisation Strategies: An Online Platform for Strategy Design and Assessment

Anastasia Panori[1], Nicos Komninos[2(✉)], Christina Kakderi[1], and Katharina Fellnhofer[3,4]

[1] Intelspace Technologies SA, Balkan Center Building C, 57001 Thessaloniki, Greece
[2] URENIO Research, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
komninos@urenio.org
[3] Research and Innovation Management GmbH, Neumarkt an der Ybbs, Austria
[4] Lappeenranta University of Technology, Lappeeranta, Finnland

Abstract. Regions in the European Union (EU) are called to design and implement Research and Innovation Strategies for Smart Specialisation (RIS3), as a prerequisite to receive funding for research and innovation from the European Regional Development Fund (ERDF). To facilitate and streamline this process, the European Commission (EC) has published a Guide to RIS3 and a handbook for implementing Smart Specialisation, providing a set of methodological steps on how to design a RIS3 strategy. Although these publications provide valuable resources to facilitate RIS3 design and implementation, their inputs are focused mostly on the methodological framework, without pointing out any operational directions that could support an undertaking of the proposed methodological tasks in a streamlined and user-friendly way. The Online-S3 project, funded under the Horizon 2020, tries to address this challenge, by developing an online platform for policy advice. This study explores the information links amongst a set of methodologies, across the six phases of RIS3 design process, highlighting underlying relationships in a logical manner, based on the information flows that are detected. The results reveal parts of the overall mechanism for RIS3 policy making processes, providing guidance to regional authorities and encouraging them to use additional methods throughout their RIS3 strategy-design process, that could be managed and delivered through online platforms and applications. This prepares the grounds for future, empirical investigations of this currently under-researched topic, which appears to be crucial for policy-makers.

Keywords: Smart growth · Regional Development · Smart Specialisation Strategy · Platform · Information flow

1 Introduction

Strategy design and implementation is a complex and demanding effort and takes multiple forms depending on the organisation and the context of the initiative. Large companies, non-governmental organizations (NGOs), utility companies, cities and regions,

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_1

governments and international institutions, all these organisations use strategic planning methods and design strategies to succeed in their mission. Strategy design is characterised by uncertainty and ambiguity and requires transdisciplinary knowledge and skills, as there is a plurality of values and opinions to bridge within the organisation, many possible futures, and power games between interest groups internally and externally.

Information systems that support strategy design are becoming mainstream, but also more and more complex [1]. A literature review of information systems for strategic planning reveals a series of factors that influence their success and shortcomings [2]. Apart from pure information systems and dataset feeding the strategy with data, many other IT-based strategy design tools have become available, offline and online. They are used either as e-learning assistants or as step-by-step roadmaps to strategy elaboration. Within this framework, the present paper focuses on Research and Innovation Strategies for Smart Specialisation and on online, web-based, environments that can support the design, implementation, and assessment of such strategies.

Smart Specialisation Strategies (S3) constitute the main growth approach of the EU for the period 2014–2020. These strategies should be formulated by a process of discovery and innovation: as a process of 'choosing races and placing bets' rather than 'picking the winners'. Consequently, strategy interventions should be informed and precise as possible, guided by evidence appropriate to the context, and outcomes that should be monitored and evaluated using quantitative and qualitative metrics and data. The elaboration of Smart Specialisation Strategies is an ex ante conditionality for the ERDF investments of the Thematic Objective 1 (Strengthening research, technological development and innovation), but also it is relevant to the ex ante conditionality of the Thematic Objective 2 (Enhancing access to, and use and quality of information and communication technologies) and Thematic Objective 3 (Enhancing the competitiveness of small and medium-sized enterprises). Ex ante conditionalities are commitments that should be fulfilled to get financial support from the European Structural and Investment (ESI) Funds.

To date, various contributions and preliminary RIS3 evaluation reports have highlighted the difficulties in designing and implementing a RIS3 strategy [3–6]. The initial European Commission's RIS3 planning documents provided some guidance to regional policy makers in the rather complex process of RIS3 design policy [5, 7]. Furthermore, even though entrepreneurs are in better place to identify opportunities, still, the bottom-up approach of the entrepreneurial discovery process (EDP), which is one of the main pillars of the RIS3 strategic planning, requires conscious moderation and careful guidance [5, 8–10]. Both [5] and [11], explain different methodological ways to overcome the theoretical vagueness of the RIS3 guide in selecting priority sectors, while [9] discuss how technological relatedness can provide significant input to the EDP process. Finally, we recently see the development of online tools, through the JRC S3 platform, offering the opportunity to policy-makers to detect any emerging landscape of specialisations more effectively and benchmark regions for improved cross-border learning.

Under this framework, digital platforms have been considered as a key element for enhancing capacity-building for policy-making activities, aiming to upgrade institutional capabilities [12]. Added value, when developing policy-making platforms, can be found on the strengthening of the stakeholder engagement processes, as well as the analytical skills of the users. Both of these issues, are strongly related to a higher degree of RIS3 effectiveness, in terms of better identifying regional assets and features, as well as promoting opportunities for transferring good policy practices between regions.

It has been argued that quality of government, alongside with its contextual and structural characteristics, is related to the quality of outcomes of processes constituting key pillars of the RIS3 approach, such as the EDP process [13, 14]. Moreover, a recent evaluation of a number of implemented RIS3 strategies, highlighted a set of governance-related challenges, including the lack of capability to design and implement regional policies, as well as to actively engage actors in EDP processes [3]. At the same time, regions illustrating a satisfying level of implementation of the RIS3 policies indicate stronger possibilities to reinforce that kind of policy-making processes, through the development of novel toolsets and policy practices.

Given the fact that the original concept of the RIS3 approach has been based upon an accurate and targeted governmental intervention logic to support a number of promising activities [15], the definition of potential areas of intervention should be made through an extended set of methods, including descriptive, benchmarking and discovery exercises. As a result, any existing gap between the EU regions, in terms of high analytical skills and thus, administrative capabilities, could lead to increased levels of inequality regarding the RIS3 effectiveness. Under this context, ICT tools and online platforms target to minimize this gap, in order to reinforce the opportunities, even for less developed regions, to design an evidence-based policy, tailored to their regional specificities.

Under this scope, the Online-S3 Platform (www.onlines3.eu), being developed in the framework of a Horizon 2020 project (ISSI-4-2015), has been designed to address challenges and shortcomings of S3 implementation and assessment. Deploying a connected intelligence approach, the Online-S3 platform uses smart assistants and roadmaps to standardise and automate the tasks of strategy elaboration; give access to databases guiding the strategy formulation by evidence and datasets; and enable participatory design that awakes the potential for collaboration among users and organisations. With all these features, the Online-S3 Platform creates a community of actants (people, organisations, machines) of higher creativity, effectiveness and collective intelligence. It is a web environment that enables a number of stakeholders and users to go through the six steps/phases of strategic planning, proposed by [16, 17], elaborate an informed RIS3 strategy, and monitor its implementation and impact.

The paper tries to shed light on the process of development of the Online-S3 Platform, through investigating the information links that arise between the different applications that are being used. The development of the applications is based on the selection of a number of methods for each phase, that has been performed throughout the early stages of the project [18]. Some of the key questions that are being explored in this study, include the interoperability between the different applications, as well as the existence of any sequential orders that might arise within each phase, or between different phases.

The structure of the paper is the following: Sect. 2 presents a short overview regarding the selected methods, being used as a baseline for the application design, pointing out some of their key features and functionalities. Section 3 provides a more detailed analysis of the three most critical applications within the Online-S3 Platform, analyzing the ways in which a set of different tools are interconnected, through the investigation of information flows during a RIS3 design process. Finally, some discussion and further potentials for the Online-S3 Platform are presented in Sect. 4.

2 Selection of Methods for the Online-S3 Platform

Figure 1 illustrates the key phases penetrating an entire RIS3 policy-design process. As it is shown, the six phases included in this process try to capture a comprehensive set of components that are essential to a RIS3 strategic planning procedure. Starting from Phase 1, governance refers both to government and stakeholder engagement, implying a quadruple helix approach, as the key process of the innovation production. This phase is essential to be placed at the start of a RIS3 strategy design, setting the framework of the entire procedure, as user engagement and participation penetrate the whole policy-making process.

Fig. 1. Phases included in the RIS3 policy-design and implementation process on the Online-S3 Platform. (Source: Authors' elaboration based on [16, 17]).

At the same time, analysis of the context (Phase 2) is a common process for retrieving background information, necessary for any strategic planning process to identify regional specificities and provide information regarding the existing institutional setting to be considered. This phase includes a broad set of methods, targeting to a descriptive, as well as a comparative analysis of a region. Analysis of the regional context targets on pointing out the strengths and weaknesses of a region, when compared to other regions, similar to it. Phase 3, including shared vision and strategy formulation, denotes the strategic and project oriented character of RIS3, highlighting the existence of a bottom-up approach in defining the vision, as well as the priority setting objectives (Phase 4). Policy mix (Phase 5) refers to the definition of the implementation process of the strategy through action plans' design, stressing the need for a structured project-driven approach to RIS3 implementation. Finally, monitoring indicates the need for developing a set of tools for data collection and processing, as key instrument for evaluation of the implemented actions.

It is important to highlight at this point, that the RIS3 policy-design process is not a linear procedure. In many cases, information coming as an output from the implementation of a method, might be used as input to others. Therefore, potential links of information exist between several tools, belonging not only to the same phase, but also to different phases. Thus, it is important to clarify that information flows, as in many cases they overcome the RIS3 phases' sequential logic. Stakeholder engagement, intervention logic, as well as monitoring constitute three characteristic cases of methods, receiving and transmitting information from and to a wide number of other methods.

The selection of the 29 methods, corresponding to the 6 phases of the RIS3 strategic planning process, has been based on a set of methods that have been collected through a mapping exercise, as well as a gap analysis between these methods and a review for good practices [18]. The mapping exercise revealed that regions did not follow the RIS3 steps [16] as a baseline for the design of their methodological approaches, and thus, the robustness of the implemented methods in many cases is questionable, as even the key concepts of the various RIS3 steps were not fully understood. These results also point out that there is no real link between the level of innovativeness of a region and the methodological sophistication of RIS3 design. Hence, it cannot be claimed that moderate and modest innovator regions generally use fewer and less rigorous methods, than leading innovation regions.

At the same time, literature review on good practices has indicated several emerging methodologies, that still have not been used by the regions during their RIS3 design, but could possibly enhance the overall effectiveness of the process. These include foresight exercises and diagnostic tools to identify new activities, possible synergies and complementarities that may arise within the regional context [19]. Furthermore, the use of unstructured data could reveal potential emerging areas of technological and economic activity in a more accurate way [20]. Focusing on strengthening the evidence-based and participatory character of the RIS3 design, policy-makers could also include crowdsourcing priority setting methods and social media analysis for assessing stakeholders' views, through opinion mining and sentiment analysis techniques. Finally, the lack of policy intelligence tools and methods, reflecting the ways in which the monitoring process could be used with a view of a continuous RIS3 update process, was noticeable throughout the literature review on good practices. The use of open data could work on a positive way towards this direction, as it would allow to track progress in terms of objectives and visions, as well as to see how they match with the overall RIS3 approach.

The results, alongside with a short description of the application that has been developed for each method, are presented in Table 1. The description includes the main functionalities of each application, which have been derived based on the key concepts of the corresponding methodology. Information presented in Table 1 provides a baseline, upon which we can further understand the main features of the Online-S3 Platform mechanism, as well as the arising links between the developed applications. A detailed analysis regarding these issues is given in the following section. It should be noted that in all cases, the name of each method corresponds to the name of the developed application. Moreover, the terms method and tool are being used interchangeably in this paper.

Table 1. List of the selected methods/ tools for the Online-S3 Platform. (Source: [18] and authors' elaborations).

Name of the method	Short description
1	2
Phase 1: Governance	
1.1 RIS3 Vision sharing	An application that allows RIS3 managers to create visually attractive infographics that can be used to communicate to a broad audience what RIS3 is about, what are the priority sectors, roadmap and action lines
1.2 Stakeholder engagement	Application providing the opportunity to invite RIS3 stakeholders to use online deliberation functionalities, specifically tailored for entrepreneurial discovery process. Key features of the tool focus on: (1) facilitating discussions; (2) co-creation of procedures including provision of feedback; and (3) a reputation management system
1.3 RIS3 debate at a glance	Application enabling participatory deliberation, in order for policy makers and stakeholders to visualize and share networks of thought, make their reasoning transparent and open to collaborative and iterative reflection
1.4 RIS3 legal and administrative framework related to ESIF	An application providing an overview of ERDF regulations and EU processes of selecting and funding projects in the framework of national/ regional Operational Programmes (OPs)
Phase 2: Analysis of the context	
2.1 Regional assets mapping	Application that draws together information on key regional assets. The objective is to support descriptive analysis of regional assets including a number of key categories
2.2 Research infrastructure mapping	An application for mapping the existing research infrastructures across the EU regions, providing basic background information for regional policy makers in their RIS3 process
2.3 Clusters, incubators, and innovation ecosystem mapping	An application for mapping the innovation ecosystem of a region, in terms of existing clusters, incubators, co-working spaces, start-up support, and the challenges of openness, funding and sustainability of such 'soft' innovation infrastructure
2.4 Benchmarking	An application for comparing the performance of a region with regions that are structurally similar, through providing comparative measures for a series of indicators imported by the user
2.5 Science and technology profile and performance	Application producing 'scientific profiles' for regions, based on Web of Science (WoS), Scopus and Google Scholar data

(*continued*)

Table 1. (*continued*)

Name of the method	Short description
2.6 Specialization indices	The application produces technological and economic specialisation indexes, for understanding the position of regional technological and economic activities in global value chains
2.7 SWOT analysis	An application for completing the SWOT analysis results, including regional strengths, weaknesses, opportunities and threats
Phase 3: Shared vision/Strategy formulation	
3.1 Collaborative vision building	This application capitalises on the outputs obtained in 3.2 and 3.3 and provide tailored online guidelines on the necessary additional phases to arrive at a shared vision for regional smart specialisation strategy
3.2 Scenario building	An application supporting RIS3 scenario building exercises, through the development of baseline scenarios and data projections for scenarios building
3.3 Delphi - Foresight	An application to provide a supportive function to run Delphi-type methods for RIS3
Phase 4: Priority setting	
4.1 EDP focus groups	Application providing a roadmap for the implementation of EDP, including the definition of industry activities and groups, the selection of stakeholders to be involved, the communication of conclusions about opportunities and emerging innovation ecosystems, and the use of EDP conclusions by the regional and national authorities for drafting calls for actions
4.2 Extroversion analysis	Application to detect possible industry segments in which regions present increased extroversion, in terms of exports, attraction of FDI, or other forms of regional openness
4.3 Related variety analysis	Application for calculating the Related/ Unrelated variety entropy indexes, estimating whether specialisation or diversification objectives should be given priority
Phase 5: Policy mix/ Action plan implementation	
5.1 Intervention logic	Application where users can build intervention logic roadmaps, essential to achieve the regional vision and priorities. It is a central application that gathers and comprehensively illustrates information from a number of other tools
5.2 RIS3 action-plan co-design	Application for enhancing collaboration between citizens and policy-makers, throughout the design of an action-plan

(*continued*)

Table 1. (*continued*)

Name of the method	Short description
5.3 RIS3 budgeting	Application for providing a framework for using different budgeting methods to capture the funding dimension of the RIS3 action plan and the needs for funding across the defined implementation period
5.4 RIS3 administrative framework conditions	Application for helping the user/ policy-maker to identify, if their policy instruments included in its RIS3 policy mix/action-plan is eligible for State aid
5.5 RIS3 calls consultation	Application enabling RIS3 stakeholders to assess calls for projects under SF operational programmes that are made by regional authorities
5.6 RIS3 innovation maps	Visualisation tool teasing out information about regional technological trends, using grant data that is collected from S3 programmes and initiatives
5.7 RIS3 actions tracker	Application allowing to track relevant actions from EU research projects available on CORDIS
Phase 6: Monitoring	
6.1 RIS3 monitoring	Application to define the overall process/ roadmap for RIS3 monitoring
6.2 Definition of RIS3 output and result indicators	Application to provide an online guidance on indicator selection and data processing of the results from the implemented actions
6.3 RIS3 beneficiaries and end users' satisfaction online survey	Application for collecting information regarding levels of satisfaction of the RIS3 beneficiaries
6.4 Balanced scorecard	Application drawing together all monitoring indicators together with the results and outcomes achieved to date
6.5 RIS3 social media analysis	Application analysing RIS3 data coming from social media

In many cases, policy-makers and regional authorities do not perfectly understand the rationale behind the design process of a RIS3 strategy [8], not only from a theoretical perspective, but also in terms of data processing and management. This results to the development of fragmented approaches regarding the implementation regional RIS3 strategies, which try to combine many outcomes deriving from several methodologies, characterized by missing links between them. This, of course, is an essential parameter that decreases the overall effectiveness of the decision-making process, as it does not take full advantage of the underlying information hidden in the existing data.

The next section tries to further explore the main functional characteristics of three of the abovementioned set of applications, in order to provide a deeper analysis, which could reveal the rationale behind the selection and use of a concrete set of methods, instead of random selections, based mostly on data availability.

3 The Online S3 Applications: Three Cases

Within the Online-S3 Platform, stakeholder engagement, intervention logic and monitoring are, perhaps, three of the most critical applications, penetrating the whole process of evidence driven policy-design and assessment. In this section, we further explore their internal structure and highlight their role throughout the design of a RIS3 strategy.

3.1 Stakeholder Engagement

The stakeholder engagement tool, together with the RIS3 vision sharing application, provide the users with all the essential information and utilities, related to definition of the stakeholders' groups, as well as engagement and dissemination activities. Given that the RIS3 strategic planning process is largely characterized as a bottom-up approach, in terms of stakeholders' participation, these two tools constitute vital ingredients for establishing an effective link between policy-makers and stakeholders. Previous policy and innovation processes research [21] has emphasized that successful stakeholder participation is characterized by an interplay of quality dimensions which enables to better design public participatory processes of new online service development projects. Figure 2 presents relevant information flows starting from these two applications, towards several other tools, belonging to the same or different phases of the overall RIS3 process. As it can be seen, stakeholder engagement and RIS3 vision sharing apps provide feedback to almost all phases of the design process.

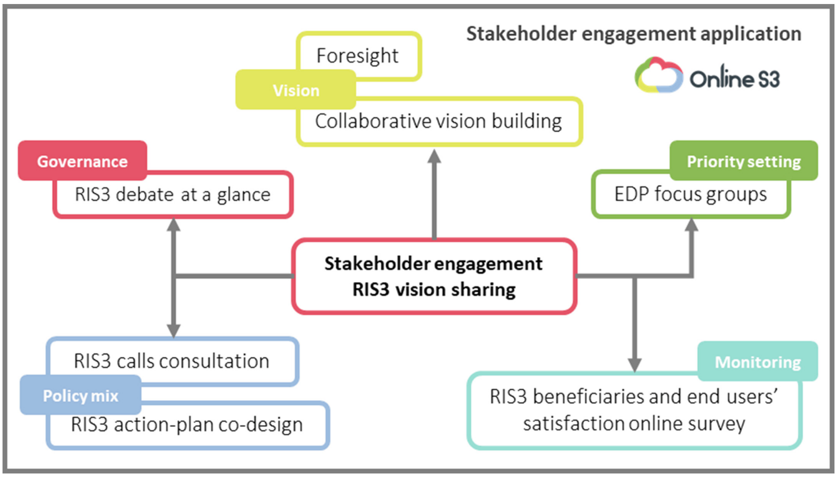


Fig. 2. Main structure of the Stakeholder engagement application. (Source: Authors' elaboration).

Looking more carefully at the type of applications linked to stakeholder engagement, someone can find traditional methods, such as Foresight and EDP focus groups, as well as novel approaches, like collaborative vision building, calls consultation and action-plan co-design. All these strengthen the evidence-based and participatory character of the RIS3 design. At the same time, RIS3 debate at a glance and RIS3 end users' satisfaction survey both target on increasing the overall regional policy intelligence, in terms of crowdsourcing priority setting, through opinion mining and sentiment analysis.

It is essential to notice, that the added value from developing a stakeholder engagement application refers to the cumulative positive feedback that can be raised through public participation. Users are given the opportunity to systematically organize their stakeholder engagement strategy, without being experts in that field, and thus, expand the administrative capabilities of their region.

3.2 Strategy Intervention Logic

The intervention logic application is in the heart of the policy-design process. It is the only application characterized as a solely input application, since it acts as a central information point. It has a dual character. First, it collects information from a number of peripheral phases, in order to depict the overall rationale behind the RIS3 strategy design. Second, it works as a basis upon which policy-makers can detect possible vulnerability points, regarding the implementation of RIS3 actions, based on the feedback that it receives from the monitoring application. Knowing the rationale behind the vision and the implemented policy actions, can be extremely useful to better understand and correct any existing variations between the expected (result indicators) and the actual results (output indicators).

Its main structure is given in Fig. 3. As it is shown, the existing discrete building blocks of the tool guide the user to provide input regarding the regional context, the vision and priorities' setting, the policy mix, as well as the monitoring process. Information referring to the context of a region comes from the outcomes of Phase 1, including descriptive analyses, benchmarking and SWOT analysis. These should then be related to the overall vision and priority setting of the region, which have been defined through EDP, foresight, and extroversion and related variety analyses. The set of result indicators is connected to the selected priorities, as they express the overall vision of the region. In addition to this, the main outcomes of Phase 5, referring to the policy mix, including actions and ways of implementation, are also part of the Intervention logic application. Moreover, the user should relate these with the corresponding output indicators, to get a clear picture of the connection between the policy mix and the monitoring process.

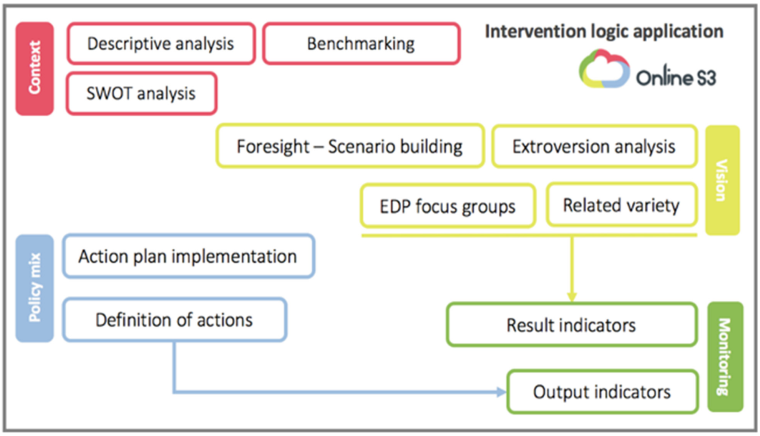


Fig. 3. Main structure of the Intervention logic application. (Source: Authors' elaboration based on [22]).

3.3 Monitoring and Assessment

The monitoring application is characterized as an input/output application, in terms of data interaction. Its main goal is to monitor and evaluate the overall RIS3 implementation process, providing a comprehensive and continuous feedback to decision-makers, regarding the degree to which the actual results are in line with the expected outcomes.

This application works as an umbrella for Phase 6, using as its main input the outcomes of the other four tools included in this phase. Thus, the evaluation process, offered by the Online-S3 Platform, goes one phase further, incorporating a set of different sources and methods for data processing. These consider, not only the output and result indicators, but also the RIS3 beneficiaries' satisfaction and social media analysis outcomes, strengthening in this way the constructive involvement and participation of the stakeholders.

Figure 4 illustrates the abovementioned information flows. Amongst them, the relationship between Monitoring and Intervention logic applications has been previously described. At the same time, RIS3 innovation maps uses monitoring results as input, to visualize available information about regional technological trends funded by RIS3 initiatives, whereas the RIS3 open data tool is a form of a data repository, including information about RIS3 projects linked to specific priorities. Both these applications, are means for communicating the achieved results of the RIS3 strategy, making it comprehensive to the public.

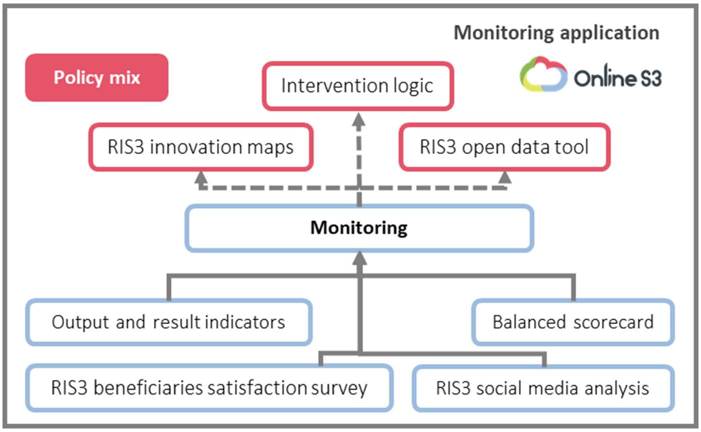


Fig. 4. Main structure of the Monitoring application. (Source: Authors' elaboration).

4 Discussion and Further Potentials for the Online-S3 Platform

Enhancing the analytical skills of regional authorities, as well as the participatory character of the RIS3 policy-design process, should be kept in the spotlight of innovative actions towards a more comprehensive evidence-based strategic planning. Integrated solutions for policy-design, such as the Online-S3 Platform, should be considered as key elements reinforcing this vision.

Although the initial design of the RIS3 approach included a wide set of methodologies to be applied throughout its implementation, the robustness of the applied methods in many cases is questionable, as even the key concepts of the various RIS3 steps were not fully understood by the regional actors. Lack of empirical guidance, as well as low levels of analytical skills of the policy-makers, both resulted in low levels of effectiveness regarding the RIS3 outcomes. Given this, the Online-S3 Platform aims to expand and facilitate administrative capabilities of regional institutions, and thus, become an essential tool for improving the effectiveness and efficiency of decision-making processes.

Taking into consideration the 29 applications that have been developed throughout the Online-S3 project, and are related to the 6 phases of the RIS3 design-process, this study has tried to reveal the existing information flows between a set of crucial tools within the platform. The three examples presented here indicate that there is a complex underlying network of information that links these applications to a number of other platform tools. More specifically, the design of the stakeholder engagement strategy, through the corresponding applications in Phase 2, provides essential feedback to applications based on public participation. Collaborative vision building and EDP are both further strengthened through the use of the vision sharing tool, as it becomes possible to better communicate the aims and targets of RIS3 to the groups of stakeholders involved. Moreover, the development of applications related to opinion mining and public assessment of the RIS3 implemented policies, contribute towards a more effective stakeholder-driven monitoring and evaluation mechanism. As a result, the role of public participation is fostered, not only in priority setting, but also in monitoring the RIS3 process, following its bottom-up principles.

In terms of controlling the overall strategic planning and monitoring RIS3 process, the intervention logic application can be used as a central point for collecting the key outcomes of all phases, included in the Online-S3 Platform. Having a control panel for the strategy, including its rationale for selecting the vision, priorities, policy mix and measurement indicators, can help policy-makers to have an overall picture regarding the expected outcomes of the process. Thus, any deviation of the actual results could be translated and explained through the intervention logic application. Definition of possible corrective actions may can always refer back to this application, in order to update the overall intervention logic or a part of it.

In order to better understand the overall added value of the Online-S3 Platform, it is important to further explore and highlight the existing information flows, as we have done in this paper for the three applications. Given the fact that the abovementioned applications have been developed as standalone apps, further potentials of this platform include the development of an interoperability mode. This could provide a missing link between the 29 applications, facilitating the information flows between them, giving the opportunity to policy-makers to design even more comprehensive RIS3 strategies. Analytical capabilities and public participation should be further enriched and expanded in that case, promoting the effectiveness of the implemented RIS3 strategies.

Acknowledgements. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 710659, ONLINE-S3. We are also indebted to the Finnish regions and their RIS3 responsible persons for their views and time given in the interviews. The article reflects the authors' views only.

References

1. Komninos, N., Tsarchopoulos, P., Kakderi, C.: New services design for smart cities. A planning roadmap for user-driven innovation. In: Proceedings of the 2014 ACM International Workshop on Wireless and Mobile Technologies for Smart Cities, pp. 29–38. ACM (2014)
2. Fellnhofer, K.: Evidence revisited: literature on Smart Specialisation calls for more mixed research designs. Int. J. Knowl.-Based Dev. **8**(3), 229–248 (2017)
3. Capello, R., Kroll, H.: From theory to practice in smart specialization strategy: emerging limits and possible future trajectories. Eur. Plann. Stud. **24**(8), 1393–1406 (2016)
4. Komninos, N., Musyck, B., Reid, A.: Smart Specialisation strategies in south Europe during crisis. Eur. J. Innov. Manage. **17**(4), 448–471 (2014)
5. Iacobucci, D.: Designing and implementing a Smart Specialisation strategy at regional level: some open questions. Sci. Reg. **13**(1), 107–126 (2014)
6. Reid, A., Komninos, N., Sanchez-P, J.A., Tsanakas, P.: RIS3 National Assessment: Greece Smart Specialisation as a means to foster economic renewal. A report to the European Commission, DG Regional Policy, Slovenia (2012)
7. Cooke, P.: Knowledge economy spillovers, proximity, and specialization. In: Interactive Learning for Innovation, pp. 100–111. Palgrave Macmillan, UK (2012)

8. Kroll, H.: Efforts to implement smart specialization in practice-leading unlike horses to the water. Eur. Plann. Stud. **23**(10), 2079–2098 (2015)
9. Boschma, R., Gianelle, C.: Regional branching and smart specialization policy. JRC technical reports, (06/2104) (2013)
10. Fellnhofer, K.: Facilitating entrepreneurial discovery in Smart Specialisation via stakeholder participation within online mechanisms for knowledge-based policy advice. Cogent Bus. Manage. **4**(1), 1296802 (2017)
11. Iacobucci, D., Guzzini, E.: Relatedness and connectivity in technological domains: missing links in S3 design and implementation. Eur. Plann. Stud. **24**(8), 1511–1526 (2016)
12. McCann, P., Ortega-Argilés, R.: The early experience of smart specialization implementation in EU cohesion policy. Eur. Plann. Stud. **24**(8), 1407–1427 (2016)
13. Rodriguez-Pose, A., Wilkie, C.: Institutions and the entrepreneurial discovery porcess for Smart Specialisation. In: Kyriakou, D., Martínez, M.P., Periáñez-Forte, I., Rainoldi, A. (eds.) Governing Smart Specialisation. Routledge, Abingdon (2016)
14. Rodríguez-Pose, A., Garcilazo, E.: Quality of government and the returns of investment: examining the impact of cohesion expenditure in European regions. Reg. Stud. **49**(8), 1274–1290 (2015)
15. Foray, D.: From Smart Specialisation to Smart Specialisation policy. Eur. J. Innov. Manage. **17**(4), 492–507 (2014)
16. Foray, D., Goddard, J., Beldarrain, X.G., et al.: Guide to Research and Innovation Strategies for Smart Specialization (RIS3). Publications Office of the European Union, Luxembourg (2012)
17. Gianelle, C., Kyriakou, D., Cohen, C., Przeor, M. (eds.): Implementing Smart Specialisation: A Handbook. European Commission, EUR 28053 EN, Brussels (2016). https://doi.org/10.2791/53569
18. Griniece, E., Panori, A., Kakderi, C., Komninos, A., Reid, A.: Methodologies for Smart Specialisation Strategies: a view across the EU regions. In: 10th ICEIRD Conference Proceedings, Thessaloniki. (Forthcoming) (2017)
19. Ortega A.R., McCann, P., Perianez-Forte, I., Cervantes, M., Larosse, J., Sanchez, L.: Innovation-driven growth in regions: the role of Smart Specialisation. In: OECD Science, Technology and Industry Policy Papers, vol. 12. OECD/ODCE, Paris (2013)
20. Bakhshi, H., Mateos-Garcia, J.: New data for innovation policy. In: OECD Blue Sky Conference 2016, Ghent, Belgium (2016)
21. Abelson, J., Gauvin, F.-P.: Assessing the Impacts of Public Participation: Concepts, Evidence and Policy Implications. Ottawa: Canadian Policy Research Networks Research Report P/06 (2006)
22. Gianelle, C., Kleibrink, A.: Monitoring mechanisms for Smart Specialisation Strategies, European Commission, JRC-IPTS, Seville. S3 Policy Brief Series no. 13/2015, April 2015 (2015)

A Multistakeholders Multicriteria Decision Support Platform for Assessing Urban Freight Transport Measures

Eftihia Nathanail(✉)

Department of Civil Engineering, University of Thessaly, Pedion Areos, 38334 Volos, Greece
enath@uth.gr

Abstract. Urban road freight transportation significantly affects the quality of life in urban areas. City logistics constitute a major component in the urban economy, and when smartly implemented they may contribute in the city sustainability and livability by alleviating traffic congestion and mitigating emissions and noise impacts. The present research aims at deepening the knowledge and understanding of urban freight transportation (UFT), and enabling guidance provision for selecting and implementing effective and sustainable policies and measures in a city. This is seen from the prism of an integrated evaluation framework, which has been developed for city logistics policies and measures. The framework assesses the complexity of UFT systems, through selected performance indicators, taking into account divergent stakeholder interests, conflicting business models and operations. Evaluation components are formulated in a hierarchical process; sustainability disciplines (economy and energy, environment, transportation and mobility, society), applicability enablers (policy and measure maturity, social acceptance and user uptake), criteria and indicators, capturing the lifecycle impact of the policies and measures. An index is estimated and reflects each evaluation component and stakeholder category; the overall performance of the policy and measure for the city is depicted in the Logistics Sustainability Index.

Keywords: Freight transportation · City logistics · Smart solutions
Multicriteria evaluation · Decision making

1 Introduction

During the last two decades, the booming increase of passenger and freight transport has resulted in deep impact in human and natural environment. Urban areas represent utmost challenges for freight transport, both in terms of goods distribution performance and environmental impacts (air emission, traffic congestion, road safety and accidents and noise and visual nuisance). Urban freight is deemed as high contributor in urban economy but in parallel, urban road freight transport significantly affects the quality of urban environment as it encompasses the processes of transportation, handling and storage of goods, the management of inventory, waste and returns as well as home delivery services and e-commerce [1].

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_2

The contribution of freight movements to urban traffic is 10–18% of urban road traffic and the proportion of air and noise emissions of freight vehicles to total urban traffic reaches almost 40% [2]. Also, urban transport is responsible for about 69% of road accidents that occur in cities revealing the vital dimension of safety in urban freight systems [3].

Policies in global context mostly aim at the alleviation of the most significant bottlenecks that emerge in the urban freight transportation with view to enhance the level of service without penalizing the urban living conditions [4]. According to the EU's Action Plan for Urban Mobility [5], action 19 Urban Freight Transport, some action should be taken with a view to:

- Optimizing urban logistics efficiency;
- Improving the links between long-distance, inter-urban and urban freight transport;
- Incorporating freight transport in local policies and plans; and
- Better managing and monitoring transport flows.

City logistics have been introduced as an efficient concept to address the intricate problems arising from the multidimensional character of urban areas, formulated by environmental considerations, economic growth, new and smart technologies, legal and institutional frameworks, but also by congestion, air pollution, noise, crashes and reduced accessibility due to obsolete infrastructure or environmental and traffic restrictions. The knowledge exchange in this sector is vital in order to capitalize the benefits of urban logistics good practices and assimilate the lessons learnt from the unsuccessful practices around the world, through series of symposia, which will address issues by taking researchers from both sides of the Atlantic to discuss specific topics, achieve intense information sharing, with the opportunity for follow-on collaboration at the project or institutional level [6].

According to the latest Transport White Papers of the European Commission, the achievement of CO_2-free city logistics by 2030 was set as an intermediate objective towards a reduction of 60% in Greenhouse Gas (GHG) emissions [3]. Thus, city logistics constitute a key catalyst in the urban economy but, in parallel, urban road freight transport significantly affects the quality of life in the urban environment. In addition, the optimization of Urban Freight Transport (UFT) can make a significant contribution to the sustainability and livability of cities, alleviating traffic congestion and mitigating emissions and noise impacts.

In Europe, there is an extended reference on the integration of logistics operations in the transportation sector. This regards policy-making and directives developed by European policy-making bodies that delineate the framework of such integration. The European Commission (EC) documenting the Transport White Paper of 2011 [3] and several concrete Action Plans has set the grounds for research towards innovation in the transport sector intending to link research with industry. Among the Commission's concerns in freight transport are insufficient standardization, tackling legal impediments and uneven market actors' capabilities to adopt smart solutions [7].

The existence of diverse stakeholders and their conflicting interests require a safe way of planning and implementation taking into account all stakeholders that are impacted by the solution that is promoted. As such, the investigation of logistics performance should integrate this kind of diversity in order to take forward sustainable logistics solutions. Moreover, the evaluation tools of such solutions should be developed in order to enable better integration and assessment of these interests and concerns of all involved actors.

To this extend, the salient scope of this paper is to demonstrate a decision-making framework which enables knowledge and understanding of freight distribution and service trips by providing guidance for implementing effective and sustainable policies and measures. One of its objectives, is set to develop a common evaluation platform for city logistics measures, which assesses the complexity of UFT systems, through selected performance indicators, divergent stakeholders' interests, conflicting business models and operations. Evaluation components are formulated in a hierarchical process; sustainability disciplines (economy and energy, environment, transportation and mobility, society), applicability enablers (policy and measure maturity, social acceptance and user uptake), criteria and indicators, capturing the lifecycle impact of the policies and measures. An index is estimated and reflects each evaluation component and stakeholder category; the overall performance of the policy and measure for the city is depicted in the Logistics Sustainability Index.

2 A Brief Review

Why Evaluation. Evaluation is a technique that critically examines a process, program or project. It involves collecting and analyzing information about activities, context and results. Its purpose is to enable judgments on effectiveness and efficiency and lead to the improvement of a process, program or project, through facilitating decisions for corrective actions [8].

As finding out about everything is in most cases not feasible, evaluation can help find about the things that really matter. It may reveal accountability, demonstrate impacts (positive and negative), and guide decision-making.

On this direction, the evaluation of UFT measures provides ground to continually improve a project or idea or initiative, both during its implementation (adaptive management), but also after its completion, through the replication of good practice (DOs) and the information provision on the avoidance of mistakes (DONTs). In other words, it helps stakeholders to find out "what works" and "what does not work" – enabling them to answer basic questions about city logistics measures' effectiveness and performance, including:

- Has the UFT measure reached the intended goals? To what degree, how and why?
- Do all the involved stakeholders or part of them benefit from the provided services?
- Are the involved stakeholders or part of them satisfied and to what degree?
- Does the UFT measure constitute a good practice to be replicated by other similar city cases?
- Which are the adaptability and transferability preconditions and which are the potential risks?

In addition, evaluation contributes in the identification of how efficient the project, policy or measure has been in converting resources (funded and in-kind) into activities, objectives and goals. It showcases the effectiveness to the community providing validation through results and findings. Evaluation findings demonstrate to community that the application of a UFT policy or measure is worthwhile. Moreover, sharing findings within the community can serve as a good outreach tool for attracting collaborative partners, recruiting participants and volunteers, while also building trust and organisational resilience.

Evaluation can increase a project's capacity to conduct a critical self-assessment and plan for the future incorporating respective adaptations where and when needed. Conducting an evaluation either internally or with an outside evaluator leads to the assessment of the degree the initial goals and objectives have been met, also investigating how sustainable and meaningful the project has been for the involved or affected stakeholders.

In the end, evaluation can also build knowledge and knowledge sharing, contributing to the evidence-base on what works, in order to avoid mistakes and that successful and effective strategies are replicated, providing information and guidelines to decision makers on how to build on or improve a project.

Evaluation Experiences. There are a lot of projects dealing with city logistics systems and how improvements can be achieved through the implementation of new measures-solutions. The selection and implementation of the appropriate measure(s) as well as the expected impacts of a proposed/implemented measure is subject of the evaluation process. Over the last years, tendencies for adoption of a uniform evaluation method exist, however there is not yet a uniform and robust evaluation method for UFT [9].

The evaluation process and the method that will be selected to assess the performance of a measure is rather a complex issue since it has to incorporate various factors, such as different stakeholder categories often with contradicting interests, difficulties regarding data acquisition, heterogeneity of cities as well as different types of measures, goals etc. For this reason, frequently, measures are assessed by several analyses depending on the objectives that have been set and/or the realized effects. In the past, many projects have evaluated UFT measures.

Within STRAIGHTSOL a general assessment framework for the evaluation of measures has been introduced. The first stage of the framework consisted of four impact areas (economy, environment, society and transport) along with the required data sets for the criteria and indicators. The second stage of the assessment framework focused on the evaluation of alternatives and the current situation and was conducted based on three analyses [10]:

1. A social cost benefit analysis (SCBA) that considered the benefits and costs of measures to society.
2. A Business Model Analysis (BMA) for the financial viability of the operator.
3. A Multi Actor Multi Criteria Analysis (MAMCA) which integrates stakeholders' opinions.

In the context of SMARTFUSION the evaluation methodology was divided into two parts. The first part followed the so called 'Design and Monitoring Framework' (DMF) which address primarily two aspects [10]:

1. A discussion among the stakeholders over the problems that need to be solved and the expected impacts and benefits of the solution;
2. Determination of key indicators to measure the benefits, impacts and success of the solutions.

Given that the first two evaluating settings are addressed, the data collection and data analysis follow. More specifically:

1. Collection of information regarding each company's operation and market's characteristics;
2. Collection of information regarding each city's characteristics;
3. Development of a calculation model to evaluate impact and benefits of the solutions.

SMARTFUSION's evaluation on UFT refers to 'before' and 'after' analysis of every measure-solution in each pilot-city. Collection of data before and after the analysis, considered cost data in respect to logistics and external social cost data.

The scope of CITYLOG was to raise the sustainability and efficiency standards of city logistics through adaptive and integrated solutions. This was achieved by the introduction of a structured evaluation framework for the impacts and interrelationships (based on their measurement to the field test) of:

1. Vehicle design concepts,
2. Structural urban development and
3. Urban transport policies logistic operational models.

The impact assessment considered four impact areas with respective sets of indicators based on data that were derived before, during and after field tests. 'Ex ante – ex post calculation' and 'target and achievement level' were the methods for indicators' analysis [11].

Two methodologies have been used in order to cross-evaluate the measures adopted by the ENCLOSE cities [12]. The first methodology examines the implementation of the measures. For this purpose, six criteria together with more sub-criteria were chosen. Those sub-criteria were weighted and depending on the sensitivity level that was given by partners' survey, a factor was attributed to every measure in order to score it. The second methodology takes into account local assessment results of each city and based on them forecasts expected savings from a measure's implementation. In order this to be achieved, measures were categorized based on their type. The cities that showed interest to adopt such measures were compared with cities that have already applied them. The average of the factors of the cities that have implemented a measure is used to produce the future measure result for a city.

The SMILE project's scope was to develop strategies, plans and measures for city logistics with focus on energy for Mediterranean cities. To evaluate project's output a cost-benefit analysis along with a sensitivity analysis and development of business model took place. The conduct of the analysis was made through the tool City Goods model 2.0 which analyses the demand and the corresponding external effects of UFT through analytic description of city's logistic system, definition of indicators and allowance of planning and evaluation [13].

The BESTUFS project used a set of indicators taken by the CIVITAS-MIRACLES project. Those indicators were used to evaluate different legislative and technological measures, in a Multi Criteria Analysis (MCA). Except for the Multi Criteria Analysis in the context of BESTUFS, a Strength Weakness Opportunities Threats (SWOT) analysis was carried out in order to evaluate more specific strategic and operational measures [14].

The BESTFACT project's scope was to enhance and disseminate best practices concerning UFT evaluation methods [15]. In order this to be achieved, an evaluation tool was developed for impact assessment and transferability evaluation. The sui genesis of this tool was that it examines the appropriateness of a measure for implementation instead of comparing the measures. The impact assessment was carried out based on criteria comprised of strategic targets and topics, while the transferability evaluation was based on four criteria. Furthermore, transferability was assessed separately since the main focus of the project was to 'transfer' the proposed solutions to other interested parties. Finally, a cost-benefit analysis was also performed where applicable.

Three main components form the impact evaluation that was followed in the C-LIEGE project [16]. The first component was a multi-stage impact chain analysis. Aim of this analysis was to pinpoint the pathway from the implementation of a measure up to its realized effects. The second component concerned a comparison with reference cities that have already implemented specific (soft) measures same as the ones that are tested from the city pilots. The last component of the impact evaluation consisted of a scenario-based impact assessment. The two scenarios that were considered for the ex-ante scenario assessment were the zero scenario 'do nothing' and the 'C-Liege scenario'. The latter scenario basically makes an assumption that a specific soft measure is chosen for implementation from the pilot-city and shows indeed a great impact to the city's logistic system.

Due to similarities of FREILOT with preceding Field Operation Tests (FOTs), the project used the FESTA methodology in order to evaluate the impacts of the pilots on the UFT [17]. According to FESTA guidelines four main steps should go through:

1. Function Identification and Description;
2. Use Cases definitions;
3. Identification of Research Questions & Hypotheses;
4. Definition of Performance Indicators & Measurements.

The research questions at step 3 are associated with the effects of a service on selected indicators (e.g. on fuel consumption).

The CITYLAB project follows an evaluation framework, which consists of evaluation methods and indicators. Those evaluation methods and indicators can be distinguished into four fields of evaluation (Adoption, Process, Context and Impacts) that substantially decide whether a measure can be regarded as satisfactory or not and transferable or not. All evaluation activities within CITYLAB are conducted aiming to fulfil two main objectives:

1. Identify cost-effective strategies, measures and tools for emission-free city logistics;
2. Roll out and scale up these strategies, measures and tools.

One of the two innovations of CITYLAB concerns the transferability analysis. Transferability constitutes a primary objective of the whole project and therefore, is perceived as an equal part in the evaluation framework [18].

3 Evaluation Framework

The performance of UFT measures is assessed through an integrated evaluation framework, based on stakeholder objectives, priorities and perceptions, and multi-stakeholder multi-criteria decision-making techniques have been elaborated and implemented, to estimate a Logistics' Sustainability Index (LSI), based on life cycle impacts and societal costs.

Stakeholders in Freight Transportation and Logistics. Supply chain involves a number of stakeholders with different objectives, interests and roles. To demonstrate this, a triangular pyramid is used in Fig. 1. Each stakeholder category is located on each of the four edges of the pyramid. On one of the three edges of the pyramid base, are placed the shippers and receivers, who constitute the initiators of the supply chain. Their needs and requirements are serviced by the operators, who are also placed on one edge of the pyramid base. This stakeholder category includes all supply chain facilitators, mainly the logistics service providers, and the infrastructure and equipment providers. Civilians represent the society, and play a dual role; first, they are the passive recipients of the impacts of any operation being realized within the context of the supply chain;

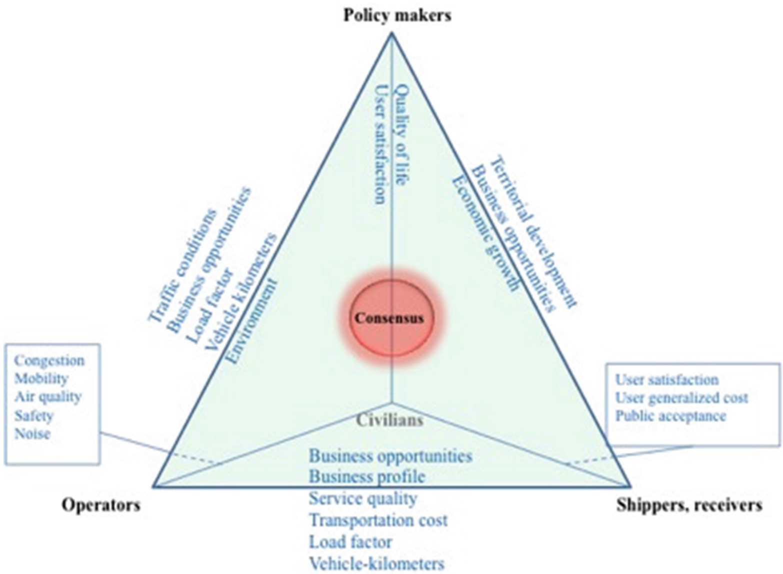


Fig. 1. Freight transportation and logistics stakeholders: interconnections and impacts [19].

second, they are the final users of the goods transported through the supply chain. All these three stakeholder categories interact with one another and share common, different and even contradictory objectives. Figure 1 depicts some major criteria, related to their objectives, set by the interactions between any two stakeholder categories. For example, shippers and receivers, on one hand, and operators, on the other, aim at business opportunities, operational cost reduction and increase of service quality. Shippers and receivers aim at increasing their customers' satisfaction, decrease their cost and gain public acceptance. Civilians need operators to reduce all possible negative impacts freight transportation may bring to their everyday lives.

Policy makers, in sense of local authorities and other public bodies, setting regulations and restrictions in the supply chains and freight transportation, aim at providing better conditions for the business stakeholders and satisfactory life standards for the society. Consensus is required for the smooth coexistence of all the above stakeholders, and to achieve this, all affecting factors and impacts, should be clear, visible and comprehensible by all. An integrated evaluation framework, which contains as components, involved stakeholders, their objectives, criteria and indicators, as well as a comprehensive and sound methodology for enabling quantitative or even qualitative analysis, is important for building consensus. Towards this direction, the research work in this paper demonstrates the formulation of the evaluation methodology used for assessing impacts occurring from the implementation of "smart" solutions in the supply chain, taking into account each and all stakeholders' perspectives.

Assessing logistics processes: Life-Cycle Sustainability approach. City logistics have been introduced as an effective concept for addressing the complicated problems that arise from the multidimensional character of urban areas, formed by environmental concerns, new, smart and innovative technologies, economic considerations and legal frameworks, but also by congestion, air pollution, noise, crashes and limited accessibility due to out-of-date infrastructure or environmental and traffic restrictions [20].

The optimization of UFT can significantly contribute to the increase of the sustainability and livability level of cities, through the limitation of the emissions and noise impacts, the alleviation of traffic congestion and the introduction of new sustainable logistics measures, schemes and incentives.

Life-Cycle Analysis is considered to be a valuable decision-making tool towards sustainability, which takes into account the emerging environmental concerns and is capable of measuring potential environmental impacts, throughout the entire life cycle of a process, system or product, avoiding the crucial errors caused by limited scope work [21]. Initially, the life cycle analysis process was introduced in Europe and the United States of America in the late 1960's, and since then it has been mainly applied to the estimation of energy requirements for the production of chemical products [22], and to the evaluation of the environmental impacts of beverage containers [23]. Through this analysis, the environmental impacts of product's life-cycle are quantified from cradle to grave, divided into five phases: raw material extraction, manufacturing, transportation, use, and end-of-life [24].

The last years, two new terms are being used and have partly replaced the original term of life cycle analysis to focus on different stages of the life cycle analysis process: "Life Cycle Assessment" and "Life Cycle Inventory". The Life Cycle Assessment

(LCA) is a robust tool to assess and, particularly, reduce the potential environmental loads of industrial activities [25]. Three LCA approaches are met in literature, including process-based LCA (P-LCA), input-output based LCA (IO-LCA), and hybrid LCA, which combines P-LCA and IO-LCA [26].

Extending their research focus, a number of studies in addition to environment aspects have included to social and economic concerns life cycle analysis, [27]. The extended quantification of environmental, economic and social impacts has resulted to Life Cycle Sustainability Assessment (LCSA) [28]. The LCSA process has been used in transport for the assessment of vehicles; indicators based on vehicle life cycle impacts that measure individual vehicle features and contribute to sustainability maximization have been used in sustainable transportation assessment of different vehicle types [29]. The LCSA was initially formulated by Klöpffer [30], followed by Finkbeiner et al., [31], and it is foreseen that the assessment of the sustainability performance of a product, process or measure should be implemented by the simultaneous consideration of the three life cycle techniques. The LCSA approach is based on the already standardized technique of the life cycle assessment [32, 33], and can be beneficial for businesses, decision makers and consumers, by contributing to the appropriate organization of complex environmental, economic and social data in a well-structured framework, the clarification of the trade-offs between the sustainability dimensions, life cycle stages and impacts, and the facilitation of further improvements over the life cycle of the product, process or measure [34].

Evaluation in 4-steps. The framework is based on a transparent and consensual decision-making model, expanded though the components of life cycle sustainability assessment, and, eventually, structured as a multi-stakeholder multi-criteria decision-making tool. The structure of the overall evaluation framework is presented in Fig. 2 [20].

It consists of four modules, namely impact assessment, social cost-benefit analysis, adaptability and transferability analysis, and risk analysis, while behavioral modeling is also integrated in order to support the modules in the qualitative data collection (indicators and weights), as well as to enable measuring the potentiality of behavioral change towards the proposed measures on achieving sustainability in cities. The "diamond" reflects the four modules and behavioral modeling, from which life cycle sustainability assessment components originate, and also, addresses the interrelation among these components [20].

Based on the four lifecycle stages (Creation – construction, Operation, Maintenance, and Closure – disposal), according to the International Standards Organization-ISO [32, 33], LCSA acts as the umbrella of the overall framework, realized within four discrete steps, which are taken into consideration in each of the modules and the behavioural modelling. The steps are presented in the following paragraphs.

Step 1: Identification of urban logistics components. Step 1 refers to the identification of urban logistics components, including key influencing factors, measures, logistics scenarios, and urban freight and service trips activities. In fact, in this step, the logistics scene is set and the interaction among key factors, measures, scenarios and activities. Based on previous studies, projects and initiatives focusing on the UFT environment, an initial list of key influencing factors has been determined, organized under

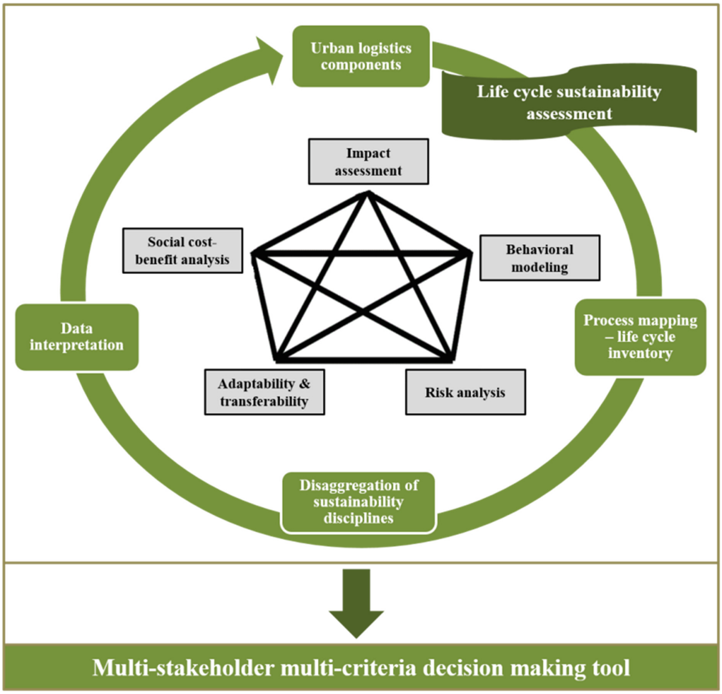


Fig. 2. Structure of the evaluation framework [20].

five main categories: economy and demographics, ecology and social responsibility, logistics' solutions, new technologies, and consumer requirements [35]. Scenarios and activities reflect the UFT measures, which are distinguished into two main categories: cooperative logistics, and administrative and regulatory schemes and incentives [36]. Furthermore, scenarios are also built based on the interested stakeholders, who are grouped in supply chain stakeholders (including supply and demand side), public authorities, and other stakeholders [36].

Step 2: Process mapping-life cycle inventory. In this step, the processes appearing in each of the UFT measures, are described analytically under each of the four stages of the lifecycle sustainability assessment thus: creation-construction, operation, maintenance, and closure-disposal (back logistics).

Step 3: Disaggregation of sustainability disciplines and applicability enablers. This step includes the disaggregation of the impact areas; thus, *sustainability disciplines*: economy and energy, environment, transport and mobility, and society; and *applicability enablers*: policy and measure maturity, social acceptance and users' uptake. In this step, for each sustainability discipline and applicability enabler, the relevant criteria are indicated, and then, for each criterion, respective key performance indicators are defined, which are further described and justified according to the data needed for their calculation or estimation, their units, the stakeholder category they apply, as well as their matching to modules.

Step 4: Data interpretation. This, final, step refers to data interpretation, and in fact relies on the estimation of the Logistics Sustainability Index (LSI). For the clear understanding of the appropriateness, feasibility, viability, efficiency and effectiveness of the tested solutions, and to cope for the conflicting objectives and perceptions of the involved actors, a multi-stakeholder decision making approach has been formulated and implemented, enabled by the evaluation framework. The evaluation methodology incorporates results from travel demand forecasting and traffic and environmental impact assessment models, and produces a set of Key Performance Indicators (KPI) which reflect the effectiveness of the smart solutions.

Evaluation incorporates a multiple weighting scheme, and elimination and ranking techniques and models, for the facilitation of "shared" decision-making, taking into account the participation, viewpoint and contribution of all involved stakeholders to the conformation of the final decision made on the measures. The functions of the evaluation, following the concept of multi-stakeholder multi-criteria assessment methodologies, are depicted in Fig. 3 [20]. Analytical Hierarchy Process has been selected as one of the most practical methods of Multi-Criteria Decision-Making [37]. One of the advantages of Analytical Hierarchy Process (AHP) is that it allows a hierarchical structure of the criteria. This provides better view on objectives, alternatives, criteria and sub-criteria and more efficient allocation of weights. Criteria are assigned weights and the associated indicators are estimated. The output is total scores given for each optional action generating their ranking and prioritization.

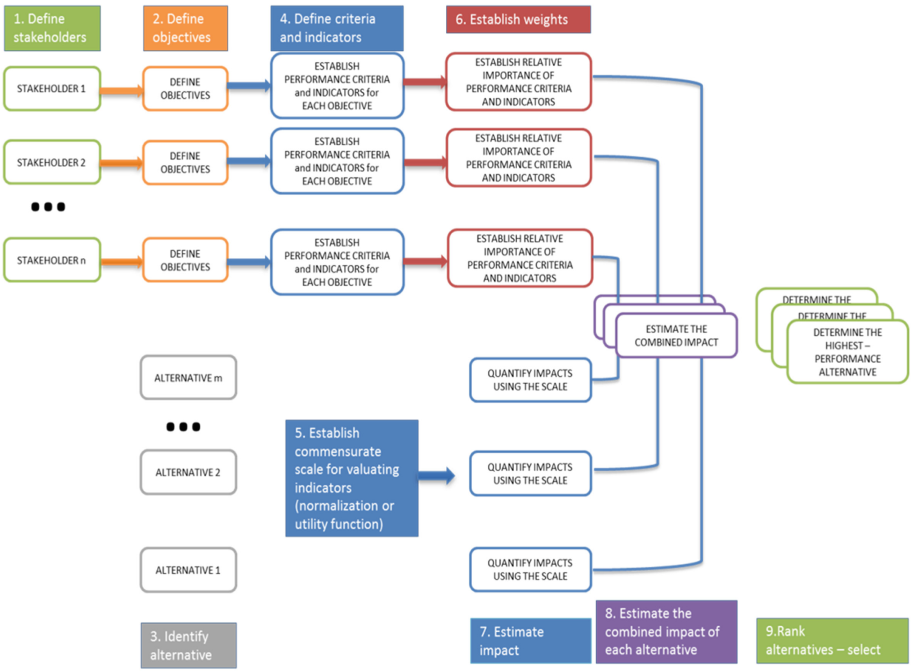


Fig. 3. Structure and function of the Evaluation Tool [20].

AHP has been used in logistics, where an impact assessment analysis was performed to investigate the impact of ICT in a supply chain [10, 38], to solve location problems [39], to evaluate the adoption of e-commerce by SMEs [40, 41], to assess the logistics construction industry [42], to analyze electric mobility in last-mile logistics [43], to evaluate intermodal transportation alternatives for manufacturing firms [44], urban distribution [45], urban construction [46].

There have been many cases of the assessment of sustainability performance of urban transport modes, regarding passenger vehicles [47] and freight vehicles as well [48]. But this analysis has focused mostly on assessing the performance of urban transport vehicles. As the lack of such an analysis has been identified as a gap in the up to now assessment frameworks [49], a life cycle analysis (LCA) has been adopted to assess the wider system by advancing the research in LCA to freight transport and logistics processes rather than vehicles and products.

The first function (function 1) includes the definition of the involved stakeholders, while the determination of specific objectives per stakeholder category is part of function 2. In parallel, alternatives in terms of different scenarios modelling each situation are built (function 3). Each scenario is tested against a number of representative performance criteria and respective key performance indicators, which are established and associated with the stakeholders' objectives (function 4). A commensurate scale is developed for the valuation of the key performance indicators through normalization or utility function (function 5). In parallel, weights per impact area, criterion and key performance indicator are estimated, following specific processes, e.g. analytic hierarchy processes, budget allocation processes, conjoint analyses, etc. (function 6) and in combination with the values of the key performance indicators, the estimation of impacts is feasible (function 7). In function 8, the combined impact of each alternative is estimated in the sense of the Logistics Sustainability Index (sum of weighted performance of all impact areas, criteria, indicators) for each stakeholder category, or for all stakeholder categories, when a weight may be allocated to each one of them (function 9).

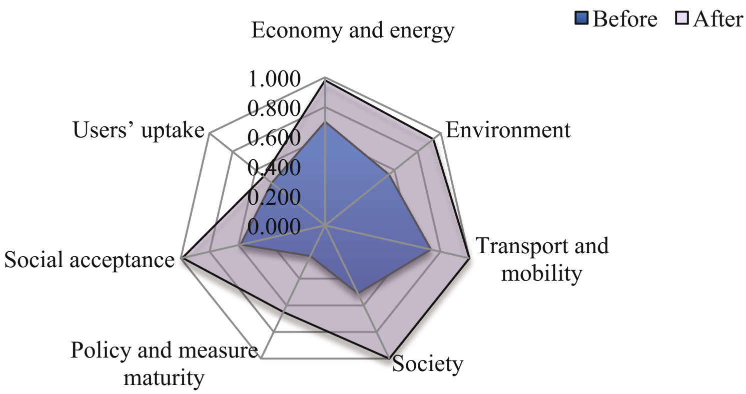


Fig. 4. Assessment results of a before and after analysis.

A regular seven-sided polygon is used to summarize the outputs for all impact areas. Each corner of the seven-sided polygon represents one of the impact areas used and illustrates weighted scores of assessed measures for each impact area, as shown in Fig. 4. The web is an illustrative tool which can be used by decision makers in transportation planning to demonstrate potential tradeoffs between the sustainability impact areas for different choices or transportation policies.

4 Conclusion

Cooperation between actors in supply chain is a key dimension of its operational performance. However, the involvement of different actors within such complicated context sometimes results in conflicting interests driven by the market competition. As such, decision-making is much more challenging. The interaction between stakeholders generates the need to devise ways to promote ‘trading-off’ of actors’ interests aiming at win-win strategies without sacrificing the attainment of general objectives such as high level of service, environmental concerns, cost efficient operations, etc. However, as the operational environment of supply chains is complex and challenging, it becomes crucial to use the proper instruments for efficient decision-making and evaluation of policies.

In this paper, an innovative framework for assessing performance of UFT measures has been developed based on life-cycle sustainability and multi-criteria multi-stakeholder principles. The result is the generation of a tool that offers a balanced comparison between indicators of different units, providing clear understanding of the attitude and impacts of UFT measures. The use of multi-stakeholder multi-criteria assessment combined with life-cycle sustainability analysis offers the opportunity of an integrated tool, clarifying the pros and cons for each party, either directly involved or indirectly, and opens the dialog for mutual decision-making.

Acknowledgements. This work has been supported by the NOVELOG project and has been funded within the European Commission’s H2020 Programme. This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

References

1. European Communities: Urban Freight Transport and Logistics. An overview of the European research and policy (2006)
2. Korver, W., Stemerding, M., Van Egmond, P., Wefering, F.: CIVITAS Guide for The Urban Transport Professionals: Results and lessons of long-term evaluation of the CIVITAS initiative, CIVITAS, Delft (2012)
3. European Commission: “White paper – Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system”, European Commission, Brussels (2012)
4. Campanai, M.: SUGAR: Round Table Best Practices in Urban logistics solutions in Italy: ICT & Organizational Models. First Good Practices Round Table, Athens, Greece, June 2009 (2009)

5. European Commission: Action Plan for Urban Mobility (2009)
6. European Commission: Together towards competitive and resource-efficient urban mobility – A call to action on urban logistics. Brussels, 17 December 2013. SWD, 524 final (2013)
7. Consultation of the Horizon 2020 Advisory Groups, Response of the Transport Advisory Group (2014)
8. Patton, M.Q.: How to Use Qualitative Methods in Evaluation. Utilization-Focused Evaluation. SAGE Publications Inc., Saint Paul (1987)
9. Graindorge, T., Breuil, D.: Evaluation of the urban freight transportation (UFT) projects. In: Transport Research Arena 2014, Paris (2014)
10. Balm, S., Browne, M., Leonardi, J., Quak, H.: Developing an evaluation framework for innovative urban and interurban freight transport solutions. Procedia – Soc. Behav. Sci. **125**, 386–397 (2014)
11. CITYLOG, Final Report D0.4. Sustainability and efficiency of city logistics (2012)
12. ENCLOSE, Deliverable 5.1. Cross-evaluation of energy efficient, sustainable urban logistics measures in the ENCLOSE towns. Evaluation and policy tools (2014)
13. SMILE, SMILE Deliverable 5.3. Post evaluation and definition of actions plans (2015)
14. BESTUFS, Deliverable 5.2. Quantification of Urban Freight Transport Effects II (2008)
15. BESTFACT, Deliverable 3.1. Best Practice Factory for Freight Transport in Europe (2013)
16. C-LIEGE, Final publishable report: Clean Last mile transport and logistics management for smart and efficient local Governments in Europe (2014)
17. FREILOT, Urban Freight Energy Efficiency Pilot – Deliverable 4.1. Evaluation methodology and plan (2011)
18. CITYLAB, Deliverable 5.1. Definition of necessary indicators for evaluation (2015)
19. Nathanail, E., Gogas, M., Papoutsis, K.: Multi-stakeholder assessment of smart solution in urban-interurban freight transportation interfaces. In: 7th International Congress on Transportation Research (2015)
20. Nathanail, E., Adamos, G., Gogas, M.: A novel framework for assessing sustainable urban logistics. In: 14th World Conference on Transport Research, Shanghai, China (2016)
21. Cellura, M., Longo, S., Mistretta, M.: The energy and environmental impacts of Italian households consumptions: an input output approach. Renew. Sustain. Energy Rev. **15**, 3897–3908 (2011) https://doi.org/10.1016/j.rser.2011.07.025
22. Ciambrone, D.F.: Environmental Life Cycle Analysis. Lewis Publishers, New York (1997)
23. Hunt, R.G., Franklin, W.E.: LCA how it came about personal reflections on the origin and the development of LCA in the USA. Int. J. LCA **1**, 4–7 (1996)
24. Rebitzer, G., Ekvall, T., Frischknecht, R., Hunkeler, D., Norris, G., Rydberg, T., Schmidt, W.-P., Suh, S., Weidema, B.P., Pennington, D.W.: Life cycle assessment part 1: framework, goal and scope definition, inventory analysis, and applications. Environ. Int. **30**, 701–720 (2004). https://doi.org/10.1016/j.envint.2003.11.005
25. Finnveden, G., Hauschild, M.Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., Koehler, A., Pennington, D., Suh, S.: Recent developments in life cycle assessment. J. Environ. Manage. **91**(1), 1–21 (2009)
26. Suh, S., Lenzen, M., Treloar, G.J., Hondo, H., Horvath, A., Huppes, G., Jolliet, O., Klann, U., Krewitt, W., Moriguchi, Y.: System boundary selection in life-cycle inventories using hybrid approaches. Environ. Sci. Technol. **38**, 657–664 (2004)
27. Norris, G.: Integrating life cycle cost analysis in LCA. Int. J. LCA **6**(2), 118–120 (2001)
28. Onat, N.C., Kucukvar, M., Tatari, O.: Conventional, hybrid, plug-in hybrid or electric vehicles? State-based comparative carbon and energy footprint analysis in the United States. Appl. Energy **150**, 36–49 (2015)

29. Mitropoulos, L.K., Prevedouros, P.D.: Assessment of sustainability for transportation vehicles. Trans. Res. Rec. J. Trans. Res. Board **2344**, 88–97 (2013)
30. Kloepffer, W.: Life cycle sustainability assessment of products. Int. J. Life Cycle Assess. **13**(2), 89–95 (2008)
31. Finkbeiner, M., Schau, M.S., Lehmann, A., Traverso, M.: Towards life cycle sustainability assessment. Sustain **2**, 3309–3322 (2010)
32. ISO14040, Environmental Management life Cycle Assessment Principles and Framework (2006)
33. ISO14044, Environmental Management life Cycle Assessment Requirements and Guidelines (2006)
34. Valdivia, S., Ugaya, C., Hildenbrand, J., Traverso, M., Mazijn, B., Sonnemann, G.: A UNEP/SETAC approach towards a life cycle sustainability assessment – our contribution to Rio + 20. Int. J. Life Cycle Assess. **18**, 1673–1685 (2013). https://doi.org/10.1007/s11367-012-0529-1
35. NOVELOG, Deliverable D2.3. "Understanding Cities" Tool (2016)
36. NOVELOG, Deliverable D3.1. Integrated Assessment Framework for UFT Solutions (2016)
37. Saaty, T.L.: Theory and Applications of the Analytic Network Process. RWS Publications, Pittsburgh (2005)
38. Nathanail, E.G., Gogas, M.A., Papoutsis, K.N.: Investigation of stakeholders' view towards the introduction of ICT in supply chain using analytic hierarchy process. J. Traffic Logistics Eng. **2**(2), 113–119 (2014)
39. Regmi, M.B., Hanaoka, S.: Application of analytic hierarchy process for location analysis of logistics centers in Laos. In: 91st TRB Annual Meeting, Washington D.C. (2011)
40. Bunno, T., Idota, H., Ogawa, M., Tsuji, M., Miyoshi, H., Nakanishi, M.: Index of the diffusion of information technology among SMES: an AHP approach. In: The Proceedings of the 17th European Regional ITS Conference, Amsterdam, Holland (2006)
41. Yang, L., Zhang, Y.: Application of fuzzy-AHP method in the evaluation of logistics capability in e-commerce environment. In: 2010 International Conference on Communications and Intelligence Information Security (ICCIIS), pp. 228–231 (2010)
42. Cho, K.T.: Multicriteria decision methods: an attempt to evaluate and unify. Math. Comput. Model. **37**, 1099–1119 (2003)
43. Klumpp, M., Witte, C., Zelewski, S.: Information and process requirements for electric mobility in last-mile-logistics. In: Information Technology in Environmental Engineering. Environmental Science and Engineering, pp 201–208 (2014)
44. Aguiar, M.S., Lima, J.P., Lima, R.S.: Analysis of intermodal transportation alternatives using analytic hierarchy process in Brazilian manufacturing. In: 93rd TRB Annual Meeting, Washington D.C., 12–16 January 2014
45. Macharis, C., Milan, L., Verlinde, S.: A stakeholder-based multicriteria evaluation framework for city distribution. Procedia Res. Transp. Bus. Manage. **11**, 75–84 (2014)
46. Macharis, C., Kin, B., Balm, S., Ploos van Amstel, W.: Multi-actor participatory decision-making in urban construction logistics. In: TRB 95th Annual Meeting Compendium of Papers (2016)
47. Mitropoulos, L., Prevedouros, P., Nathanail, E.: Life-Cycle Assessment through a comprehensive sustainability framework: a case study of urban transportation vehicles. In: XXIVth World Road Congress in Mexico City (2011)
48. Nathanail, E., Papoutsis, K.: Towards a sustainable urban freight transport and urban distribution. J. Traffic Logistics Eng. **1**(1), 58–63 (2013)
49. Jeon, C.M., Amekudzi, A.A., Guensler, R.L.: Evaluating plan alternatives for transportation system sustainability: Atlanta metropolitan region. Int. J. Sustain. Transp. **4**, 227–247 (2010)

Transport for Smart City

From Travel Time and Cost Savings to Value of Mobility

Tatiana Kováčiková(✉), Giuseppe Lugano, and Ghadir Pourhashem

ERAdiate Team, University of Žilina, Žilina, Slovakia
{tatiana.kovacikova,giuseppe.lugano,
ghadir.pourhashem}@erachair.uniza.sk

Abstract. Research on Value of Travel Time (VTT) is perhaps the most developed area within the studies on value of time. This knowledge has traditionally been regarded particularly valuable by decision-makers, transport planners, engineers, and economists in the context of projects aiming at enhancing transport infrastructure. As everyone spends much time on the move, engaged in leisure or work activities, travel time represents one of the largest costs of transport.

Current VTT definitions and methodologies for its assessment and subsequent recommendations focus on time and cost savings related to the personal "Travel Time Budget" (TTB), the constant amount of time one invests in daily mobility. Less known is instead what value of travel time means for the end users, in relation to their needs, expectations, and lifestyles. Travel needs and preferences vary, for instance, people do not always consider more meaningful or pleasant the time that is spent more efficiently or productively. One's time valuation fluctuates, also for the same activity performed in different circumstances: time remains a largely subjective entity influenced by endogenous and exogenous factors. As perceived quality of time influences individual well-being, it is important to understand and reflect on own time-use, for instance to adjust habitual behaviour and to consider alternative choices that would better define individual's needs, goals, and expectations.

The objective of this paper is to present how the "Mobility and Time Value" (MoTiV) H2020 project addresses time value following the emerging approach of estimating VTT from the perspective of a single individual with a unique combination of personal characteristics, habits, preferences, and expectations. This approach, in contrast with the classical viewpoint of the economic theories and utility maximization hypothesis, aims at achieving a broader and more interdisciplinary conceptualisation and understanding of VTT emphasising its "behavioural" component.

Keywords: Value of Travel Time (VTT) · Value Proposition of Mobility
Daily travel patterns · Time use

1 Introduction

Time is an intangible and limited resource that everyone possesses. Although one can plan when to use it, time cannot be stopped or stored, and nobody can increase or decrease his/her total stock of time by selling or buying it. As such, it is often felt as a

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_3

scarce and therefore precious resource. Time can be used more efficiently, or just experienced differently.

It is not straightforward to tackle the issue of value of time (VOT), since there is no general definition of this concept. A classic study by Becker [1] regards the study of the allocation of time as an area of consumer economics in which time value is assessed in relation to working and non-working hours, related respectively to people's role as "producers" and "consumers" of commodities. Becker [1] argues that "*the allocation and efficiency of non-working time may now be more important to economic welfare than that of working time […] the cost of a service like the theatre or a good like meat is generally simply said to be equal to their market prices, yet everyone would agree that the theatre and even dining take time, just as schooling does, time that often could have been used productively*". This approach has influenced subsequent research, which has for instance compared use of time to the use of money [5, 7, 8, 15]: indeed, workers' salary is calculated in relation to the number of working hours (i.e. hourly wage). Following this logic, time savings are a key objective of projects addressing value of time because implying higher efficiency and productivity.

The Value of Travel Time (VTT) is one of the most important factors of transport and mobility planning and a number of countries and international organisation have official values that transport project and policies on a consistent basis [19, 10]. On the other hand, research on VTT is perhaps the most developed area within the studies on value of time. The current VTT definitions and methodologies for its assessment and subsequent recommendations focus on time and cost savings related to the personal "Travel Time Budget" (TTB), the constant amount of time one invests in daily mobility. This knowledge has traditionally been regarded particularly valuable by decision-makers, transport planners, engineers, and economists in the context of projects aiming at enhancing transport infrastructure. As everyone spends much time on the move, engaged in leisure or work activities, travel time represents one of the largest costs of transport.

Traditional approaches to VTT estimation regards travel time as unproductive, thus separating "activity time" from "travel time". On the other hand, it has been acknowledged that travel can be an activity in itself [13]. Hence, a more complete VTT estimation should consider activities within mobility, as well as mobility within activities. This view, which goes beyond economic cost considerations, can be linked to the emerging behavioural perspective of "happiness economics" in transport planning. As underlined by Duarte et al. [6], "*existing behavioural travel choice models should be enhanced with regards to their behavioural validity incorporating the impacts of travelling happiness/satisfaction*". It follows that value of travel time as well should be investigated in relation to subjective wellbeing [4]. Although the role and importance of motivational and behavioural factors in VTT research start to be well recognised, these factors do not usually represent the cornerstone of VTT projects. Indeed, further research is needed to describe what value of travel time means for the end users, in relation to their needs, expectations, and lifestyles. Travel needs and preferences vary, for instance, people do not always consider more meaningful or pleasant time that is spent more efficiently or productively. One's time valuation fluctuates, also for the same activity performed in different circumstances: time remains a largely subjective entity influenced by endogenous and exogenous factors. As perceived quality of time influences individual well-being [11], it is important to understand and reflect on own time-use, for instance to adjust habitual behaviour and to consider alternative choices that would better define individual's needs, goals, and expectations. In line with this research need, the project "Mobility and Time Value" (MoTiV[1]) has been recently granted by the European Commission (EC) within the Horizon 2020 (H2020) programme. The aim of this paper is twofold: first, to present the holistic concept of value proposition of mobility, on which to develop a behavioural view of VTT; and secondly, to illustrate the MoTiV conceptual framework and its expected contribution to advance VTT research and applications.

2 VTT and the Value Proposition of Mobility

The "behavioural shift" of studies on VTT calls for the integration of models and frameworks of individual needs, motivations and preferences adapted to the mobility context. In this respect, which values and expectations should be generally fulfilled and therefore addressed by mobility solutions? The conception, development and deployment of mobility infrastructure, services and solutions from the perspective of individual motivations, needs and expectations defines and shapes a Value Proposition of Mobility. This represents a promise of value to be delivered, communicated, and acknowledged to the individual traveller. Group of travellers with similar needs, aspirations, motivations, and expectations are likely to have also a similar general judgment for different transport options. Being a complex ecosystem, there is no single actor in charge of shaping the Value Proposition of Mobility. It is rather a joint outcome of actors co-creating meaning and value to transport and mobility options through policy, implementation, deployment, and participation.

When referring to motivations and needs, a classic reference is Maslow's hierarchy of needs [12], which has been widely used in the transport context as well. An adapted version of this model has been recently used in the context of Mobility as a Service (MaaS) in the attempt to describe its value proposition. For instance, a recent study from UK Catapult describes how MaaS value propositions from different providers should address emerging mobility challenges (see Fig. 1).

It is worth noting that Fig. 1 includes a challenge related to VTT, namely "*enable faster journeys and increase confidence in arrival times*". This challenge is associated to the broader goal of "*enhancing end-to-end journeys by improving mobility choice*", which is not the highest goal of the hierarchy. To a closer look, the complete fulfilment of the Value Proposition of Mobility would require achieving the objective of "*enabling lifestyles by improving mobility fit*". The associated challenges combine both general sustainability aspects and individual well-being.

When considered from the perspective of the hierarchy of travel needs (adapted to MaaS), the economic view of VTT therefore addresses only the mid-level of the Value Proposition of Mobility. Indeed, the value of travel time cannot be always adequately assessed in terms of travel time savings: as shown by Mokhtarian and Salomon [13], in the case of leisure travel under some circumstances people travel just for the sake of

1 www.motivproject.eu.

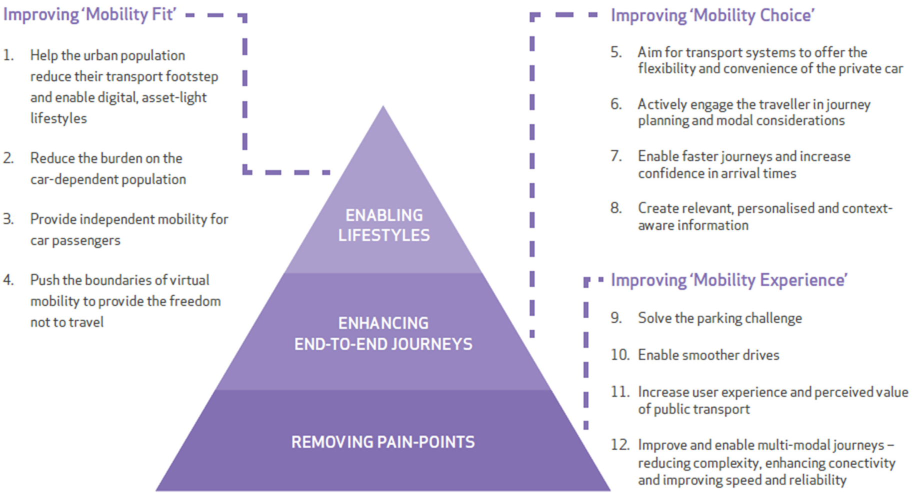


Fig. 1. The traveller needs capability challenges [21].

traveling, because it is "fun". Indeed, it is not the activity to be carried out at destination that represents utility, and therefore value to the traveller, but the journey itself. The authors of the study include several types of activities falling under this category such as driving an off-road vehicle, recreational walking, jogging, cycling, and hiking. These activities are "undirected" in the sense that they do not necessarily have a specific objective or destination point. On the contrary, value of travel time is often associated to utilitarian travel, such as dropping off/picking up children from school, going shopping or to a medical appointment. The study goes further, describing also utilitarian travel situations in which travellers may decide to travel further (therefore, not minimising travel time) because of intrinsic reasons, such as a "variety-seeking" orientation or just curiosity. These are not exceptional situations: a common decision as dining out instead of eating at home (although food is available and quick to cook) could be included under this category.

New concepts of VTT are therefore necessary to acknowledge and fulfil the highest level of the hierarchy of travel needs, dealing with individual lifestyles and well-being. To further understand individual preferences and motivations in travel choices, a classic model by Sheth [17] distinguishes five utility needs corresponding to motivational dimensions:

1. **Functional motives:** related to the technical functions the product performs. The combination of product attributes forms the total functional utility of a product.
2. **Aesthetic-emotional motives:** style, design, luxury, and comfort of a product (class). These motives are not only important for the specific (brand) choice but also for the generic (product) choice. The product class is evaluated in terms of the fundamental values of the consumer in the emotive areas of fear, social concern, respect for quality of life, appreciation of fine arts, religion, and other emotional feelings. Thus, it may be contended that individuals tend to select those product

classes that match with their life styles and enable them to express their fundamental values.

3. **Social motives:** related to the impact that consumption makes on relevant others. Status, prestige, and esteem may be derived from the possession and usage of products and their conspicuous features. Some products are selected for their conspicuousness only ("conversation pieces"), sometimes in combination with aesthetic motives.
4. **Situational motives:** these are not motives in the sense of long-term desires to reach a certain goal. The selection of a product may be triggered by situational determinants such as availability, price discount, and/or accessibility. These situational factors apply usually for a specific brand or type. The brand choice is usually made in these cases without a careful evaluation of the product class.
5. **Curiosity motives:** motives that are supposed to prompt trials of new and/or innovative products. The consumer may try a new product; however, his repeat-purchase may be independent of such trials.

Although Sheth's model was conceived more than forty years ago, it is still current as it acknowledged both intrinsic and extrinsic motivations. A recent study by Mokhtarian et al. [14] underlines that by "*focusing exclusively on the extrinsic motivations to travel runs the risk of substantially underestimating the demand for travel*". In conclusion, the traditional view of VTT allows only addressing a part of the travel needs and motivations. Conventional models such as the ones from Maslow [12] and Sheth [17] are still current and can be applied to establish a framework of assessing and measuring VTT in a way that it covers the whole Value Proposition of Mobility.

3 MoTiV Conceptual Framework

MoTiV is a 30-month research project that started in November 2017. Its primary goal is to contribute to advance research on Value of Travel Time (VTT) by introducing and validating a conceptual framework for the estimation of VTT at an individual level based on the value proposition of mobility. Its approach aims at achieving a broader and more interdisciplinary conceptualisation and understanding of VTT emphasising its behavioral component. The choice of the project acronym has been chosen in line with this perspective, as motive refers to "*something that causes a person to act in a certain way*".

This project will introduce an enlarged conceptual framework for the estimation of VTT based on the idea that each transport mode, or combination of transport modes, provides a different value proposition to the traveller in a specific mobility situation. Time and cost savings represent only one of these factors, not necessarily the ones contributing the most to VTT. Depending on the situation, other factors such as increased comfort or well-being may influence traveller's choice more than time and cost, therefore considered more valuable.

Building on the observations made by Lyons [9] on the "orthodoxy of travel time valuation", the project focus will be on the value of travel time itself (as perceived, experienced and reported almost in real-time via a smartphone app) rather than on the value of travel time saved (as estimated based on assumptions and rather limited survey data). In other words, quoting two key recommendations by Lyons [9], "*investing in schemes to save travel time should be weighed against investing in schemes to make sure travel time is well spent*" and "*trend data are needed to better understand and monitor travel time use phenomena*".

Specifically, MoTiV will go beyond the assessment of time and cost savings and will deliver a multi-dimensional framework for VTT estimation. In addition to cost and time, many other indicators play significant role in decisions on travel and mobility choices, thus representing value for the traveller. Accordingly, a new definition for "Value of Travel Time" will be introduced to acknowledge the dimension of individual "well-being" that incorporates these other indicators relevant to the individual traveller such as calories burnt, carbon footprint and overall satisfaction for the use of travel time.

The perceived value proposition of a certain travel option may not match the actual value delivered to the traveller. When the actual experience has a lower value than the perceived one, this could affect future mobility choices toward the use of other transport modes in similar situations or trip chaining. Knowledge on barriers and factors playing a role in the traveller's choice is therefore key to align expectations and actual experience.

The project will build on latest methodological approaches for collecting travel behaviour patterns via smartphone applications. The use of smartphones for collecting individuals travel behaviour and activity participation over a rather extended period and from a large number of subjects, allows in-depth behavioural analysis that was not possible with traditional survey methods such as paper based travel diaries or telephone surveys [2, 3, 18, 20]. One may argue that despite their limitations, these latter methods were more "inclusive". Although users without smartphone could be left out from a smartphone-based data collection, the penetration of this technology is constantly increasing, especially in Europe (e.g. 70% of the population already owns a smartphone in Spain, and 60% in Italy) that can be considered widely accessible [16]. Nevertheless, particular care will be devoted during local data collection campaigns to reach a homogenous sample that involve all relevant categories of users (including, for instance, senior citizens).

4 Expected Impact

Based on the aforementioned objectives, the expected impact of the MoTiV project can be outlined as follows:

1. *To broaden the definition and assessment of VTT beyond time savings consideration, based on a multidimensional time "value proposition" for the user.*

This will be achieved by introducing a conceptual framework based on a multi-dimensional "value proposition" that can be associated to a travel option in a specific mobility context. Travel option goes beyond the idea of "travel mode", since it may involve a combination of travel modes and because the same mode may feature different value propositions depending on the details how it is provided (e.g. a train with/without wifi) and characteristics of the mobility context. The value proposition of a travel option plays a crucial role in the traveller's choice, which is not only based on the purpose of the travel (e.g. commuting, accompanying children to school, get merchandise, or leisure travelling) and the time needed. Knowledge on the role that these latent factors have (e.g. curiosity, comfort, safety & security, cost) will be valuable to assess travellers' perceived value of time in that context.

A robust conceptual framework for the definition and estimation of VTT will build upon the emerging shift from a purely "economic" view of VTT to a broader and more interdisciplinary conceptualisation of VTT emphasising its "behavioural" component. In this view, time savings are not necessarily the main objective of VTT projects, especially when these are focused on individual perception/use of travel time aiming at maximising individual happiness/well-being. Accordingly, the conceptual framework includes a broadened definition of VTT, a comprehensive description of the Value Proposition of Mobility, and how these concepts are expected to influence use of personal Travel Time Budget (TTB) in mobility contexts.

Beside, a methodology for the estimation of VTT for people work and non-work activity engagements (e.g. maintenance and leisure) in line with the conceptual framework will be developed. The methodology will link micro and macro levels of analysis, describing the process of identifying similar behavioural patterns and quantifying value propositions of mobility in such patterns.

2. *To gain an understanding of traveller's reasons for his/her travel choices in line with the perceived value proposition of mobility.*

This will be achieved by gathering data, via a smartphone application, on the reason for preferring a travel option rather than another in a specific mobility context. This qualitative information will be analysed in relation to other relevant variables such as location, time of day, working and nonworking days and weather condition to identify behavioural mobility patterns. Additionally, choices will be analysed in relation to previous travel information seek or given to the user. For example, in the scenario of daily commuting to a work place at walking distance, a walk could be more attractive than using public transport on a sunny day. In this case, the value proposition is "well-being", while at rainy weather, public transport is the option due to the value proposition "comfort". The same value proposition "comfort" could explain the choice of using one's own car. An attractive public transport system could compete with private cars because of comfort. Additionally, how do preferences and choices change when additional information is provided first-hand to the user?

3. *To assess to what extent ICT connectivity and transport services/infrastructure affect VTT across leisure and work activities and within cultures and generations.*

Knowledge on these factors will be obtained by gathering data on activities carried out by the traveller while on the move, considering the available supporting infrastructure, including its performance (e.g. Internet connectivity, advanced transport connections, frequency of connections, available seat places, environmental design, availability of shared mobility facilities, exclusive bicycle lanes), and wish-list of possible improvements.

4. *To elaborate specific actions and recommendations for mobility policy makers and solution developers that shape the value propositions of travel time.*

Specific actions and recommendations for short-medium term and longer term will be delivered and include social, economic and environmental considerations on the role of ICT, particularly smartphones and tablets, as well as on transport systems and supporting infrastructure in shaping VTT. A cost-benefit analysis relevant to the European context will be carried out: among others, this is expected to identify areas in which the shift away from the "speed paradigm" has already happening and will become more relevant in the coming years. In these areas, the employment of the MoTiV conceptualisation of VTT would provide useful indications to policy makers for the assessment and development of policies as well as to businesses for delivering new or improved mobility solutions. These indications will be complemented by an assessment of the role and importance of significant factors in VTT, addressing also the question of how relevant actors should modify value propositions to foster sustainable mobility behaviour.

5 Conclusion

Value of travel time is highly variable, including a small portion of travel with very high time values, to a significant portion of travel with little or no cost, since travelers enjoy the experience and would pay nothing to reduce it. The MoTiV project introduces a broadened definition and methodology for estimating VTT, acknowledging the shift away from a purely economic view of VTT and the incorporation of behavioural aspects such as personality, preferences, and expectations in its assessment.

To do so, the MoTiV conceptual framework builds on Sheth [17] model to investigate motivational factors behind systematic transport mode choices. These factors will be analysed thanks to a European-wide mobility and behavioural data collection through a smartphone application during the project. This dataset will allow, among others, comparisons across gender, age, and geographical contexts. The dataset also will incorporate "qualitative" input from travellers (e.g. "purpose of travel") that will be used to derive the general mobility context (e.g. leisure/work), activities carried out within mobility, to what extent ICT and transport services/infrastructure supported (or disrupted) such activities, and overall satisfaction/dissatisfaction.

The challenge of an integrated transport and mobility planning is the understanding of the complexity of the parameters involved and their influence on people's choices when they travel. Therefore, understanding effective factors in people decision-making process about travel and activity participation could help planners, policy makers and authorities to make them more attentive to the consequences of their policies in short, medium and long-term. Several managerial implications will emerge from this project: the results will highlight the importance of taking a holistic approach to travellers experience management in terms of preferences and expectations. Outcomes of this study will allow drawing specific policy and business recommendations to signify the role of ICT, transport systems and infrastructure and influence factors in shaping VTT in European context. The expected outcomes of MoTiV would provide useful indications to policy makers for the assessment and development of policies as well as to businesses for delivering new or improved sustainable mobility solutions in European countries.

References

1. Becker, G.: A theory of the allocation of time. Econ. J. **75**(299), 493–517 (1965)
2. Berger, M., Platzer, M.: Field evaluation of the smartphone-based travel behaviour data collection app “SmartMo”. In: 10th International Conference on Transport Survey Methods, Transportation Research Procedia, vol. 11, pp. 263–279 (2015)
3. Yianni, S.: Mobility as a Service. Exploring the Opportunity for Mobility as a Service in the UK. Catapult Transport Systems (2016)
4. Choi, J.T., Coughlin, J.F., D’Ambrosio, L.: Travel time and subjective well-being. Transp. Res. Rec. J. Transp. Res. Board **2357**, 100–108 (2013)
5. DeSerpa, A.C.: A theory of the economics of time. Econ. J. **81**(324), 828–845 (1971)
6. Duarte, A., Garcia, C., Giannarakis, G., Limao, S., Polydoropoulou, A., Litinas, N.: New approaches in transportation planning: happiness and transport economics. Netnomics **11**, 5–32 (2010)
7. Jacoby, J., Szybillo, G.J., Berning, C.K.: Time and consumer behavior: an interdisciplinary overview. J. Consum. Res. **2**(March), 320–339 (1976)
8. Leclerc, F.H., Schmitt, B.H., Dube, L.: Waiting time and decision making: is time like money? J. Consum. Res. **22**(June), 110–119 (1995)
9. Lyons, G.: Briefcase travelling: time use and value. inaugural learned society lecture of the institution of highways & transportation. In: Inaugural Learned Society Lecture of the Institution of Highways & Transportation, London, UK (2008)
10. Mackie, P.J., Worsley, T., Eliasson, J.: Transport appraisal revisited. Res. Transp. Econ. **47**, 3–18 (2014)
11. Mogilner, C., Norton, M.I.: Time, money, and happiness. Curr. Opin. Psychol. **10**, 12–16 (2016)
12. Maslow, A.H.: A theory of human motivation. Psychol. Rev. **50**(4), 370 (1943)
13. Mokhtarian, P.L., Salomon, I.: How derived is the demand for travel? Some conceptual and measurement considerations. Transp. Res. A **35**, 695–719 (2001)
14. Mokhtarian, P.L., Salomon, I., Singer, M.E.: What moves us? An interdisciplinary exploration of reasons for traveling. Trans. Rev. **35**(3), 250–274 (2015)
15. Okada, E.M., Hoch, S.J.: Spending time versus spending money. J. Consum. Res. **31**(September), 313–323 (2004)
16. Poushtner, J., Stewart, R.: Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center (2016)
17. Sheth, J.N.: A Psychological Model of Travel Mode Selection, Bureau of Economic and Business Research of the University of Illinois, Working Paper #291, Urbana, IL. (1975)
18. Stopher, P., FitzGerald, C., Xu, M.: Assessing the accuracy of the Sydney household travel survey with GPS. Transportation **34**, 723–741 (2007)
19. Wardman, M., Chintakalyala, V.P.L., de Jong, G.: Value of travel time in Europe: review and meta-analysis. Transp. Res. Part A **94**, 93–111 (2016)
20. Wolf, J., Loechl, M., Thompson, M., Arce, C.: Trip rate analysis in GPS-enhanced personal travel surveys. In: Stopher, P., Jones, P. (eds.) Transport Survey Quality and Innovation, pp. 483–498. Elsevier Science, Oxford (2003)
21. Catapult Transport Systems: Mobility as a Service. Exploring the Opportunity for Mobility as a Service in the UK (2016)

Synchronisation of Timetables for Public Bus Lines Using Genetic Algorithms and Computer Simulations

Vitalii Naumov(✉)

Cracow University of Technology, Str. Warszawska 24, 31155 Kraków, Poland
vnaumov@pk.edu.pl

Abstract. In this paper, we propose a model of the bus lines synchronisation based on simulation of the public transport system with a genetic algorithm as a tool to obtain some rational solution. The proposed approach considers stochastic nature of the public transport technological processes and provides in a short time a solution close to optimal. The total waiting time for passengers at all the nodes of a public transport network is used as the objective function in the synchronisation problem. Synchronisation is implemented due to time shifts at the schedules for public transport line; these time shifts are represented as chromosomes of a genetic algorithm. In order to evaluate the objective function, simulations of a public transport network were provided. The developed mathematical model is implemented in Python within the frame of a class library for modelling of public transport processes. A case of a public transport system of Bochnia city is applied to illustrate the procedure of synchronisation on the grounds of the developed model.

Keywords: Public transport · Timetables synchronisation · Genetic algorithms

1 Introduction

Synchronisation of timetables for public transport is one of the main direction to enhance a quality of transportation services. Synchronised schedules of public transport lines could significantly reduce time spent by passengers waiting at transfer nodes for the next trip. Besides the increased quality of services, it leads to improvement of the vehicle's performance.

A complexity of the synchronisation problem is conditioned by random parameters of demand for public transport services and stochastic nature of the transportation technological process. These features of a public transport system couldn't be considered properly in the bounds of an analytical model due to a big dimension of the system. Thus, the only practically reachable way to obtain correct estimates of a public transport system efficiency for a set of input parameters (such as numeric characteristics of schedules) is to provide computer simulations of the system.

The goal of this paper is to present an approach to synchronise schedules of public transport lines based on genetic algorithm and computer simulations of the transportation process.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_4

The paper has the following structure: the most relevant approaches to synchronisation of public transport are discussed in Sect. 2; in Sect. 3, the mathematical model of the public transport network is briefly described and the synchronisation problem is formulated; Sect. 4 depicts the genetic algorithm proposed for solving the stated problem; Sect. 5 introduces a case study of the schedule synchronisation for the public transport system of Bochnia city; the last part offers brief conclusions.

2 Literature Review

The main objectives of transfer optimisation usually are minimisation of the passengers waiting time due to synchronisation of public transport vehicles at the transport system interchanges. The existing literature considers the trade-off between the passenger waiting time and operating costs as well [1, 2]. There also exist multi-objective optimisation approaches, such as a model formulated for the multi-objective re-synchronizing of bus timetable problem [3].

According to the commonly used approach, the problem of the timetables' synchronisation is being presented as an integer programming optimisation problem [4, 5]. Recently, in such a formulation, this problem was solved with respect to fluctuating passenger demand and different capacity of vehicles in order to minimise both the expected total passenger waiting time and the observed passenger load discrepancy [6], in order to maximise passenger transfers and minimise bus bunching along the network [7], and with objective to achieve a maximal synchronisation amongst the buses and metro [8]. Due to the complexity of the synchronisation problem related to its big dimension, standard methods of an integer programming optimisation not always could be used for solving it in the real-world conditions. For this reason, different technics, such as Tabu search method, simulated annealing, iterated local search, branch-and-bound method, local search algorithms are applied to obtain some rational solution. Also, artificial intelligence methods are being applied in order to obtain an acceptable heuristic solution, such as the ant colony model in combination with fuzzy logic methods [9] or genetic-based algorithms [2, 5, 10].

In order to develop realistic and adequate models, it is necessary to take into account randomness of different public transport parameters. Stochastic disturbances appear due to the variation of traffic intensity over time, traffic jams, weather condition, driver's behaviour, etc. Moreover, any demand change results in the dwell time deviation. The mentioned uncertainties lead to increasing the variability of the travel time and diminishing service reliability. Some published timetable synchronisation models consider uncertainties in the travel times. Focusing on transfer optimisation, authors of the paper [11] propose to combine a simulation procedure and an optimisation model based on the relaxation of the quadratic assignment problem; the developed model considers travel time as a stochastic variable.

At the planning stage, the travel time variability is commonly compensated by adding a slack time onto a schedule [12], while a holding strategy is used in case of dynamic synchronisation. However, the main drawback of the first strategy is that it leads to increase the in-vehicle time for passengers. The other strategy needs precise real-time data of vehicles arrivals at a transfer node. The research [13] is aimed at optimisation of slack time that is added to the schedules in order to avoid miss connections. The paper [14] also discusses a timed transfer concept and attempt to optimise the headways and slack times by using a heuristic algorithm. Authors of the publication [15] use real-time control actions to reduce bus bunching and maximise transfer synchronisation: the vehicle travel time and passengers arrive moments are considered to be random variables in the proposed simulation model. The paper [10] addresses the stochastic travel time at the stages of planning and operation.

Although the existing approaches allow researchers to obtain a rational solution of the synchronisation problem for a bunch of cases, a lack of versatility and scalability in the known methods should be mentioned. In the next sections, our attempts to develop a general but scalable mathematical model of a public transport system and the respective genetic-based approach to solve the timetables synchronisation problem will be described.

3 Mathematical Model of the Timetables Synchronisation Problem for a Public Transport System

As far as synchronisation procedures are being determined in the frame of a public transport network containing public transport lines, the mathematical model aiming an implementation of these procedures should be defined in the bounds of the model of a public transport system. Here we present a basic description of such the model developed in order to solve public transport optimisation problems on the grounds of computer simulations.

3.1 Model of a Public Transport Network

At the higher level, a public transport network $\mathbf{\Omega}$ could be presented as the set containing a set $\mathbf{\Lambda}$ of lines operating within the bounds of the system and a set $\mathbf{D}$ of passengers using the public transport system to satisfy their needs in trips.

$$\mathbf{\Omega} = \{\mathbf{\Lambda},\ \mathbf{D}\}. \tag{1}$$

As elements of the i-th public transport line λ_i, $\lambda_i \in \mathbf{\Lambda}$, the following objects should be mentioned: a set $\mathbf{L}_i$ of the route segments from which the i-th line is composed, a set $\mathbf{V}_i$ of vehicles operating on the i-th line:

$$\lambda_i = \left\{\mathbf{L}_i,\ \mathbf{V}_i\right\},\ i = 1 \ldots N_{\Lambda}, \tag{2}$$

where N_{Λ} – number of lines in the public transport network.

Elements of the set $\mathbf{L}_i$ characterise the begin and end points (respective stops at the beginning and at the end of the network segment) and a weight (a length of the route segment):

$$l_{ij} = \left\{n_{ij},\ m_{ij},\ w_{ij}\right\},\ l_{ij} \in \mathbf{L}_i,\ j = 1 \ldots N_{L(i)}, \tag{3}$$

where l_{ij} – the j-th segment of the i-th line route; n_{ij} and m_{ij} – the beginning and the end stops of the j-th route segment, $n_{ij} \in \mathbf{N}_i$, $m_{ij} \in \mathbf{N}_i$; w_{ij} – weight of the j-th route segment [km]; $N_{L(i)}$ – number of the route segment for the i-th public transport line; $\mathbf{N}_i$ – a set of all bus stops for the i-th line.

A vehicle v_{ij} as an element of the set $\mathbf{V}_i$ ($v_{ij} \in \mathbf{V}_i$, $i = 1 \ldots N_{V(i)}$, where $N_{V(i)}$ is the number of vehicles servicing the i-th line) is first characterised by a capacity and the timetable on the i-th public transport line:

$$v_{ij} = \left\{c_{ij},\ s_{ij}\right\}, \tag{4}$$

where c_{ij} – capacity of the j-th vehicle [pas.]; s_{ij} – timetable of the j-th vehicle.

The timetable item (its k-th position) for the j-th vehicle of the i-th line could be presented as a tuple $s_{ijk} = \left\langle p_{ijk},\ t_{ijk}\right\rangle$, for which the first component p_{ijk} is the position on the schedule (a stop of the public transport line), and the second component t_{ijk} is the moment of arrival of the j-th vehicle at the p_{ijk} stop. Thus, the timetable s_{ij} of the j-th vehicle at the i-th line is a set of tuples:

$$s_{ij} = \bigcup_{k=1}^{N_{s(ij)}} s_{ijk} = \bigcup_{k=1}^{N_{s(ij)}} \left\langle p_{ijk},\ t_{ijk}\right\rangle,\ t_{ijk} < t_{ij(k+1)}, \tag{5}$$

where $N_{s(ij)}$ – number of positions in the sequence of stops, which the j-th vehicle passes operating at the i-th bus line.

The moment of arrival at the first stop in a sequence of all positions in a timetable for the j-th vehicle of the i-th line should be determined in a following way:

$$t_{ij(1)} = t_0 + \psi_i + \Delta t_{ij}, \tag{6}$$

where t_0 – the moment when the transport system starts servicing the clients [min.]; ψ_i – time-shift of the servicing start for the i-th line [min.]; Δt_{ij} – time-shift of the servicing start for the j-th vehicle of the i-th line [min.].

Demand for services of a public transport we propose to present as a set of elements that describe passengers intending to use the bus service. Each element of this set could be described on the grounds of a number of parameters:

$$\pi_i = \left\{\eta_i,\ \mu_i,\ \mathrm{P}_i,\ \tau_i\right\},\ \pi_i \in \mathbf{D},\ i = 1 \ldots N_D, \tag{7}$$

where π_i – the i-th passenger; η_i and μ_i – original and destination stops of the i-th passenger trip, $\eta_i \in \mathbf{N}$, $\mu_i \in \mathbf{N}$; P_i – a set of transfer stops where the i-th passenger changes lines within his trip; τ_i – moment of time when the i-th passenger appears at the bus stop η_i in order to perform a trip [min.]; N_D – the total number of passengers using the public transport system [pas.].

In order to simulate demand for bus services, it is quite convenient to divide all the elements of the set **D** into groups according to the stops of the public transport network where the trips begin:

$$\mathbf{D} = \bigcup_{j=1}^{N_L+1} \mathbf{D}_j, \tag{8}$$

where $\mathbf{D}_j$ – a group of passengers traveling from the j-th stop of the bus line:

$$\mathbf{D}_j = \{\pi_i\colon \eta_i = j\}. \tag{9}$$

The time interval ξ_j between the moments of passengers' arrivals at the j-th stop is a random variable and in that case, it is considered in the model as a characteristic of a bus stop. Then for each group $\mathbf{D}_j$, the parameters τ_i of the certain elements of the set could be defined on the grounds of realisation of the random variable ξ_j of intervals between appearances of passengers at the j-th stop:

$$\tau_i = \begin{cases} \tilde{\xi}_j, \ i = 1, \ \eta_i = j, \\ \tau_{i-1} + \tilde{\xi}_j, \ i > 1, \ \eta_i = j, \end{cases} \tag{10}$$

where $\tilde{\xi}_j$ – realisation of the random variable of an interval between the passengers' appearances at the j-th stop [min.].

The software implementation of the described model is presented in the paper [16] and could be found in the repository available at [17].

3.2 Formulation of the Timetables Synchronisation Problem

As input variables for the timetables synchronisation problem, a set of parameters could be used in the frame of the described model: the number of vehicles on the public transport lines, the vehicles capacity, and the schedule numeric parameters – time-shifts ψ_i and Δt_{ij}.

In the simplest case, characterised by minimal changes of existing public transport system, synchronisation of timetables could be accomplished by setting values of the servicing start time-shifts for the public transport lines. These shifts could be presented as a vector Ψ of integer numbers:

$$\Psi = \langle \psi_1, \psi_2, \ \ldots, \psi_{N_\Lambda} \rangle. \tag{11}$$

The basic technological parameter reflecting the result of synchronisation is total time spent by the passengers waiting for vehicles at bus stops. The respective objective of synchronisation would be the minimisation of total waiting time:

$$T_w(\Psi) = \sum_{i=1}^{N_D} t_{wi} \rightarrow \min, \tag{12}$$

where t_{wi} – waiting time of the i-th passenger of a public transport system [min.].

It should be mentioned, that other goal functions could be used in the described synchronisation problem, e.g. maximisation of the transport company's profit, minimisation of total operation and waiting costs, maximisation of the servicing level, etc.

4 Genetic Algorithm for Timetables Synchronisation

As far as the synchronisation problem in the presented form cannot be solved analytically, the respective heuristics should be proposed. The use of genetic algorithms is a convenient way to obtain solutions for such complex stochastic models as the one presented in this paper. Main features of a genetic algorithm application are the used approach to the chromosomes representation and the procedure of the fitness function evaluation.

4.1 Chromosomes Representation

A chromosome as an element of a genetic algorithm presents a set of independent variables. In the case of the formulated problem, the vector Ψ representing an input variable should be coded.

If a chromosome is presented in a binary form, the respective data should be defined as a sequence of 0 and 1 (the chromosome genes). Such a sequence could be easily obtained if the vector Ψ would be transformed in accordance with the principle shown in Fig. 1: the value of each element in the time-shifts vector is coded in a binary form for the given number of genes per value (at the presented example each vector element is coded with 5 genes), then all the coded values are merged into a sequence representing the vector Ψ. The high bound of the possible time-shift value is determined by the number of genes per the element representation: 3 genes corresponds to the range of the time-shift value between 0 and 7 min inclusively, 4 genes allow to consider the range between 0 and 15 min, 5 genes could code the range between 0 and 31 min, etc.

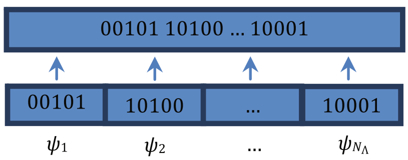


Fig. 1. Coding a chromosome for the timetables synchronisation problem.

The chromosome shown in Fig. 1 contains the value of 15 min for the time-shift at the first line of the public transport system, 20 min for the second line and 17 min for the last element in a list of public transport lines.

4.2 Evaluation of the Fitness Function

The fitness function value should be evaluated on the grounds of computer simulations: within the frame of the presented model, the total waiting time for all the passengers of public transport lines would be a random variable. The realisation of this variable could be obtained as the result of the simulation model single run.

Using the developed software implementation [17] of the described mathematical model, the simulation model of the public transport system should be built in the following way:

- the network configuration of the public transport system should be provided on the grounds of the *Net* class as the graph model: bus stops (the graph vertices) have to be defined as entities of the *Node* class and segments of the routes (the graph edges) should be set as the class *Link* entities;
- the public transport line should be defined on the basis of the developed graph as entities of the *Line* class;
- models of the buses serving the public transport lines should be defined on the grounds of the *Vehicle* class as the fields of the respective entities representing the transport lines;
- numeric parameters of demand for the public transport services should be described as random variables with the use of the *Stochastic* class as the respective fields of the bus stops' models.

Having the implemented model of the public transport system and the generated demand for transport services (with the use of the method *gen_demand* of the *Net* class), the fitness function (total waiting time) is being evaluated as the first element of the tuple returned by the *simulate* method of the *Net* class.

5 Case Study: Timetables Synchronisation for the Public Transport System of Bochnia City

The presented approach for synchronisation of the public transport lines was used to minimise the passenger's waiting time in Bochnia (Lesser-Poland Voivodeship, Poland). This case was chosen in order to illustrate the calculative abilities of the model and the respective software on the example which doesn't require many resources: Bochnia is a town of about 30 thousand inhabitants with 4 public transport lines. The total length of all the routes of Bochnia is 42.3 km. There're 43 bus stops at the public transport lines of the city. The fleet of the Bochnia's public transport system accounts 8 vehicles with a capacity of about 45 passengers. More detailed characteristics of the city public transport lines are presented in Table 1.

For schedule synchronisations of the Bochnia's public transport lines, the genetic algorithm with the following parameters was implemented on the basis of the *GA* class of the developed library [17]:

- number of generations: 30 (number of the algorithm iterations);
- the population size: 50 (number of entities in a pool – a set of the alternative values of the time-shifts vector);
- the chromosome size: 20 bits (5 bits per a line: the time-shift values are considered in the range between 0 and 31 min inclusively);
- crossover probability: 0.5 (probability that genetic code of a new entity would be replicated from the code of both parents);
- number of mutation turns per a chromosome: 3 (number of genes in a chromosome being chosen for mutation);
- mutation probability: 0,1 (probability that the chosen gene will be inverted);
- survivors rate: 0.2 (20% of species in a generation with the best characteristics are chosen for further reproduction).

Table 1. Characteristics of public transport lines in Bochnia.

Bus line ID	Route length [km]	Number of stops	Number of vehicles
#1	8.1	16	2
#3	9.9	22	2
#5a	6.6	16	2
#9	17.7	38	2

In the result of calculations for 100 runs of the described genetic algorithm, a set of alternative solutions was obtained (Fig. 2).

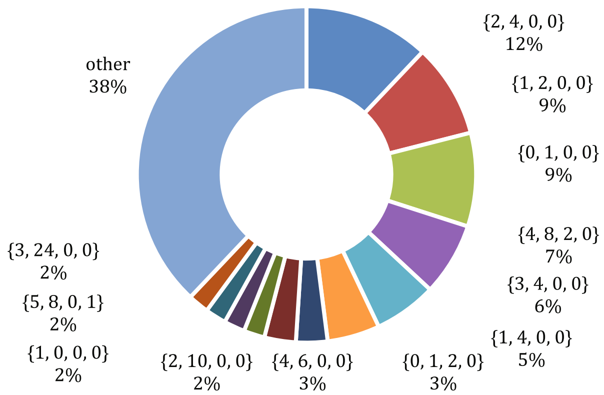


Fig. 2. Values of the time-shifts vector obtained as a result of calculations.

As it could be concluded from the data presented in Fig. 2, the most frequent solution determined by the developed algorithm is $\Psi = \langle 2,\ 4,\ 0,\ 0\rangle$, i.e. the schedule of the bus lines #1 and #3 should be shifted on 2 min. and 4 min. respectively. However, as far as the genetic algorithm is a heuristic procedure, this solution does not guarantee a minimum of the total waiting time.

Using the developed model of the Bochnia's public transport system, the system simulations for the repeatedly arising solutions were performed in order to estimate the expected value of the total waiting time. The simulation results obtained on the grounds of 100 model runs per an alternative solution are shown in Table 2.

Table 2. Waiting time estimations for the alternative solutions.

Values of the time-shifts vector Ψ [min.]				Expected value of the total waiting time [min.]
$\psi_{\#1}$	$\psi_{\#3}$	$\psi_{\#5a}$	$\psi_{\#9}$	
0	1	0	0	59 549
0	1	2	0	58 974
1	0	0	0	57 892
1	2	0	0	57 039
1	4	0	0	56 347
2	4	0	0	57 343
2	10	0	0	57 594
3	4	0	0	58 174
3	24	0	0	59 010
4	6	0	0	60 262
4	8	2	0	58 945
5	8	0	1	58 808

The simulation results allow us to conclude that in the set of considered alternatives the best solution would be $\Psi = \langle 1,\ 4,\ 0,\ 0 \rangle$, which is characterised by the minimal expected value of the total waiting time.

6 Conclusions

The public transport system is a complex system characterised by stochastic demand for transport services and random parameters of technological processes. This conditions a need for simulations of the system while solving the problems of its optimisation. The developed general mathematical model of a public transport system is a scalable tool which allows researchers to formulate a range of problems in the area of public transportation including the timetables synchronisation problem. The proposed software implementation of the model includes a number of methods providing simulations of the public transport operation processes.

The developed software in combination with the described genetic algorithm contributes to solving the timetables synchronisation problem. The proposed approach could be expanded by changing the chromosomes representation: e.g. other decision variables could be encoded in the chromosome, such as the number of vehicles, the vehicle's capacity or individual time-shifts for buses in a schedule. However, it should be noted that such changes would affect the calculation speed of the algorithm and could limit the applicability of the proposed approach.

References

1. Bruno, G., Improta, G., Sgalambro, A.: Models for the schedule optimisation problem at a public transit terminal. OR Spectr. **31**(3), 465–481 (2009)
2. Shrivastava, P., Dhingra, S.L.: Development of coordinated schedules using genetic algorithms. J. Transp. Eng. **128**(1), 89–96 (2002)
3. Wu, Y., Yang, H., Tang, J., Yu, Y.: Multi-objective re-synchronizing of bus timetable: model, complexity and solution. Transp. Res. Part C **67**, 149–168 (2016)
4. Ceder, A., Golang, B., Tal, O.: Creating bus timetables with maximal synchronisation. Transp. Res. Part A **35**(10), 913–928 (2001)
5. Wihartiko, F.D., Buono, A., Silalahi, B.P.: Integer programming model for optimizing bus timetable using genetic algorithm. IOP Conf. Ser.: Mater. Sci. Eng. **166**, 012016 (2017)
6. Liu, T., Ceder, A.: Synchronisation of public transport timetabling with multiple vehicle types. Transp. Res. Rec. **2539**, 84–93 (2016)
7. Ibarra-Rojas, O.J., López-Irarragorri, F., Rios-Solis, Y.A.: Multiperiod bus timetabling. Transp. Sci. **50**(3), 805–822 (2016)
8. Shen, Y., Wang, S.: An adaptive differential evolution approach for the maximal synchronisation problem of feeder buses to metro. J. Comput. Theor. Nanosci. **13**(6), 3548–3555 (2016)
9. Teodorović, D., Lučić, P.: Schedule synchronisation in public transit using the fuzzy ant system. Transp. Planning Technol. **28**(1), 47–76 (2005)
10. Wu, W., Liu, R., Jin, W.: Designing robust schedule coordination scheme for transit networks with safety control margins. Transp. Res. Part B **93A**, 495–519 (2016)
11. Bookbinder, J.H., Désilets, A.: Transfer optimisation in a transit network. Transp. Sci. **26**(2), 106–118 (1992)
12. Dou, X., Yan, Y., Guo, X., Gong, X.: Time control point strategy coupled with transfer coordination in bus schedule design. J. Adv. Transp. **50**(7), 1336–1351 (2016)
13. Lee, M., Schonfeld, P.: Optimal slack time for timed transferred at transit terminal. J. Adv. Transp. **25**(3), 281–308 (1991)
14. Ting, C., Schonfeld, P.: Schedule coordination in a multiple hub transit network. J. Urban Planning Dev. **131**(2), 112–124 (2005)
15. Nesheli, M., Ceder, A., Gonzalez, V.: Real-time public-transport operational tactics using synchronized transfers to eliminate vehicle bunching. IEEE Trans. Intell. Transp. Syst. **17**(11), 3220–3229 (2016)
16. Naumov, V., Samchuk, G.: Class library for simulations of passenger transfer nodes as elements of the public transport system. Procedia Eng. **187**, 77–81 (2017)
17. Class library for modeling a public transport network. https://github.com/naumovvs/publictransportnet. Accessed 15 July 2017

Modelling the Location of Charging Infrastructure for Electric Vehicles in Urban Areas

Larisa Grackova[1(✉)], Irina Oleinikova[2], and Gaidis Klavs[2]

[1] Riga Technical University, 11 Krivu Street, Riga 1006, Latvia
larisa.grackova@inbox.lv
[2] Institute of Physical Energetics, 11 Krivu Street, Riga 1006, Latvia

Abstract. This research attempts to develop a model for the optimal location of charging stations and distribution network in urban areas from the point of view of a user of an electric vehicle. The proposed model focuses on the interaction between behaviour of people, urban infrastructure and the power supply system. To evaluate the optimal connection points for charging stations the transportation theory algorithm was used. Having received the final version of the location of the infrastructure at the considered geographical location in the Calculation phase, the low, medium and high penetration demand for charges is estimated. This methodology has been applied to a residential low-voltage system. The main finding of this work is that the number of fast charging stations needed for the urban district is quite low. Also, the results indicate that the increase share of electrical vehicles in the urban road parks leads to a reduction in harmful emissions.

Keywords: Electric vehicle · Charging station · Location assignment problem

1 Introduction

The European Commission adopted the Transport 2050 roadmap (Transport 2050) for a transport system that sets goal to increase cars with cleaner fuels, mainly focusing on the urban transport. Since, first of all, the urban traffic, noise exposure and poor air quality harmful emissions from vehicles are having a negative effect on people's health. The goal of Transport 2050 is to halve the use of vehicles that run on traditional petroleum fuels (petrol or diesel fuel) in urban environments by 2030 and also to reduce carbon emissions and other pollutants [1]. To accelerate this process, appropriate charging infrastructure for widespread deployment of electrical vehicles (*EVs*) are being created in cities around the world. A number of major cities and regions around the world are actively pursuing deployment goals through a variety of innovative policy measures and programs, which are tailored to particular circumstances of each city. For instance, there are financial and non-financial consumer incentives to boost the demand for vehicles and charging infrastructure. Financial incentives include: rebates or tax credits on vehicles, exemptions from vehicle registration taxes or license fees, discounted tolls and

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_5

parking fares, as well as discounts for recharging equipment and installation. Non-financial incentives include: preferential parking spaces, access to restricted highway lanes, expedited permitting and installation of electric vehicle supply equipment (EVSE). Cities are also using the *EVs* as municipal fleet and hybrid buses are being added as public transportation [2]. The density of population and its development of urban power system are well positioned to benefit from *EVs*. The research from different countries to establishing algorithms and models of *EV* infrastructure, which are used as constraints with the attributes being: providing the population with vehicles, the average daily mileage of a vehicle, the distance between the destinations, number of individual trips, opening hours and parking lot availability, technical indicators of the distribution network and etc.

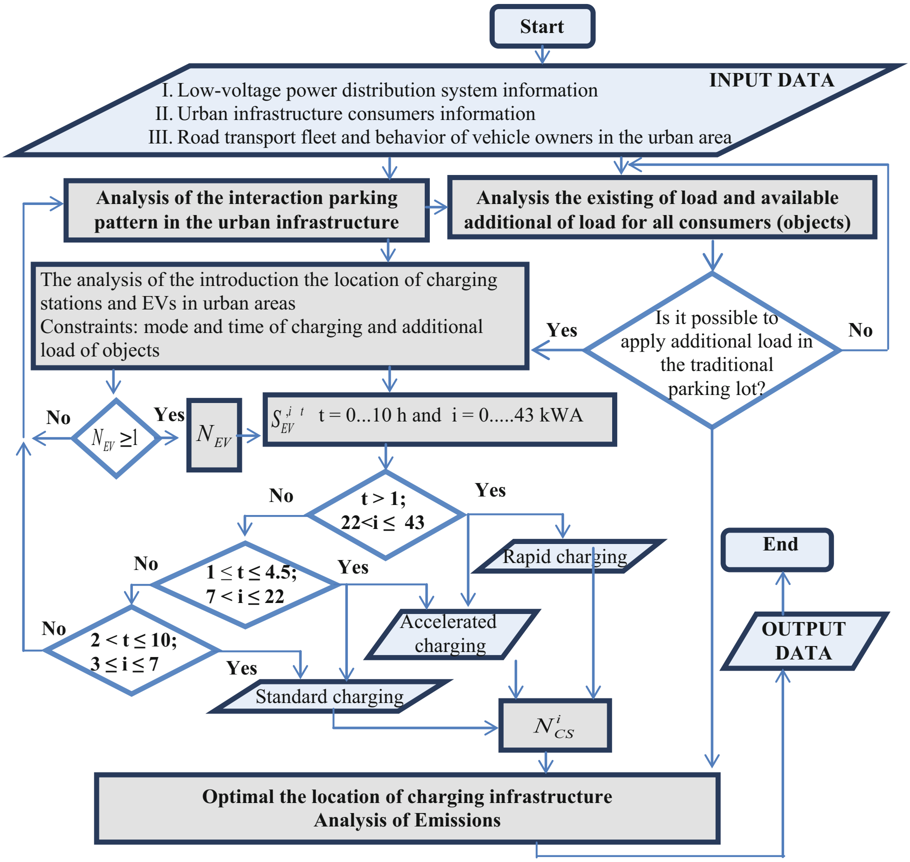


Fig. 1. Block-scheme of the algorithm.

Thus, the aim for the present research is to develop an algorithm for studying a local distribution network in accordance with the needs of electric vehicles by applying mathematical statistics methods, probability theory and the transportation theory methods.

The block-scheme of the algorithm of assessing the location of charging stations is presented in Fig. 1. The implementation of the algorithm is available from the next section of the paper.

2 Methodology of the Algorithm

In this section, the sequence of the algorithm and the basis for choosing the limitations are described. The input data are represented in three parts. Data analysis, accepted constraints and calculation methods are also considered here.

2.1 Part I: Low-Voltage Power Distribution System Information

This part is presented power supply system information. This for modelling of the low-voltage distribution network, the following input data must be identified selected: scheme of the distribution network, the transformer substations technical specifications, the objects (consumers) which have a connection to the transformer substations and the map of territory. In the process for testing of the model is used to a residential low-voltage system of urban district.

2.2 Part II: Urban Infrastructure Consumers Information

The second part is the information about all consumers, namely:

- the number and type of the objects, and also their annual (or daily) power consumption and daily load schedules are used as input data;
- the analysis of the power consumption of the objects and their daily load schedules are estimated by means of two methods – statistical and calculating;
- the quantitative indicators and load graphs: active power (P, kW); reactive power (Q, kVar); apparent power (S, kVA); the coefficients of reactive power of residential buildings; parameters of lighting installations of the transport and pedestrian system of streets and roads; total daily load of consumers of the urban district, etc.

2.3 Part III: Road Transport Fleet and Behaviour of Vehicle Owners in the Urban Area

The location of the charging stations and their integration into distribution power networks directly depends on electrical vehicle driver behaviours. Since the fleet of electric vehicles in Latvia is at an early stage of formation, the initial data for modelling are the behaviours of users of road vehicle with internal combustion engines by gross weight less than 3.5 tons. The collection of information was conducted by traditional methods, such as monitoring and studying existing legislative documents, forms and reports on the location and duration of parking near the house, work, shop, etc. [2, 3]. The information about parking lots was processed by cluster analysis.

The location and duration of parking. In the city, the passenger car is in motion only a small percentage of the day time. During the remaining time it is parked near a dwelling house, an office building, and special parking facility or another convenient place. The hours and the length of time that a car remains at a parking space vary.

Therefore, the location and duration of parking is a necessary condition for studying the behaviour of vehicle users. The behaviour of car owners is interpreted as follows: more than 50% of cars from 20:00 to 07:00 are at the parking places near dwelling houses and at big parking lots, and more than 50% of cars were between 07:00 and 20:00 near the office building and the medical clinic, lastly – near the grocery store between 07:00 and 21:00, Fig. 2.

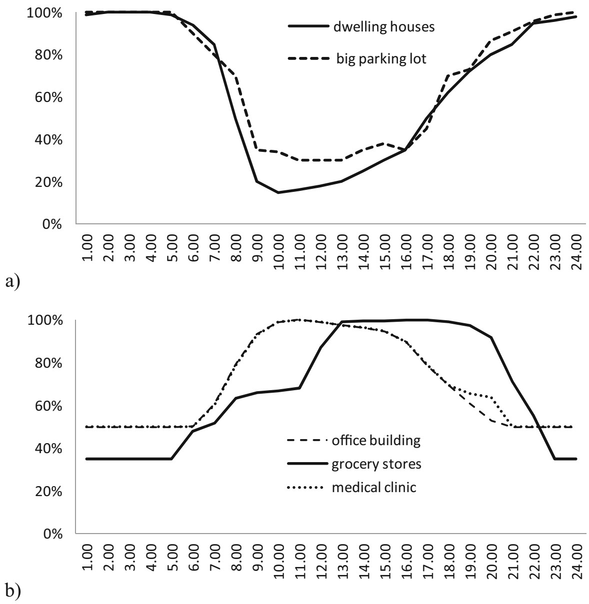


Fig. 2. The distribution of the parking duration: (a) the dwelling houses and big parking lot; (b) the office building, the grocery store and the medical clinic parking.

The average of daily mileage for vehicles till 3.5 tonnes gross weight. In the European Environment Agency (EEA) Guidebook, a standardised was adopted the categories of road vehicles till 3.5 tonnes gross weight: passenger cars (PC) and light commercial vehicles (LCV) [5]. In their turn, as part of a study conducted for the Riga City Council, they vehicles were divided into two groups: owned by private persons and owned by legal persons, as the annual mileage and the daily mileage for each group varies very much. In Riga, according to the 2008–2015 data was revealed, that the daily mileage does not exceed 28 km for private persons and 88 km for the legal persons, Table 1 [6].

Table 1. The average of daily mileage of passenger cars, km, 2008–2015[a].

	2008	2009	2010	2011	2012	2013	2014	2015
Private persons	25.6	25.7	25.5	22.3	22.4	22.4	27.8	27.8
Legal persons	74.3	73.4	87.8	65.7	76.7	76.7	57.4	57.4

[a]taxi cars are not included in this the study.

Electric vehicle battery charging. The main characteristics of Electric Vehicle Battery (BEV) are charging time, power demand and the rate of charge [7]. Therefore, the process of charging the battery is carried out in accordance with the selected Charging Mode (CM), the initial state of charge ($SOC_{initial}$) and the actual state of the charge (SOC_{actual}) of the battery at the end of the process. The communication between BEV, Plug-In Hybrid Electric Vehicles (PHEV) and the EVSE is determined by the ISO 15118 which is designed to support the energy transfer from an EVSE to all types of EVs [8, 9]. By the Eq. 1, each of the CMs, the EVs are charges until it's the required state of charge (*SOC*) reached.

$$SOC_{actual} = SOC_{initial} + SOC(t), \tag{1}$$

where $SOC(t)$ is the state of battery charging process at an instant of time, minute or hour. The time depends on the selected mode and power level of the charging station, Table 2. It is important to note that the model range of vehicles with electric or hybrid types of engines is expanding every year, so it is impossible to consider the entire spectrum of existing battery. The BEVs and PHEVs technology assessment conducted by the California Environmental Protection Agency present an increase in the average mileage per full charge of the battery by 45% by the end of 2016 compared to 2010, and it is expected to still increase substantially by 2025 [10].

Table 2. Charging station modes. Standard charging power levels.

Charging mode	Connection mode (Grid connection)	AC voltage (V)/AC current (A)
Standard charging – SCh (slow): 3 kW (1-phase), 7 kW (3-phase), 6–12 h	Mode 1 (1 phase/3 phase) and Mode 2 (1 phase/3 phase)	230/16 and 400/16 230/32 and 400/32
Accelerated charging – ACh (medium): 7–22 kW (1- or 3-phase, 32 A), 1–4 h	Mode 3 (1 phase/3 phase)	230/32 and 690/250
Rapid charging – RCh (fast): 50 kW (DC) and 43 kW (AC), 80%, less than an hour	Mode 4 (3 phase)	600/400

Based on an analysis of the behaviour of drivers, the following data were received for modelling: the location and duration of parking; the average of daily mileage of vehicles both for private and legal persons; CM stations. Also, taking into account the popular EVs brands in Latvia, in accordance with kWh/100 km of consumption, vehicles can be further subdivided into three groups (small, medium, large) Table 3 [3]. For each group, the following data is of essence: the maximum charging and maximum time of the charge from empty to full. But since, at present, real information about these popular brands is not enough, in this study, only medium group was considered for modelling. It is assumed that all the public charging stations (ChSts) allow to charge via all type charging sockets.

Table 3. Current details and charging times for the ranges per 100 km (BEV and PHEV).

Charging mode	SCh – mode 1 (3–7 kWh)	ACh – mode 3 (7–22 kWh)	RCh – mode 4 (43 kWh)
Small (EV > 16 kWh)	2–5 h	1–2.5 h	0.35 h
Medium ($16 < EV \leq 22$ kWh)	3.5–8 h	1–3.5 h	0.5 h
Large ($22 < EV \leq 30$ kWh)	4.5–10 h	1.5–4.5 h	0.65 h

2.4 Data Analysis, Accepted Constraints and Calculation Methods

Analysis of the interaction parking pattern in the urban infrastructure. The demand of urban consumers and the periods of time intervals of the vehicle user behaviours at parking lots are presented in the analysis, comparing results of the periods of time intervals. The results of the analysis of the interaction parking pattern in the urban infrastructure show the parking availability pattern that lets assume a possible scheduling of EV charging.

Analysis the existing of load and available additional of load for all consumers. Depending on the presence of input data and objectives (requirements may be for season, month and working day or weekend, etc.) the maximum ($S_{\max}$), average (S_{aver}) and minimum ($S_{\min}$) demand and their daily time of use, the load factor (f_{load}) and the irregularity factor (f_{irreg}) are calculated for the each object of a customer. In case of the absence of data, the calculating methods on the basis of mathematical statistics and probability theory are used instead. By use of these analyses, the following question can be answered: Is it possible to apply additional load in the traditional parking lot? If we get the time intervals at which the parking demand and urban consumers' the demand of urban consumers coincide or will be in the same or near time intervals, as well as have an additional load in this time is performed then the transition to the next stage is executed. If such coincidences are not enough for further research, then it is necessary to conduct a more detailed analysis of the power supply system. The next stage is defining the priority of the location of charging stations and number of EVs where is set of constraints: the mode and time of charging and additional load of the urban consumers. As a result of the above review the constraints shown in Table 3 were established. According to the penetration demand and possible additional load the number of EVs (N_{EV}) is calculated for further study by the Eq. (2).

$$N_{EV_i}^t = \frac{S_{add_i}^t}{S_{EV_i}^t}, \tag{2}$$

where $S_{add_i}^t$ - additional load in i period of time, $S_{EV_i}^t = \sqrt{P_{EV_i}^2 + Q_{EV_i}^2}$ - load that represents charging for one *EV* in the period of time. At calculations $Q_{EV} = 0$ was accepted.

Having received the number of EVs allowed for determining what charging constraints will be used to calculate the mode and time of charging. This calculation will be identifying which criteria for the public ChSts as a whole are performed, and which of the criteria performed are absent.

Constraint 1: $t > 1$ (h) is period of time interval and $i \leq 43$ (kWh) the rated at load of charging station. If criteria are performed then this means that in the considered period of time the public RChSts and AChSts can be loaded. If No, then Constraint 2 shall be used.

Constraint 2: $1 \leq t \leq 4.5$ (h) and $7 < i \leq 22$ (kWh). If criteria are performed then this means that in the considered period of time the public AChSts and SChSts can be loaded. If No, then Constraint 3 shall be used.

Constraint 3: $2 < t \leq 10$ (h) and $3 \leq i \leq 7$ (kWh). If criteria are performed then this means that in the considered period of time the public SChSts can be loaded.

In case of failure to meet the constraints, the calculation returns to the analysis of the introduction of the location of charging stations and electric vehicles in urban areas. Further on, modes and numbers of the public charging stations with the power balance are estimated by the Eq. (3).

$$N_{CS}^i = \frac{N_{EV} \cdot S_{EV} \cdot \dfrac{l}{100}}{S_{CS}^i \cdot 24}, \tag{3}$$

where N_{CS}^i – types of public *CSs*; N_{EV} – number of *EVs*; S_{EV} – energy consumption of an *EV*, kWh/100 km; l – daily driven distance of an *EV*, km; S_{CS}^i – power of public *CS* of type i, kWA; 24 – h in day.

Data analysis, accepted constraints and calculation methods. The optimisation process involves a large number of possible solutions by methods of linear programming (LP), which are characterized by an objective function (maximize or minimize) subject to a number of constraints. The following assumptions must be satisfied to justify the implementation of the said methods: the functions must be linear in nature (linearity), parameters are assumed to be known (certainty), and negative values are unacceptable (nonnegativity). The aforementioned occurs in the present model. The problem posed in this study is solved with the Transportation theory method by using MSExcel (Simplex for problems that are linear). To determine the optimal places, types and number of public chargers, it is also necessary to take into account the tariffs for electricity which are differentiated by the times of day, since owners who have installed *EV* charging stations must offer mutually beneficial rates to attract customers. At present, in Latvia, according to the connections and time zones, the objects are subjects to the tariffs for electricity which are differentiated by the times of day: for private persons are based on 1 or 2 time zone tariffs and for legal persons are based on 3 time zone tariffs. The tariffs are applied in weekday: the maximum tariffs are from 8:00 to 10:00 and from 17:00 to 20:00; day tariffs (cost = 77% from maximum tariff) are from 7:00 to 8:00, from 10:00 to 17:00 and from 20:00 to 23:00; the night and weekend tariffs (cost = 46% from maximum tariff) are from 23:00 to 7:00. The single-phase connection in one time zone and three-phase connection with one and two time zones are used for private and legal consumers [11]. Another important part of this algorithm solution is the so called Analysis of Emissions, using the COPERT IV program. This analysis shows the dynamics of harmful emissions process from changes in the road transport fleet.

COPERT is a software tool used to calculate air pollutant and greenhouse gas emissions from road transport, for official road transport emission inventory preparation, relevant research, scientific and academic applications [12].

3 Algorithm Testing

The simulation for the charging stations has been applied to urban low-voltage distribution networks, where transformer substation TS-1 (TR 2 × 1250/10/0.4 kV) provides electricity for 1 601 individual customers: 1 584 (37 objects) are private persons (Household sector) and 18 are legal persons (Tertiary sector). The consumers are: 2 office buildings, 4 grocery stores, 12 dwelling houses (3-storey house), 24 dwelling houses (5-storey house), large parking place (for 500 vehicles), 2 catering services, medical clinic, police-station, post office and a kindergarten and the length of the streets is 3.1 km. The fossil fuel vehicles (*FVs*) owners in this urban area are: 832 – private persons and 208 – legal persons. Analysis of the existing load and available additional load for objects show the temporary daily intervals from 10:00 to 21:00 less than 20%. From 23:00 to 07:00 more than 40% are available additional of load for consumers and for EVs charging stations on the traditional parking lots. The behaviour of vehicle owners was discussed earlier in this article. The average $S_{add} = 418.44$ kVA was determined. As the next step, the numbers of EVs for every hour were defined, taking into account the medium group (Table 3) and the times of day. The average quantitative indicators in the case of the accelerated charging (22 kWh) by the expression (2) are as follows: $N = 19$ of *EVs* for every hour or 456 *EVs* for day (44% from cars fleet in this urban area). Furthermore, the constraints for the public ChSts are studded by an Excel IF function. The objective function of the criteria is the minimum possible number of ChSts and 19 *EVs* charged per hour a according to S_{add} min. The results are as follows: the Constraint 1 is 10 RChSts and 84 AChSts from 24:00 to 6:00 could charge additional *EVs* if necessary. Constraint 2 is 19 AChSts and 58 SChSts from 23:00 to 7:00. Constraint 3 is 60 SChSts and 58 SChSts from 23:00 to 7:00. Verification of the criteria by use the Eq. (3) confirmed the results of the number of charging stations. The transportation problem has the following formulation. Objective function for high penetration demand charging process is follows:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \rightarrow \min, \quad F = 4180x_1 + 3850x_2 + 685.1x_3 + 783x_4 + 538.3x_5 \rightarrow \min \tag{4}$$

and

$$S^j_{add}: \sum_{j=1}^{n} x_{ij=}a_i;\ S^i_{CS\,s}: \sum_{i=1}^{m} x_{ij} = b_j,\ \text{and } x_{ij} \geq 0\ i = \overline{1,m}\ ; j\ =\ \overline{1,n}.$$

$$\text{By the constraints:}\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} + x_{15} = 7796 \\ x_{21} + x_{22} + x_{23} + x_{24} + x_{25} = 2820. \\ x_{31} + x_{32} + x_{33} + x_{34} + x_{35} = 5512 \end{cases}$$

The medium and low penetrations are calculated by using a process similar to high penetration of EVs calculation.

As to results, the location and number of public charging stations were received with high, medium (max km) and low (min km) penetration demand charging process, seen in Table 4, according to the daily mileage for private persons and the legal persons, Table 1.

Table 4. The location and number of public charging stations.

	Dwelling houses	Big parking lot	Office building	Grocery stores	Medical clinic	Total
High penetration of EVs						
SCh	16	15	3	3	3	41
ACh	–	–	1	3	1	5
RCh	–	–	–	1	–	1
Medium penetration of EVs						
SCh	5	4	3	3	3	18
ACh	–	–	–	1	–	1
RCh	–	–	–	1	–	1
Low penetration of EVs						
SCh	3	3	3	3	3	15
ACh	–	–	–	1	–	1
RCh	–	–	–	1	–	1

As a result, the calculations have shown that the number of the SCh stations form 88% of the overall number of public charging stations. To analyse the possible change of the emissions in this urban area the use of *FV* and *EV* technologies, the following comparison was proposed: if the number of *FV* technologies is 100% and if high penetration of *EV* technologies (*FVs* is 66% and *EVs* is 44%). As results, an analysis of emissions was shows that the greenhouse gases (GHGs) emissions are reduced by 27.5%, NO_x by 44.7%, PM10 by 26.7% and PM2.5 by 41%. The summary is illustrated in Fig. 3.

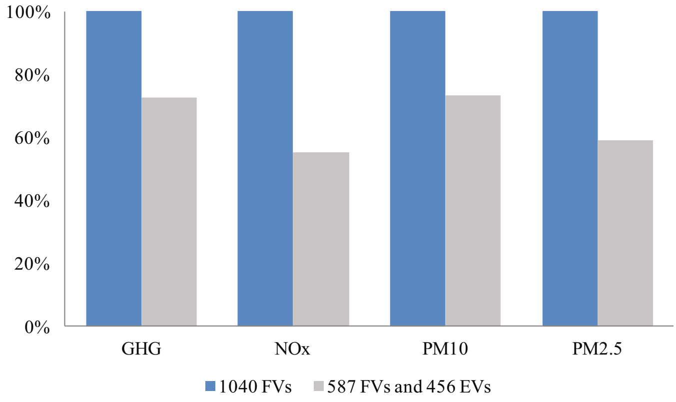


Fig. 3. The emissions analysis.

The final step is the Output data and the final choice for urban area in accordance with the needs for *EVs*.

4 Conclusions

In this study, in order to predict locations for charging stations using models, an approach based on transportation theory algorithm was used. There are several advantages to the algorithm: it allows several target functions to be considered; simultaneously calculate several constraints; and estimate a large number of decisions that are feasible for an urban area in process of the real time. To evaluate the optimal connection points for charging stations in urban areas the behaviour of car owners at the places of traditional parking, the power consumption for the *EVs* charging process and the capacity of the residential network was investigated. The obtained results illustrate that to ensure the living and working of the electric vehicles owners in the urban districts, the Standard charging stations have form a big percent of the overall number of charging stations. There is near 10% need in accelerated and rapid charging stations. At high EVs penetration, the average daily electricity consumption in urban area is increased by 19.6%. In our research, S_{add} exceeds S_{CSs} demand, which means the possibility to increase the number of public charging stations is needed.

In a follow up research our aim would be to extend the algorithm, to include the $SOC_{initial}$ and the SOC_{actual}, the tariffs for electricity which are differentiated by the times of day, increasing the applicability of our architecture to real time driving situations.

Acknowledgements. This paper has been supported by the Latvia National Research Programme 2014–2017 “LATENERGI”.

References

1. Transport 2050: Commission outlines ambitious plan to increase mobility and reduce emissions. EC-IP/11/372 (2011). http://europa.eu/rapid/press-release_IP-11-372_en.htm
2. International Energy Agency: EV city casebook, A look at the global electric vehicle movement. IEA edn., 75 p. (2012)
3. Road Traffic Safety Directorate. https://www.csdd.lv
4. Rigas satiksme. www.rigassatiksme.lv/en/services/parking-services/
5. EMEP/EEA air pollutant emission inventory guidebook. Road transport, 153 p., 30 September 2016. https://www.eea.europa.eu//publications/emep-eea-guidebook-2016
6. Riga Smart City Sustainable Energy Action Plan for 2014–2020. Decision 1358, 131 p. (2014). http://www.rea.riga.lv/files/RIGA_SMART_CITY_SEAP_2014-2020_EN.pdf
7. International Organization for Standardization (ISO): Road vehicles – Vehicle to grid communication interface – Part 1: General information and use-case definition. ISO 15118-1:2013. Technical Committee: ISO/TC 22/SC 31, ICS: 43.120, Geneva, Switzerland, 65 p. (2013). https://www.iso.org/standard/55365.html
8. International Organization for Standardization (ISO): Road Vehicles–Vehicle-to-Grid Communication Interface–Part 2: Network and Application Protocol Requirements, ISO 15118-2:2014, Technical Committee: ISO/TC 22/SC 31, ICS: 43.120 Electric road vehicles: Geneva, Switzerland, 342 p. (2014). https://www.iso.org/standard/55366.html
9. The International Electrotechnical Commission (IEC): Electric Vehicle Conductive Charging System. Part 1: General Requirement; IEC 61851-1:2010; ISBN 978-2-88912-222-6, 99 p. (2010)
10. California's Advanced Clean Cars Midterm: Review Appendix C: Zero Emission Vehicle and Plug-in Hybrid Electric Vehicle Technology Assessment. California Environmental Protection Agency, Air Resources Board, 94 p., 18 January 2017
11. AS Sadales tikls: Rates. www.sadalestikls.lv
12. Gkatzoflias, D., Kouridis, C., Ntziachristos, L. Samaras, Z.: COPERT 4 computer programme to calculate emissions from road transport – user's manual. In: ETC/AEM, European Environment Agency, 70 p. (2012)

The System of the School Routes' Development and Their Safety Assessment

Irina Makarova(✉), Ksenia Shubenkova(✉), Vadim Mavrin, and Aleksey Boyko

Kazan Federal University, Naberezhnye Chelny, Russia
kamIVM@mail.ru, ksenia.shubenkova@gmail.com

Abstract. Since the main type of regular trips of children and teenagers are the trips to school, ensuring school routes' safety is the main direction of the children's safety improving. While constructing the routes to school it is necessary to choose the safest possible one. The categorization of the route can be performed according to the presence of different factors, complicating traffic conditions. Selection of the best route relates to multicriteria optimization problems. To solve this problem it is necessary to use simulation methods as well as modern information technologies. The algorithm of the rational route selection as well as the algorithm of its safety assessment is offered in the article. Mathematical model of the school routes' planning and the conceptual scheme of the Schoolchildren's Transportation Management System are also presented in the article. Proposed system is aimed at reducing the risks of dangerous situations while pupils travelling to school.

Keywords: School route · Safety · Multicriteria assessment

1 Introduction

Urbanization – one of the today's megatrends – poses an enormous challenge. Cities currently consume some 75% of the world's energy and generate 80% of its greenhouse gases. The growth of mega-cities today is accompanied by an increase in the population's mobility and the need of cargo transportation, that causes congestions, air pollution and the risk of traffic accidents. Traffic jams cost the European Union an estimated €100 billion a year in lost economic performance, and road traffic injuries cost governments approximately 3% of GDP [1].

Children and teenagers are the most vulnerable road users among all types of city dwellers: 186 300 children become victims of the road accidents each year [2]. Since the main type of regular trips of children and teenagers are the trips to school, ensuring school routes' safety is the main direction of the children's safety improving. While solving this problem there are two main issues: the selection of the safest walking or cycling routes and ensuring safety of the school buses' routes.

The school routes' safety ensuring problem is complicated by the fact that not enough attention is paid to the needs of pedestrians and cyclists, although the number of road traffic deaths among them achieves 49% worldwide [1]. At the same time the traffic mix

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_6

in many countries means that children and teenagers on their way to school share the road with high-speed vehicles, forcing them to negotiate dangerous situations and fast moving traffic. Ensuring safety of children on the roads will not be possible unless the school routes planning and development is carried out, first, from the point of view of minimizing the number of conflict points that might arise the possibility of the traffic accident.

2 State of the School Routes' Safety Problem

2.1 National Safe Routes Programs: The Worldwide Experience

The program "Safe school route" is developed in many countries. Safe Routes to School programs began in Europe in the 1970s. Today there is a cross-border Sustainable Urban Mobility Plan (SUMP) in the city Nova Gorica (Slovenia) and five other municipalities in Slovenia and the nearby Italian city of Gorizia [3]. The Ljubljana city introduced website, the so-called portal of safe routes, with particular focus on the elementary schools [4]. The public information campaign called "Safe routes to schools" have started in December 2011 as a part of the project Pedibus [5].

In the United States, efforts to promote walking to school emerged in the late 1990s. Initial and ongoing successes led to strong national enthusiasm, inspiring Congress to establish a federal Safe Routes to School program in 2005 [6]. The Healthy Works SRTS project encourages and supports comprehensive SRTS planning, programs and coordination that actively engage local schools, cities, residents, families and other stakeholders in improving routes to school and increasing the number of kids that walk or bike to school.

In Russia, a program for improving the safety of routes to school, are implemented within the framework of the program "Passport to road safety" by educational institutions. Passport displays information about the educational organization from the point of view of ensuring the safety of children on the stages of the route "home-school-home", as well as to conduct training sessions and additional events, and contains various plan safe traffic routes [7]. There is also a created portal which contains information for children, teachers and parents about safe behavior on the road [8].

2.2 World Experience of Ensuring Safety of the School Buses

All over the world a greater focus is being placed on School Buses. According to statistics, school buses are the safest mode of transport: students are about 70 times more likely to get to school safely when taking a school bus instead of traveling by car. That's because school buses in the USA are the most regulated vehicles on the road; they're designed to be safer than passenger vehicles in preventing crashes and injuries [9]; and in every State, stop-arm laws protect children from other motorists [10]. Outside North America, purpose-built vehicles for student transport are not as common. However, major accidents with a large number of children who died on the way to school made central government of China to announce strict rules for school buses. For example, all buses must have seatbelts and GPS transponders [11].

The main field of the use school buses in Russia is the transportation of the schoolchildren from the rural areas where there are no schools, and pupils have to travel to neighboring communities for long distances. According to [12], pupils have to be transported by school buses if they live more than one kilometer from the place of study. Due to the fact that regional roads in rural areas are unsafe, the situation with the school transportation system leaves much to be desired. The federal program "School Bus" is realized to ensure safety while children transportation. This program implies to supply the schools by buses specially designed for the children transportation and that are corresponding to GOST 33552-2015. However, to ensure children's safety while transportation is still a serious problem. That's why the "Rules for the organized transportation of a group of children by buses" have been approved [13].

Transportation of pupils to schools in Europe is a lot like that in the United States, but with key operations and safety differences. According to Department of Defense Education Activity [14], safe student transportation in Europe consists of three parts: high standards for mechanically sound vehicles, qualified, trained drivers and safe student behavior on the bus. However, in the USA, school bus transportation safety has lately been considered from the point of view of the choice of specific locations for school bus stops and the planning of school bus routes [15].

3 Methods and Software Solutions to Improve Safety of the School Routes

3.1 Methods and Models to Solve the Vehicle Routing Problem

The history of solving the Vehicle Routing Problem (VRP) starts more than half a century ago. Since the mid-1990s, research has focused on the so-called metaheuristics that represent a certain method for constructing a complete heuristic for a particular task. The most interesting are the following methods: Taboo search [16], Genetic Algorithm [17], Ant Colony Optimization [18] and neural networks [19]. All these methods can also be used to solve the School Bus Routing Problem (SBRP). Although a number of studies focusing on SBRP can be found in the related literature with a variety of purposes, the following issues appear as the most frequent goals: (1) to minimize costs associated with transportation [20], (2) to minimize time spent on transportation [21], and (3) to minimize the total distance travelled by all buses [22]. The authors of the research [23] consider the SBRP as bicriterion problem, where the first criteria is minimizing the travelling time and the cost of travel simultaneously. The research work [24] is devoted to solving the SBRP as the multicriteria problem. The objectives considered include minimising the total number of buses required, the total travel time spent by pupils at all pick-up points, and the total bus travel time. Moreover, the proposed heuristic algorithm allows balancing the loads and travel times between buses.

Thus, despite a large number of studies in the field of SBRP, none of the proposed algorithms have considered the optimization of school routes in terms of their safety.

3.2 Intelligent Systems to Plan Safe Routes

Thanks to the technological advancements in areas such as the global positioning system (GPS), geographical information systems (GIS), mobile communication networks, and trac sensors, it is now possible to solve the vehicle routing problems in a dynamic and real-time manner. In the world practice for more than 40 years the problems of optimal transport management have been solved with the help of Intelligent Transportation Systems. These systems consist of modules for collecting operational information on parameters of the transport flow, analyzing imbalances and searching the causes of their occurrence with the use of GIS technologies and modeling techniques, as well as developing recommendations for the processes optimization.

To optimize the management of transport systems, decision support systems are being created. DSS, according to its functional purposes, are more similar to expert systems. The difference is that the experts make decision based on recommendations developed by DSS's intelligent heart (mathematical and simulation modeling, evolutionary computation and genetic algorithms, neural networks, situation analysis, cognitive modeling, etc.). To build models and to perform statistical analysis, monitoring data obtained with the help of satellite navigation systems (GLONASS, GPS) is usually used as initial data. For example, in the research [25] the software for improving currently available school bus routes is presented. Authors suggest storing the data on GPS locations of the students and coordinates of stops on the route with their latitude and longitude values in the proposed system's database. GPS locations of students and the school bus are transferred to a server in real time through the Android-based mobile software. The optimized routes are determined by intelligent heart of this system based on Ant Colony Optimization, genetic algorithm metaheuristics and K-means clustering method. Authors of the research have formed the most suitable routes by the locations of these coordinates on Google Maps and then have transferred them to the school bus over the server. In order to store the planned route and points at the route a SQLite database is used.

Today there are a lot of different electronic resources (route planners and navigators) that allow building a route between two points on the map. These are Yandex Maps, Google Maps, 2GIS (in Russia), jakdojade.pl (in Poland). However, despite the large number of applications for path finding, they have one significant drawback: the route is not evaluated from the point of view of its safety.

4 Proposed Solutions

Depending on the choice of the mode of transport school routes can be walking, cycling, routes of school buses and combined. The combined route usually includes areas where pupils use public transport or the school bus and they get to the bus stops on foot or by bike. Since while constructing the school routes it is necessary to choose the safest possible one, their assessment should be performed due to the different factors, complicating traffic conditions. Therefore, we propose a methodology a multicriterial evaluation of the safety of the route.

4.1 Proposed Algorithm of the Safest Route Selection

As a first stage, the factors determining the category of the route complexity are identified (Table 1). Factors that determine the safety of the route can be both objective (e.g., terrain, presence of unregulated intersections, etc.), and subjective due to the features and physical condition of the pupil. Adequacy of the assessment will depend on the correctness of the selected factors and their combined inclusion. For example, the same route can be safer in the daylight than in the darkness, in the summer than in the winter, etc.

Table 1. Factors influencing the route's safety.

Factors	Significant factor for		
	pedestrians	cyclists	school bus
The length of the route	+	+	+
The length of sections with elevation changes	+	+	+
The number of pedestrian crossings	+	+	+
The number of unregulated intersections	+	+	+
The number of lanes of the highway at the place of pedestrian crossings	+	+	+
The number of ramps	+	+	–
The number of underground/raised pedestrian crossings	+	+	–
The length of bikeways on the route	–	+	–

Factors that influence on the safety of the walking and combined routes can include, first of all, the number and the type of crossings of the road. Despite the fact that pedestrian crossings are safe places for pedestrians where they are given priority, most often the deaths of pedestrians occur precisely during the crossing the road.

The evaluation of the route can be done with the help of a complex indicator K, which is calculated according to the formula of the factors' weighted average values:

$$K = \sum_{i=1}^{n} K_i \cdot \alpha_i, \tag{1}$$

where K_i – the value of the i-th factor, α_i – weight of the i-th factor.

Since the factors influencing the route's safety have different dimensions, before adding them in formula 1 the standardization procedure needs to be performed.

First of all initial data should be specified. Possible routes are evaluated on the base of this information. To do this a matrix of standardized values of factors is constructed and then the complex indicator of each route is calculated with provision for correction coefficients that depend on the physical condition and characteristics of the user. The generic scheme of the routes safety assessment algorithm is shown in Fig. 1.

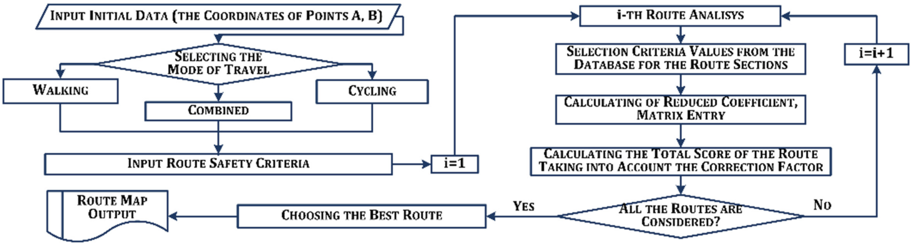


Fig. 1. The integrated algorithm of the route safety assessment.

4.2 Proposed DSS for the School Bus Routes Planning and Their Safety Assessment

The planning of school bus routes usually consists of evaluating possible variants of displacements and choosing the optimal one based on several criteria. At the same time, their safety should be taken into account as one of the essential indicators, that is, we consider the SBRP as a problem of multicriterial optimization taking into account the safety of schoolchildren. We have identified the following criteria:

1. Route safety complex indicator. While developing routes of the school buses the factors that determine the level of the route's safety can include the number of the areas of the route where accidents are likely to happen, statistical data on accidents, etc. In addition, this indicator should also take into account the safety indicator of the walking routes of the approach to the bus stop.
2. Costs associated with transportation. Transportation costs largely depend on the number of used vehicles, the number of routes and their length. Therefore, the number of buses, routes and their total length should seek to a minimum.
3. The load balance. It is the search for the minimum deviation of the scope of work of each bus. This criterion is defined as the number of transported passengers to the number of kilometers that these passengers have traveled.
4. Balance of distances. This criterion is defined as the minimum difference of the route distances.
5. Walking distance from the place of living to the bus stop. This indicator represents the total distance from the residence of all schoolchildren to the points for their collection along the safest walking routes.

In addition, the school bus network should meet the following restrictions: (1) each bus stop should be visited only once, (2) the beginning of all routes is established from the garage (g), and the end conditionally is in the school (s), (3) the number of schoolchildren transported by bus in one run should not exceed the capacity of buses, and (4) the maximum travel time spent by a schoolchild for one trip to school should not exceed (D), where D is the matrix of distances between selected bus stops.

To solve the problems of planning the processes of organization of school transportation, it is necessary to provide the possibility of operative simultaneous access to information databases (geographic, economic, technical and operational, etc.). The data in these databases are constantly updated, relevant, complete and diverse. This information should be presented in a form convenient for analysis and should ensure the adoption of the most rational decisions, both in the choice of vehicles, and the construction of an optimal route.

In our view, operational control requires a software solution that integrates modules that provide the ability to quickly collect information, store large amounts of data, and also analyze them intellectually. A conceptual scheme of the interaction of the modules of the proposed DSS is shown in Fig. 2.

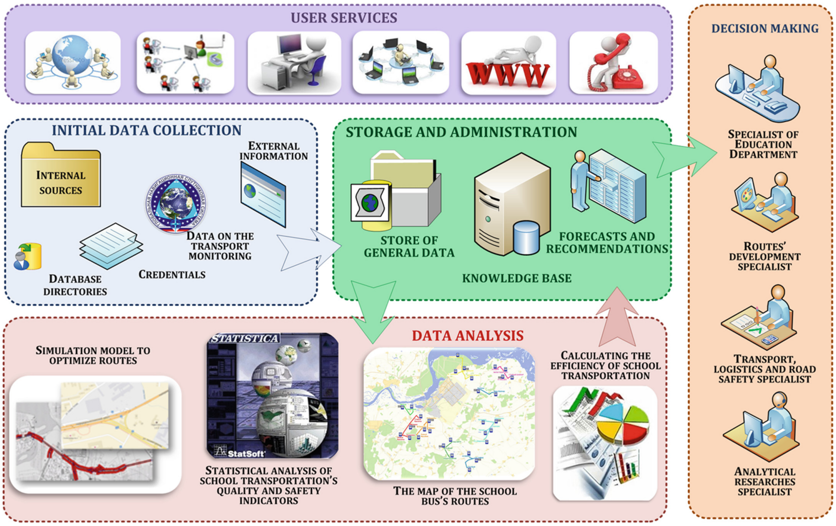


Fig. 2. Conceptual scheme of the Schoolchildren's Transportation Management System.

The proposed DSS combines following functionalities:

- an intelligent system that allows selecting optimal routes for school buses by solving multicriterial optimization problem with given constraints, and also to evaluate the effectiveness and safety of the decisions taken;
- control and monitoring system based on the use of the navigation system and GIS-technology;
- system for processing, storing and analyzing Big Data;
- informing system for different groups of users, implemented as the user services.

The routes are constructed using information about the city road network, which allows determining the cost and distance of movement between all the required nodes. The intelligent heart of the proposed Schoolchildren's Transportation Management System is the simulation model based on the mathematical model for solving the School Bus Routing Problem. The objective function is composite and it is divided into several subfunctions:

1. f_T^D – the total distance run by school buses (formula 2);
2. f_T^W – the total walking way of schoolchildren (formula 3);
3. f_T^C – the total cost associated with schoolchildren's transportation (formula 4).

$$f_T^D = \sum_{k \in K} \sum_{(j,l) \in E} d'_{jl} \cdot z_{jl}^k, \tag{2}$$

$$f_T^W = \sum_{(i,j) \in W} d''_{ij} \cdot x_{ij}, \tag{3}$$

$$f_T^C = \sum_{k \in K} [c'^k \sum_{l:(g,l) \in E} z_{gl}^k + \sum_{(j,l) \in E} c''^k \cdot d'_{jl} \cdot z_{jl}^k], \tag{4}$$

where K – total number of all buses; J – selected bus stops; I – total number of schoolchildren; c'_k – fixed costs per trip by one bus; c''_k – variable costs per 1 km of the k-bus; d'_{jl} – the shortest safe distance between bus stops j and l; $W = \{(i,j):i \in I,\ j \in J\}$ – the matrix {schoolchildren – selected stops} (if (i, j) ϵ W, then d''_{ij} – safe walking distance of the i- schoolchild to j-stopping point; E – the network graph consisting of all safe arcs between any two selected stopping points, all safe arcs starting from the point of departure (garage) and the destination (school); N – total number of graph points $N = J \cup \{g\} \cup \{s\}$. It is assumed that the graph has no loop-like sections.

The parameters of the model are the following binary variables:

$$x_{ij} = \begin{cases} 1, & \textit{if the bus stop j is selected by the student i} \\ 0, & \textit{else} \end{cases}, \forall (i,j) \in W, \tag{5}$$

$$y_j = \begin{cases} 1, & \textit{if the bus stop j is used by any student} \\ 0, & \textit{else} \end{cases}, \forall j \in J, \tag{6}$$

$$z_{jl}^k = \begin{cases} 1, & \textit{if the } k - \textit{bus passes through the arc } (j,l) \\ 0, & \textit{else} \end{cases}, \forall k \in K, \forall (j,l) \in E. \tag{7}$$

The proposed software solution consists of several modules that implement various functions. The initial conditions (characteristics of the transportation area, the number of schoolchildren in each locality indicating year of education, location of schools, the structure of the bus fleet and its technical characteristics, location of the garages and other elements of the processing and technical base, etc.) and parameters for the calculation are entered into the database. The functions of building and editing the map and the road network are realized using GIS-technologies. This module contains a set of geometric primitives for representing the elements of the road network with their attributes (number of lanes, speed limits, traffic on the road network sections is in both or only one side, etc.). After the graph construction, the elements of the real system are presented in the mathematical model, and all the necessary structures are ready for triggering the algorithms. All possible functions of the proposed system are based on routing algorithms and presented in the main menu. They return information on the constructed routes, represented as a sequence of vertex numbers. Then the user can evaluate the quality of the solutions obtained, recalculate them, if necessary, with other parameter values, or save the results in the form of reports if they suit him. The "Route" tab displays possible routes in the order of their priority (from the best to the worst), as well as the bus schedule.

The peculiarity of the school bus routing is that the determining parameter is not to minimize the cost of transportation, but to ensure safety, reliability, regularity and compliance with restrictions that are regulated by official documents. Only after determining the set of such alternative routes options, the determining parameter is the cost price. The algorithm for choosing the optimal route is shown in Fig. 3.

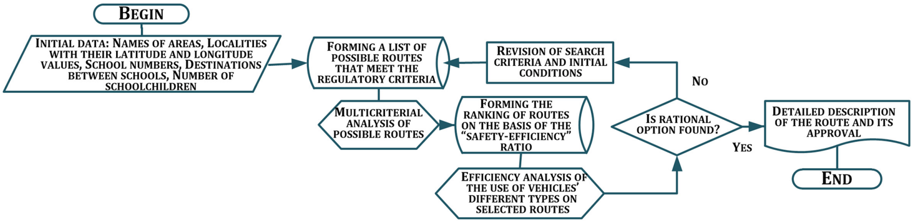


Fig. 3. Algorithm to select rational route of the school bus.

5 Conclusions

The School Bus Routing Problem in Russia is relevant, primarily, when organizing pupil's transportation from the rural areas where there are no schools. Since the garage of the buses is situated in a nearby city and there are several different localities for the children's collection, the empty runs of the buses are very large. Furthermore, route planning and development is the function of principals who do not have any necessary knowledge and skills in this field. The proposed approach and the software developed on its basis will simplify the routing process, will contribute both to improving the safety of school routes, and to reducing transportation costs.

References

1. Global status report on road safety 2015. http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/. Accessed 01 July 2017
2. Ten strategies for keeping children safe on the roads. http://apps.who.int/iris/bitstream/10665/162176/3/WHO_NMH_NVI_15.3_rus.pdf. Accessed 01 July 2017
3. Nova Gorica makes school routes safe with CIVITAS co-funding (Slovenia). http://www.eltis.org/discover/news/nova-gorica-makes-school-routes-safe-civitas-co-funding-slovenia. Accessed 01 July 2017
4. Safe routes to schools. http://civitas.eu/content/safe-routes-schools. Accessed 01 July 2017
5. Pedibus, http://pedibusbeograd.com/wp-content/uploads/2013/02/Finalni-Izvestaj-Sigurni-putevi-do-%C5%A1kola1.pdf. Accessed 01 July 2017

6. Safe Routes to Schools. http://www.saferoutestoschools.org/. Accessed 01 July 2017
7. Passport to road safety. http://www.gibdd.ru/about/social/pasport-road-safety/. Accessed 01 July 2017
8. Encyclopedia of road traffic. http://deti.gibdd.ru/. Accessed 01 July 2017
9. Standard No. 131 School bus pedestrian safety devices. https://www.gpo.gov/fdsys/pkg/CFR-2016-title49-vol6/xml/CFR-2016-title49-vol6-sec571-131.xml. Accessed 01 July 2017
10. National Highway Traffic Safety Administration Homepage. https://www.nhtsa.gov/road-safety/school-buses/. Accessed 01 July 2017
11. American-Style Schoolbuses Make Inroads In China. https://www.wired.com/2012/02/american-style-schoolbuses-make-inroads-in-china/. Accessed 01 July 2017
12. Methodical recommendations "About the organization of students' transportation to schools". http://ugnaya.depon72.ru/?p=766. Accessed 01 July 2017
13. Decree of the Government of the Russian Federation of December 17, 2013 N 1177 "Rules for the organized transportation of a group of children by buses". http://pravo.gov.ru/proxy/ips/?docbody=&nd=102170019. Accessed 01 July 2017
14. Department of Defense Education Activity. http://www.dodea.edu/Europe/offices/transportation/newcomersInfo.cfm. Accessed 01 July 2017
15. School bus operation guidelines. http://ascip.org/wp-content/uploads/2014/05/ASCIP-School-Bus-Stop-Planning-Guidelines-10-19-2015.pdf. Accessed 01 July 2017
16. Zachariadis, E.E., Tarantilis, ChD, Kiranoudis, ChT: A guided tabu search for the vehicle routing problem with two-dimensional loading constraints. Eur. J. Oper. Res. **195**, 729–743 (2009)
17. Abouhenidi, H.M.: Genetic algorithm in vehicle routing problem. Int. J. Sci. Eng. Res. **5**(5), 1517–1523 (2014)
18. Bin, Y., Zhong-Zhen, Y., Baozhen, Y.: An improved ant colony optimization for vehicle routing problem. Eur. J. Oper. Res. **196**, 171–176 (2009)
19. Hasan, L.S.: Evaluate neural network for vehicle routing problem. Int. J. Recent Trends Eng. Technol. **7**(1), 19–21 (2012)
20. Bögl, M., Doerner, K.F., Parragh, S.N.: The school bus routing and scheduling problem with transfers. Networks **65**(2), 180–203 (2015)
21. Arias-Rojas, J.S., Jiménez, J.F., Montoya-Torres, J.R.: Solving of school bus routing problem by ant colony optimization. Revista EIA Escuela de ingenieria de Antioquia **17**, 193–208 (2012). Accessed 01 July 2017
22. Kang, M., Kim, S.-K., et al.: Development of a genetic algorithm for the school bus routing problem. Int. J. Softw. Eng. Appl. **9**(5), 107–126 (2015)
23. Widuch, J.: A relation of dominance for the bicriterion bus routing problem. Int. J. Appl. Math. Comput. Sci. **27**(1), 133–155 (2017)
24. Li, L., Fu, Z.: The school bus routing problem: a case study. J. Oper. Res. Soc. **53**, 552–558 (2002)
25. Ünsal, O., Yiğit, T., Altıntaş, C.: Optimization of dynamic school bus routing problem by using metaheuristic and clustering methods. In: Proceedings of the 11th International Conference on Practice and Theory of Automated Timetabling (PATAT-2016), pp. 555–559 (2016)

The Impact of Selected Road Freight Transport Management Measures for the Society and Environment

Tatjana Vasiljeva[1(✉)] and Michael Minx[2]

[1] RISEBA University, Meza iela 3, Riga 1047, Latvia
Tatjana.Vasiljeva@riseba.lv
[2] KEDGE Business School, 680 Cours de la Liberation, 33405 Talence, France
Michael.Minx@gmail.com

Abstract. Nowadays logistics, supply chain management and usage of various transport become the key components of every company constituting a considerable percentage of the overall costs that a company incurs in its daily business. Logistics accounts amount up to 15% of total expenditure of the largest industry sectors when the proportion of transportation achieves one to two thirds of the total costs of logistics. A lot of researchers believe that effective functioning of flow of goods is not conceivable without a functional transport.

The given research is focused on the studying the importance and the irreplaceable function of the freight transport industry with focus on transportation of goods by road. Methodology of the research: Scientific and fundamental literature critical analysis; Primary data gathered by Quantitative method (questionnaire) and Qualitative method (semi-structured interviews), Data analysis and Interpretation of the results. The research period was March – April 2017, when 129 respondents had answered the questionnaire and the managers of transport companies were interviewed. The results of the research could be valuable for top management of transport companies and other practitioners giving several recommendations for setting relevant key performance indicators (KPI) as effective management measures.

Keywords: Road freight transport · Telematics system
Key performance indicators

1 Introduction

Freight transport is a vital component of the economy. It supports production, trade dealings, consumers' consumption activities, as well as efficient movement and timely availability of raw materials and finished goods. Freight transport constitutes a significant part of the final cost of products and represents an important component of the national expenditures of any country. This field of study is increasingly attracting attention as it has become one of the key issues from the political, economic and social points of view [1].

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_7

Doubtfully, freight transport and more specifically road freight transport is a drive force of the economic activity. For instance, in the EU, the road transport is responsible for 2.8% of employment and 2.2% of value added, as per the following Table 1.

Table 1. Freight transport in EU in numbers (Source: The authors based on [2]).

	Employment 2009		Value added 2010	
	Absolute figures (×1000)	Share of total economy (%)	Absolute figures (mln. Euro)	Share of total economy (%)
Land transport	6 314	2.8	232 045	2.2
Water transport	223	0.1	44 259	0.4
Air transport	371	0.2	26 528	0.4

Due to the economic recession after 2008, the behaviour of the logistics services providers and the customers has changed. Drop in sales and consequent limitation of production decreased the volume of goods transported. The transporters were forced to relocate resources and even reorganize their logistics system [3]. Logistics and supply chain management is one of the key components of every firm as it constitutes a considerable percentage of the overall costs that a company incurs in its day to day business [4].

A lot of researchers have investigated the issues of transportation and supply chains [3, 5–7] claiming that the most important operational factors of road freight transport that impact the environment and society include environmental pollution, occupation of land, traffic congestions and risks resulting from transport of dangerous goods. Environmental pollution includes the following negative impacts: air, water and soil pollution, noise, vibration and waste.

Eisler [8] states that the negative impacts on the environment depend on the type of vehicle used and on the means of transport. The effects of the various means of transport can result in different levels of air, water and soil pollution as well as noise and vibrations, having long-term and cumulative effect. Furthermore, road freight transport has a considerable impact on global CO_2 emissions and other greenhouse gases becoming a centre of attention of many critiques and are often regarded as unsustainable [9–11].

The authors [12] argue that the elements of sustainable freight transport include: "efficient scheduling and vehicle use; shortening supply chains and policies to promote shortening distances and reducing volumes of freight". Anderson et al. [13], Gros [3] add the need of governmental involvement in educating privately own companies in cost effectiveness of using more ecologically efficient options in road freight transport operations and to "address a market failure due to consumers misperceiving the benefits of improved energy efficiency."

For the purpose of this paper, the research questions about the efficiency of road freight transport and key performance indicators (KPI) in the road freight transport will be studied. The efficiency of road freight transport has a considerable impact on the economy, as inefficient transport leads to decreased company competitiveness, impaired expansion of the market and higher negative impact on the environment and the society [14]. Efficiency of road freight transport will be understood as non-financial measures.

Fuel consumption of road freight transport vehicles appears to be one of the key KPI in the industry, as it is a crucial indicator for road freight transport companies, but it is also an indicator that correlates with the negative effect of road freight transport on the environment and sustainability [15]. Fuel consumption is a “productivity measure showing the efficiency with which energy is converted into the movement of freight” [14].

Road freight transport management is a broad subject of study and may be analysed through a substantial number of research fields, management theories, concepts and approaches. Novak [16] identified the following areas of road freight transport management: customer focus, understanding and integrating into the logistic system of the customer, creating computer-based system compatible with the customer ones, preferring strategic alliances with customers, quantifying and benchmarking, applying controlling or training the personnel.

Several studies suggest that an on board Information and Communication Technologies (ICT) system, known as telematics, also have a significant effect on the drivers’ performance. The authors [17–20] stated that a vast majority of drivers would appreciate a monitoring system in their vehicles leading to an ameliorated workplace and consequently to an arguably more satisfied workforce. These findings appear to represent just the tip of the iceberg of what benefits on-board ICT bring to a haulier company. Consequently, carrier companies invest significant amount of resources in ICT and trucks are now often referred to as “computers on wheels” [19]. ICT in trucks is otherwise called as telematics and may be explained as the combination of the transmission of information over a telecommunication network and the computerised processing of this information.

However, the on-board monitoring is also a subject of controversies. The amount of information, constant monitoring and recordings made by telematics appear to be a privacy concern for many drivers, some of which even consider it as spying [21]. Peng et al. [18] noticed that 20% of truck drivers are opposed to the new telematics technologies due to the fear of losing their privacy.

2 Research Design

According to the conceptual model (Fig. 1) or the research two selected management measures, fuel efficiency management measures and the installation of on-board telematics were investigated. The researchers stated the hypothesis that the efficient management measures of a trucking company will lead to ameliorated road freight transport efficiency KPIs which lead to limiting the negative externalities of the industry and strengthen the crucial characteristics of road freight transport on the economy and companies. The crucial characteristics of road freight transport on the economy and companies, as well as the negative externalities of the industry, require focus on ameliorating the efficiency KPIs of road freight transport which could be done by using the selected management measures efficiently.

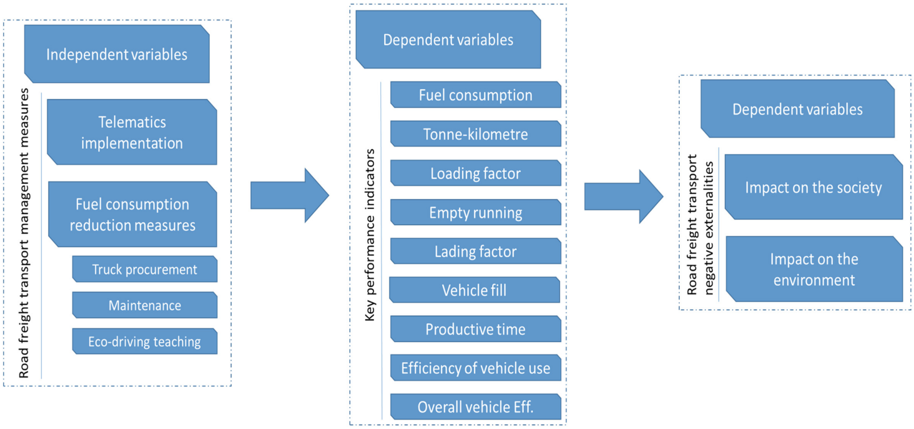


Fig. 1. The relationship between the studied variables – conceptual model of the research (by the authors).

The first research question (RQ1) aims to provide the answer to how does road freight transport affect the society and how the public perceive the incentives to lower the impact of road freight transport on the society. Therefore, in order to provide an answer to this research question, it is necessary to collect data from the public, preferably with as many people as possible in order to have statistically relevant and significant results. Online survey with quantitative questions was selected for answering the RQ1. Quantitative questions were used for the online survey allowing carry out statistical comparison and get descriptive data. As the survey questions were structured according to how positively or negatively people perceive selected negative externalities of road freight transport the scale questions method allowed to illustrate how negatively or positively the respondents perceive the industry and its influence on the society and the environment.

RQ2: what effect do telematics have on road freight transport KPIs? And RQ3: how do the selected management measures affect the impact of road freight transport on the society and the environment?

The research design for answering the RQ2 and RQ3 questions needs to be based on collecting primary data in a form of structured interviews from a road freight transport company which operates with its own fleet, has telematics installed in its trucks, as well as a fuel management system with data availability. The selected company was a transportation and logistics company that currently operates in four European countries being one of the leading Central European entities providing transportation and logistics services (Fig. 2).

The online survey helped in getting answers in the following aspects: (a) the questions regarding the respondents' profile (living environment, transportation habits, etc.) and (b) the questions regarding the impact of road freight transport on the society. The structured interviews answered to the following aspects: (a) the questions regarding fuel consumption management; (b) regarding the effects on the environment; (c) regarding improvement of truck efficiency through telematics; (d) regarding the effects on the environment. The online survey was carried out using the Google Drive forms platform, 129 respondents replied to the survey questionnaire.

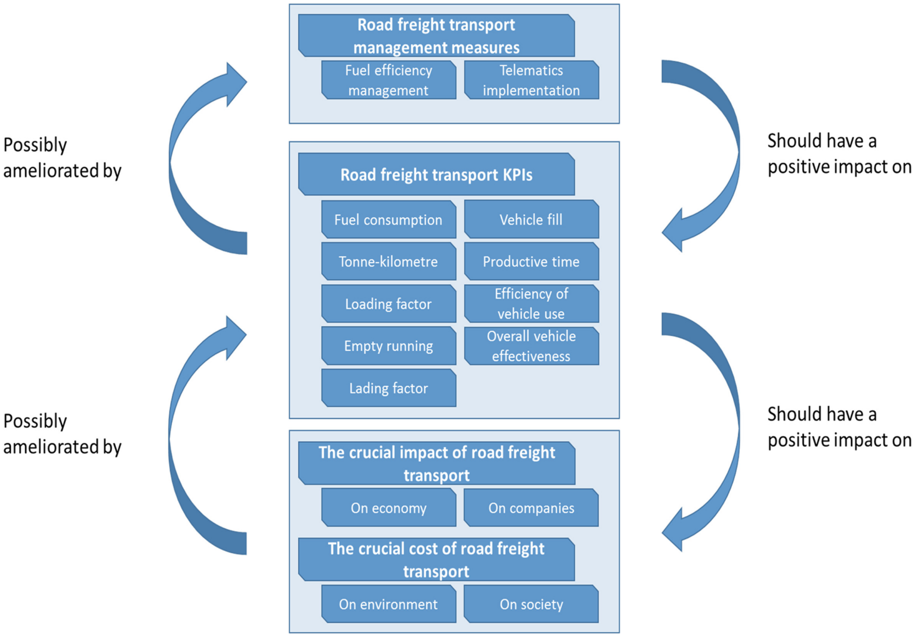


Fig. 2. The content and logic of the theoretical part of the research (by the authors).

For better checking the primary data gathered from the interviews with the managers of the international company, supplementary interviews with one of the transport company operating in Baltic countries were carried out during summer 2017.

3 Results, Analysis and Interpretation

Figures 3, 4, 5, 6, 7 and 8 below illustrate the main aspects of inhabitants' perception of the road freight transport and its impact to the life conditions.

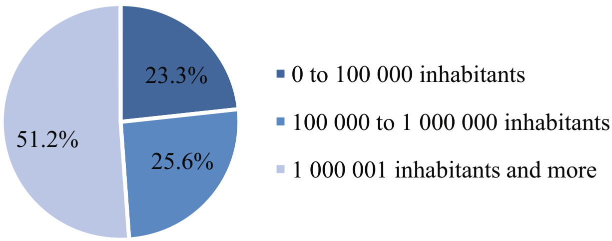


Fig. 3. Characteristics of the respondents.

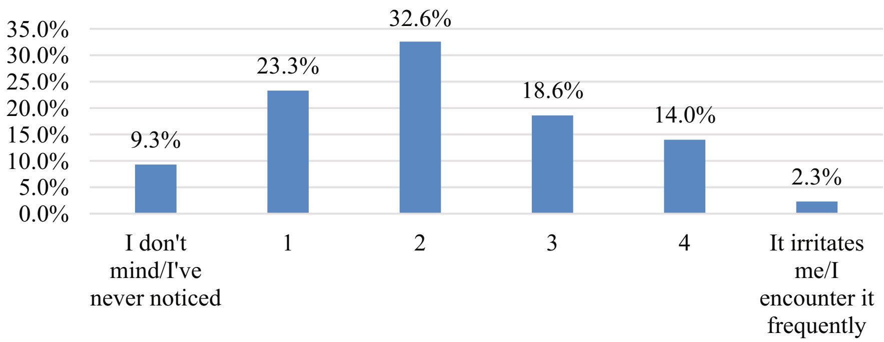


Fig. 4. The affect of the noise to the respondents.

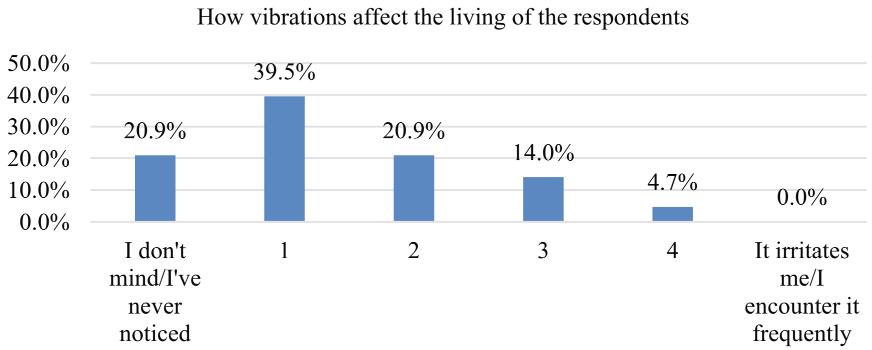


Fig. 5. The vibrations affect to the respondents.

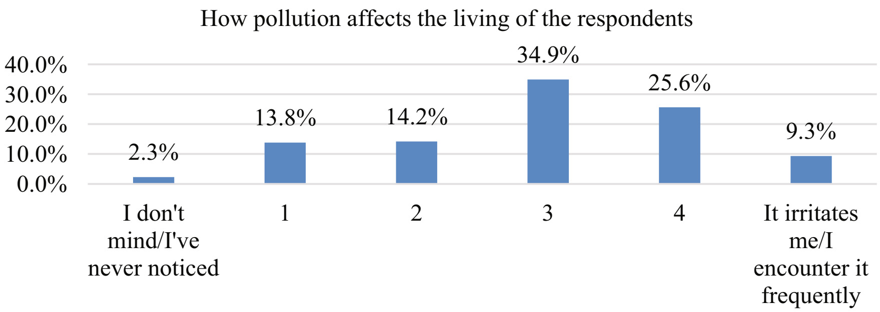


Fig. 6. The pollution affect to the respondents.

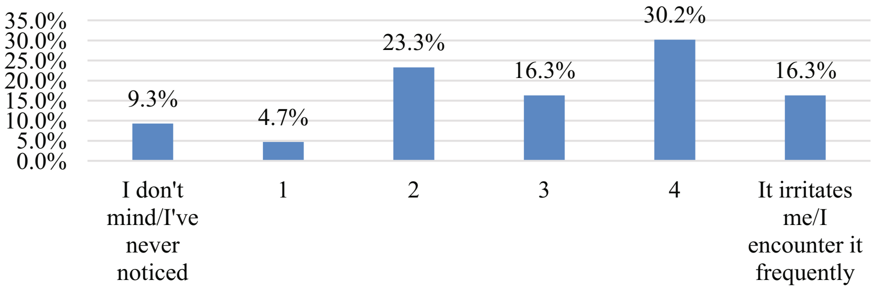


Fig. 7. The respondents' perceiving of the traffic congestions caused by road freight transport.

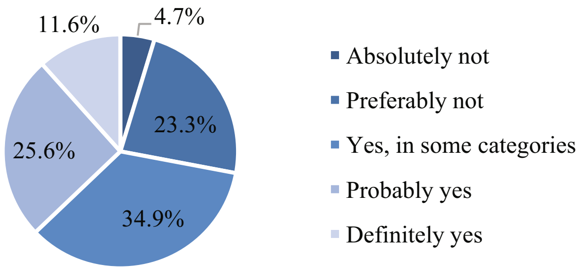


Fig. 8. The attitude to the convenience of immediate availability of the favourite products.

Respondents were characterised by the size of the city where they currently live in with the hypothesis that people living in large cities will have a different perception of the road freight transport than people living in smaller cities where the heavy trucks movement may have different characteristics.

Lastly, the respondents were asked would you sacrifice immediate availability of some of your favourite products in exchange of lessen traffic of heavy trucks.

The main objective of the online survey was to measure the impact of negative externalities resulting from road freight transport by heavy trucks based on the four main negative externalities on the society presented in the literature review – noise, vibrations, pollution and traffic congestions on a scale from 0 to 5, where 0 represents "I don't mind/ I've never noticed" and 5 represents "It irritates me / I encounter it frequently" (Table 2).

For answering the RQ3 semi-structured interviews with the managers in transport area were conducted. The questions related to the main aspects of telematics: (a) the questions regarding improvement of truck efficiency through telematics; (b) the main characteristics of telematics installed on board; (c) the effects of the telematics to following KPI like fuel consumption; tonne-kilometres; weight-based loading factor; empty running; productive time; efficiency of vehicle usage; overall vehicle effectiveness and some others; (d) the questions regarding the effects on the environment.

Table 2. Negative impact of road freight transport to the respondents.

Respondent' profile	Noise	Vibrations	Pollution	Traffic congestions	Total average
Large city	2.14	1.32	3.14	2.86	2.36
Mediums city	2.18	1.64	2.91	3.09	2.45
Small city	2.00	1.40	2.60	3.30	2.33
City centre	2.33	1.87	3.27	3.20	2.67
Outskirts	2.18	1.55	3.18	3.09	2.50
In between	1.88	0.94	2.53	2.82	2.04
Use a car	2.40	1.75	3.25	3.55	2.74
Do not use a car	1.87	1.13	2.70	2.57	2.07
Weighted average	2.12	1.42	2.95	3.02	2.38

Summarising the views about telematics usage on board it may be concluded that telematics system installed across a fleet of trucks is a crucial tool in helping to manage the trucks more efficiently and therefore, to ameliorate the KPIs especially related to fuel consumption, empty running and the productive time of the vehicles. The fuel consumption is reduced by telematics system thanks to information recollection of data which may help to improve proper maintenance of the vehicles and to set the optimal route between the point of loading to the point of unloading and whether the driver respected the set route, and even to prevent theft of the fuel. The availability of precise information regarding empty running allows the managers to see whether the trucks are being used efficiently in terms of what orders are accepted to be transported and thereafter address the issue if any arises. Lastly, the telematics helps to improve the productive time of the vehicles by identifying any inefficiency of planning the routes and/or truck drivers' lack of performance. However, telematics mainly allows only to collect the data and do not improve the KPI directly, it is up to the management to use the data collected by the telematics system and put them to use in order for the road freight transport KPI to be ameliorated.

4 Conclusions and Discussion

This paper discussed the importance and the irreplaceable function of the freight transport industry with focus on transportation of goods by road. It presented the reasons behind its importance, as well as the negative externalities that accompany the road freight transport industry and the effect it has on the environment and the society. The research part focused on evaluating to what extent the society is impacted by the negative externalities of road freight transport and on evaluating the two selected management measures (fuel consumption management and implementation of telematics) in terms of how the management measures impact the road freight transport KPIs and how they affect the impact of road freight transport on the environment and the society.

Based on research interviews, the authors could recommend the managers of the transport companies the following management measures.

- The top management of the company should consider tracking the weight and volume of the transported loads and use the information for a more efficient use of its fleet with focus on the possibility of transporting goods from more customers at a time;
- The literature claims that teaching eco-driving to drivers may lead to up to 15% improvement of fuel consumption, so the companies' management team should focus more attention to the trainings themselves to identify any potential issues and tracking the long-term effects of eco-driving trainings;
- Many of the efficiency KPI identified in the literature may lead to increased efficiency of a trucking company, the top management should therefore consider collecting the data that are necessary to track and manage these indicators.

5 Limitations and Further Research

A considerable amount of KPI exists in the road freight transport industry and most are focused on the financial aspect of the business. This paper is focusing only on the main KPI related to road freight transport efficiency in transportation of goods by heavy trucks.

The online survey regarding the influence of negative externalities generated by road freight transport was available online for the public without differentiation of the country the respondents live. However, the negative influence of trucks on the society may vary from country to country.

Acknowledgements. The authors would like to thank to management transport company participated in the research which dedicated the hours of their time discussing the subject of road freight transport and the interviews' questions. We are grateful to all the respondents of the online survey, for taking the time to answer the questionnaire and to everyone who helped sharing the online survey in the social networks.

References

1. Ben-Akiva, M., Meersman, H., Van de Voorde, E.: Recent developments in freight transport modelling. Emerald (2013). http://www.emeraldgrouppublishing.com/products/books/pdf/chapters/freight_c1.pdf
2. Meersman, M., Van de Vorde, E.: The relationship between economic activity and freight transport. In: Ben-Akiva, M., Meersman, H., Van de Voorde, E. (ed.) Freight Transport Modelling. (2013). http://www.emeraldinsight.com/doi/abs/10.1108/9781781902868-002
3. Gros, I.: Velka kniha logistiky, 1st edn. Vysoka skola chemicko-technologicka v Praze, Prague (2016)
4. Ameknassi, L., Ait-Kadi, D., Rezg, N.: Integration of logistics outsourcing decisions in a green supply chain design: a stochastic multi-objective multi-period multi-product programming model. Int. J. Prod. Econ. **182** (2016). http://www.sciencedirect.com/science/article/pii/S0925527316302274
5. Andreas, R., Holter, D.B., Grant, J.R.N.S.: A framework for purchasing transport services in small and medium size enterprises. Int. J. Phys. Distrib. Logist. Manag. **33** (2008). http://www.emeraldinsight.com/doi/pdfplus/10.1108/09600030810857193

6. Kherbach, O., Mocan, M.: The importance of logistics and supply chain management in the enhancement of Romanian SMEs. Procedia – Soc. Behav. Sci. (2016). http://www.sciencedirect.com/science/article/pii/S1877042816302038
7. Tousek, R.: Management dopravy, 1st edn. Jihoceska univerzita, Ekonomicka fakulta, Ceske Budejovice (2009)
8. Eisler, J.: Ekonomika dopravnich sluzeb a podnikani v doprave, 1st edn. Vysoka skola ekonomicka v Praze, Fakulta podnikohospodarska, Praha (2005)
9. Proost, S., Van Dender, K.: What sustainable road transport future? Trends and policy options. Joint Transp. Res. Center (2010). http://www.itf-oecd.org/sites/default/files/docs/dp201014.pdf
10. Becker, U., Gerike, R., Winter, M.: Zaklady dopravni ekologie, 1st edn. Ustav pro ekopolitiku, Prague (2008)
11. Cahill, M.: Transport, Environment and Society, 1st edn. McGraw – Hill Education, London (2010)
12. Schiller, P., Bruun, E., Kenworthy, J.: An Introduction to Sustainable Transportation – Policy, Planning and Implementation, 1st edn. Earthscam, London (2010)
13. Anderson, S., Fisher, C., Ian, P., Sallee, J.: Automobile Fuel Economy Standards: Impacts, Efficiency, and Alternatives. National Bureau of Economic Research (2010). http://www.nber.org/papers/w16370.pdf
14. McKinnon, A.: Performance measurement in freight transport: its contribution to the design, implementation and monitoring of public policy. Int. Trans. Forum (2015). http://www.itf-oecd.org/sites/default/files/docs/mckinnon.pdf
15. Szabo, M., Majdan, R., Tkac, Z., Capora, R., Hujo, L.: Evaluation of fuel consumption in road freight transport. Acta Technologica Agriculturae **16**(1) (2013). https://www.degruyter.com/downloadpdf/j/ata.2013.16.issue-1/ata-2013-0005/ata-2013-0005.pdf
16. Novak, R., Pernica, P., Medved, J.: Mezinarodni kamionova doprava a zasílatelstvi, 1st edn. C.H. BECK, Praha (2013)
17. Bartlett, G., Mastrippolito L.: Volvo trucks field operational test: evaluation of advanced safety systems for heavy truck tractors. Department of Transportation, Federal Motor Carrier Safety Administration (2005). https://ntl.bts.gov/lib/jpodocs/repts_te/14349_files/14349.pdf
18. Peng, Y., Ghazizadeh, M., Boyle, L., Lee, J.: Commercial drivers' initial attitudes toward an on-board monitoring system. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting **56**(1) (2012). http://journals.sagepub.com/doi/pdf/10.1177/1071181312561480
19. Duffy, O., Wright, G.: Fundamentals of Medium/Heavy Duty Commercial Vehicle Systems, 1st edn. Jones & Bartlett Publishers, Burlington (2015)
20. Boodlal, L., Chiang, K.H.: Study of the impact of a telematics system on safe and fuel-efficient driving in trucks. Transportation Research Board (2014). https://trid.trb.org/view.aspx?id=1312403
21. Bowman, R.: Is new truck-monitoring technology for safety – or spying on drivers? Forbes – Logist. Transp. (2014). https://www.forbes.com/sites/robertbowman/2014/02/11/is-new-truck-monitoring-technology-for-safety-or-spying-on-drivers/#6fa60b0649f8

Contribution to the Optimization of the Operation of an Urban Railway Line. Case Study: Constantine Tramway

Kahlouche Abdelaziz(✉) and Chaib Rachid

Laboratory of Transportation Engineering and Environment,
Department of Transportation Engineering, Frères Mentouri University, Constantine, Algeria
kahloucheaziz@yahoo.fr

Abstract. This paper proposes a new optimization method for the operation of the Constantine tramway. The method is based on data of operation recorded in the line since June 2013. The objective of this work is to analyze the current operation of Constantine line and its performance, in order to improve both the attractiveness of the line, through competitive offers, regularity, commercial speed and capacity, and its economic viability, through a redistribution of resources and a better matching of its offer to its demand. This work reports on the first optimization of Algerian tramway operation.

Keywords: Urban railway transportation · Adequacy offer/demand Optimization · Tram of Constantine

1 Introduction

The road transportation mode is being criticized more and more because of its negative impact on the environment and the public health. Therefore, in the context of of sustainable development, collective modes of transport have become increasingly attactive. Common urban transportation modes such as subways and tramways have been undergoing major developments over the years. In big cities, the use of the public transport is unavoidable for several reasons. First, it improves the quality of the environment and at the same time protects energy resources. The tramway, for example, generates no atmospheric pollution, unlike cars. Furthermore, the use of the public transport allows the reduction of road congestions and the costs of transportation. Globally, it generates economic and social profits. Constantine, a major city in Algeria, has strongly benefited from this modern urban means of transportation [1]. Improving the quality of the tramway services and the satisfaction of the users are the main challenges to cope with if we want to increase its use over the much preferred private means.

The paper presents the development of a real-time passenger flow distribution model to calculate the average waiting time and the total number of trams in operation for each period of time, taking into account the constraints of the transport organization and the conditions of passenger displacement (comfort, safety, etc.).

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_8

2 Problem Description

Railway operations are planned, executed and managed under four broad hierarchies; strategic, tactical, operational control and real-time control [2]. When transport capacity of urban railway line cannot meet traffic demand, operators should compress departure intervals and use extra trams to complement transport capacity [3]. But the problem found by the authors in the tramway of Constantine is that operator uses a high number of trams compared with the demand for transport, which implies negative effects on the economic side (extra costs) or on the social side (fatigue of the employees) and consequently a lack of performance by time. This article establishes a mathematical model which aims at the minimum average waiting time and the optimum numbers of trams in operation, under the constraints of transport organization.

3 Trams Scheduling Optimization Method

3.1 The Average Interval Time

The frequency of an urban line is the number of trams that serve this line in a fixed time interval (e.g. in one hour). The line optimization problem consists in choosing a set of operating lines and its frequencies to serve the passenger demand and to optimize some given objective [6]. The mathematical method in this paper is based on the calculation of the number of trams to be used in an urban line in order to meet the transport demand with a minimum waiting time in the stations on the one hand and to minimize the costs of operating on the other hand, taking into account the constraints of the capacity of the trams and the general travel time.

The method is based first of all on the calculation of the time interval between the trams (I) with respect to the number of real-time passengers V(t) generally within one hour (3 600 s) and the tram capacity (C) considering the comfort and safety of passengers. Firstly, we calculate the number of trams by a period of time (Tp), it is calculated by the number of passengers in a period of time divided over the capacity of the tram; and then the time period is divided over the result obtained, that giving the time interval (i):

$$\mathrm{Tp} = \mathrm{V(t)/C}, \tag{1}$$

$$\mathrm{I} = (\mathrm{t} \times \mathrm{C})/\mathrm{V(t)}. \tag{2}$$

3.2 General Travel Time

Secondly, we calculate the general travel time T by the addition of the travel time on the railway track 1 Tv^1 and 2 Tv^2 and the boarding passenger's time in the stations Ts and the return time in the terminals Tr including the tram driver change time.

$$\mathrm{T} = \mathrm{Tv}^1 + \mathrm{Tv}^2 + \mathrm{Ts} + \mathrm{Tr}. \tag{3}$$

3.3 The Total Transport Volume of Trams

Finally, after the calculation of the two preceding parameters, the total number of trams in operation expressed as N can be calculated by the division of the general travel time over the average interval time. The mathematical expression can be presented as follows:

$$N = \frac{T \times V(t)}{t \times C}. \tag{4}$$

4 Case Study: Constantine Tramway

4.1 Presentation of the Line

The Tramway transport system of Constantine comprises a double-track line with a length of about 8 km and comprises 10 stations, (see Fig. 1). The poles of bus exchanges and cultural centers are taken into account in the design of the system [4]. The tramway of Constantine serves urban, suburban and being urbanized areas, with an average of 24 000 passengers/day, it aims to improve the offer of transport in the city by promoting a sustainable mobility.

Fig. 1. General view of the Constantine tramway line.

4.2 Determining the Transport Demand in the Line

In order to plan urban railway transport services, one needs reliable information on the number of the passengers traveling from each station in the line within a specific time interval. Figure 2 shows the maximum number of passengers in the Constantine tram line in the normal mode of operation (excluding weekends and holidays).

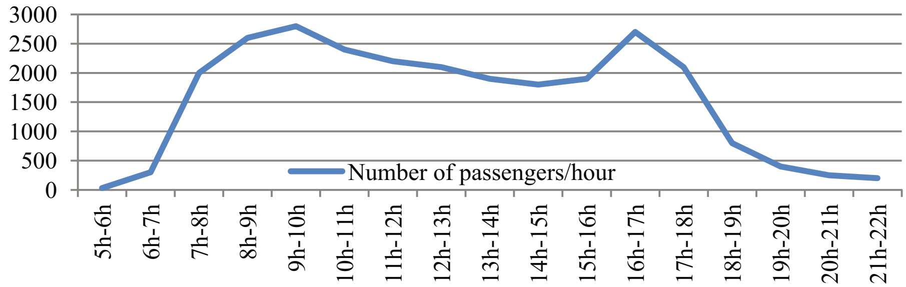


Fig. 2. Distribution of the daily number of passengers per hour.

According to the preceding graph, the hours of operation can be divided into two parts: Peak hours that record the highest number of passengers [7 am–12 pm, 4 pm–6 pm] and off-peak hours with the lowest number of passengers [5 am–7 am, 12 pm–4 pm, 6 pm–10 pm]. In order to give more flexibility to our calculation, we take for each period of time the maximum number of passengers/hour, for the period of 5 am to 6 am we notice that there is a very low number of users (>25) who use the tram in this period, on the other hand there is a fatigue of employees due to lack of sleep, that leads to lack of alertness, impaired performance, and occurrence of incidents [5]. So we propose that the operation starts at 6 am.

4.3 Operating Constraints

The tram capacity C calculations are carried out under normal load with 4 passengers/m^2, so 302 passengers/tram.

The general travel time T in the tramway of Constantine is dependent on the commercial speed and the length of the path, noted that the stopping time in the stations is the time necessary to climb and disembark of passengers in high conditions of safety and comfort; it is calculated of 30 s in each station. The result is presented in the following Table 1, and this time may decrease in certain period.

Table 1. The general travel time in Constantine tramway

Track	Time			
	Travel time without all (Seconds)	Time of return (Seconds)	Passengers boarding time (Seconds)	**Total**
Track 1	1020	240	300	**1560**
Track 2	900	360	300	**1560**
Total	**1920**	**600**	**600**	**3120**

4.4 Determination of the Optimum Number of Trams

Currently in the normal mode the operator uses 14 trams for peak hours and 8 trams for off-peak hours, we will calculate in this section the optimum number of trams for this transport request using the method described in Sect. 3. Possible objectives of this result is the minimization of the total cost of operation of Constantine tramway and the maximization of the passengers comfort satisfying certain regulations [6]. So in order to get a link between operation/demand and in order to get a high level of quality of service we will assume that the maximum waiting time in the stations is 10 min. The result is shown in the Table 2.

Table 2. The optimum number of trams/times period.

Times period	Passenger's number V/hour	Travel time T (second)	The theoretical number of trams N^t	The real number of trams N^r	The average interval time I (minute)
5 am–6 am	25	/	/	/	/
6 am–7 am	350	3 000	1	5	10
7 am–12 pm	2 800	3 120	8.03	9	5.77
12 pm–4 pm	1 900	3 120	5.45	6	8.66
4 pm–6 pm	2 700	3 120	7.74	8	6.5
6 pm–10 pm	800	3 000	2.29	5	10

5 Conclusion

The mathematical method presented in this paper offers a simple yet realistic representation of urban railway line operation. A real-time passenger flow distribution model is described, which has two objectives: calculating the minimum of the average waiting time and the optimum of transport volume for trams. The results obtained on the Constantine tramway case indicate that the current operating plan generates unnecessary costs. They also suggest ways for the minimization of the total cost of operation of Constantine tramway and for the maintenance of a high level of quality of service.

References

1. Kahlouche, A., Chaib, R.: Analysis of tram safety: case study of Algeria. Procedia Eng. **178**, 401–408 (2017)
2. Narayanaswami, S., Rangaraj, N.: Scheduling and rescheduling of railway operations: a review and expository analysis. Technol. Oper. Manage. **2**(2), 102–122 (2012)
3. Yang, J., Xu, Y., Xu, R.: Optimization method of urban rail transit train running scheme. In: Proceedings of IEEE CCIS (2011). https://doi.org/10.1109/ccis.2011.6045140
4. Alstom Transport: General description of the transport system, Tramway of Constantine (2008)
5. Tsao, L., Chang, J., Ma, L.: Fatigue of Chinese railway employees and its influential factors: structural equation modelling. Appl. Ergon. **62**, 132–141 (2017)
6. Bussieck, M.R., Winter, T., Zimmermann, U.T.: Discrete optimization in public rail transport. Math. Program. **79**, 415–444 (1997)

Sustainable Aviation and Maritime Transport

Risk Assessment of Emission Abatement Technologies for Clean Shipping

Sina Atari(✉) and Gunnar Prause(✉)

School of Business and Governance, Tallinn University of Technology,
Acadeemia tee 3, 12618 Tallinn, Estonia
{Sina.Atari,Gunnar.Prause}@ttu.ee

Abstract. The purpose of this study is recognizing and assessing the existing risks of SECA related investments of ship owners and the consideration of their risk attributes. Complying with the SECA regulations, maritime stakeholders have to choose among different abatement strategies, which are generally linked to high and risky investments. The paper focusses on the evaluation of scrubber technologies and their relationship to other abatement techniques.

Literature review reveals shortcomings in investment risk evaluation among the ship owners operating in emission control areas (ECA). The research fills this gap by presenting a comprehensive compilation of identified risks attributes in an analytical framework together with a risk assessment in the context of HFO and MGO fuel and scrubber related performance indicators comprising CAPEX and OPEX. The results in a classification framework categorize the investment risks and different elements of value at risk (VaR) as well as historical and parametric evaluation of risks. Besides that, this study contributes to new knowledge in the disciplines of green transport and shipping. For future research, the identified risk and investment must be tested in a real business case study and in different scenarios to measure and analyze its performance and efficiency.

The results of the paper are based on empiric activities, which were realized during 2017 in the frame of the EU project "EnviSuM". The empiric measures comprise primary and secondary data analysis, focus group meetings and expert interviews with specialists from shipping sector in BSR.

Keywords: Investment appraisal · SECA regulation · Payback period
Value at risk · Scrubber

1 Introduction

In order to improve the Maritime's carbon footprint and to make shipping greener, Sulphur Emission Control Areas (SECA) were implemented in Northern Europe which force ship operators to use on-board fuel oil with a Sulphur content of no more than 0.10% [1]. All global SECA regions together represent about 0.3% of the world's water surface and includes the North Sea, the English Channel (together with the coastal waters around USA and Canada) and the Baltic Sea region (BSR) [2]. Strengthened regulations and environmental awareness are of vital importance to stimulate clean shipping but it

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_9

has a number of consequences for shipping business, which are linked, directly or indirectly, to their economic decisions [3].

Since 2015, maritime stakeholders are forced to comply with SECA regulations in BSR in order to be able to run legally a shipping business by taking under account the benefits, acceptable risks and the investment costs of the available compliance options. Olaniyi and Viirmae [4] pointed out that SECA regulation compliance costs (transaction costs) are high and can interfere with the effective productivity; so, making compliance choices can be considered as being strategic for whole shipping industry. However, recent results from BSR show that due to low oil price most maritime actors tried to postpone risky investment decisions by using low Sulphur oil [5].

A large number of failures of long lead transport investments stress that risk represents an important element in all investment decisions [6]. This applies especially for investment risks related to recently implemented SECA regulations in BSR. Literature reveals that there is a lack of research on investment risk appraisals for green shipping industries. Thus, this study aims to analyze the capital budgeting practices and to measure the risks for abatement technologies of shipping companies in BSR from a comparative perspective. The research is based on quantitative and qualitative data collection, which took place in the frame of "EnviSuM" project within the last year.

The results show that shipping companies are able to face the SECA investments risks and the estimated payback periods turn out to be rather short. The right choice of bunker fuel together with the corresponding abatement solution positively influences the payback period as well as the related risks, which are assessed with methods of VaR.

2 Theoretical Background

2.1 SECA Compliance

Since 2015, the SECA regulations in BSR limit the sulfur content of fuel to 0.1% in the emission control areas. Three alternative solutions can be used for compliance (1) switching to low sulfur fuel, (2) installing LNG-compatible machinery or (3) installing an exhaust gas scrubber which all need to be assessed [7, 8]. On the other hand, Patricksson and Erikstad [9] carried out five possible initial machinery concepts, which considered as a main solution: diesel machinery, diesel machinery with a scrubber system, dual fuel machinery, pure gas engines (LNG), and dual fuel ready complete machinery. There are sets of reconfiguration possibilities available also for each alternative solution. If the ship is already running on a low Sulphur fuel with no traffic outside of ECA, and is compliant with the rules, there is obviously no need to install a secondary cleaning method. One might also want to have a look at the annual fuel usage, and compare it to the installation cost and OPEX of a scrubber system.

The crude oil itself is then refined to various products such as MGO, MDO or HFO. The price of crude oil is based on supply and demand. Prices are affected by short-term expectations depending on economic forecasts, production estimates from the oil producing countries, stock levels, seasonality, accidents, weather and force major situations [10, 11]. HFO fuel turns out to be the cheapest, the most popular and available bunker fuel in shipping industry but it usually has Sulphur content of about 3.5%, i.e. it exceeds the limits of SECA regulations.

Most well-known viable bunker fuel types are IFO 180 and 380 which are both intermediate fuel oil and they are often mixed with different portions of residual oil and distillate oil. On the other hand, MGO and MDO are distillate oils with Sulphur content less than 0.1%, thus they are more expensive for production, and i.e. they have a higher price than other bunker oils but they comply with SECA requirements.

The price spread between HFO and MGO play an important role for selecting the right investment in this research. Based on expert interviews, Olaniyi et al. [5] predicted a sharp reduction in HFO demand by 2020 which will be accompanied by a strong increase of distillates fuels such as MGO, MDO, and ULSFO (Ultra-light Sulphur fuel oil). Thus, a scrubber installation together with the use of globally available HFO becomes a viable option for ship owners. The attractiveness of scrubber use increased significantly after MEPC's announcement of the "global fuel Sulphur cap 2020" in September 2016, which limits the global Sulphur content of maritime fuel to 0.5%.

The most commonly and widely used scrubbers are still wet scrubbers which are washing the exhaust gases. The initial investment costs of scrubbers range from €2.5 to €5 million per ship. The costs depend on particular features such as ship capacity, engine and boiler type, scrubber type and new build or retrofit for used ships. Scrubbers need space for installation and extra space for all the equipment consisting of the scrubber, pumps, tanks, engines and a piping system for the wash water allowing the use of scrubbers only in large vessels [10]. Additionally increased operating costs of scrubber have to be considered due to higher energy needs for the scrubber support systems, which cause higher fuel consumption up to 3% or even more [4].

Consequently, ship owners have to concern not only about the availability of compliant Sulphur fuel, but also about the price spreads between bunker fuels, the scrubber investments costs as well as additional operating costs.

2.2 VaR in Investment Appraisal

Future operating costs and revenues in the appraisal of investment opportunities are linked to uncertainties and risks, which also apply to scrubber investments. In the case of scrubber investments, the assessment uses usually dynamic approaches comprising net present values (NPV) and payback periods [12]:

$$\mathrm{NPV} = \sum\nolimits_{t=1}^{n} \frac{\mathrm{CF}_t}{(1+r)^t} - \mathrm{Outlay},$$

or introducing CF_0 yields:

$$\mathrm{NPV} = \sum\nolimits_{\mathrm{t}=0}^{\mathrm{n}} \frac{\mathrm{CF}_{\mathrm{t}}}{(1+\mathrm{r})^{\mathrm{t}}},$$

CF_{t} Represents the cash flow during period t and the outlay the investment expenditures at period zero. Further variables are the normal lifetime of n periods of a scrubber (usually 15 years) as well as the annual average interest rate of the investment r. The cash flow of period t is calculated according to the following formula:

$$CF_t = V_t \cdot spread_t \cdot (100\%{-}e\%) - add\_cost_t,$$

with V_t annual fuel consumption, $spread_t$ = HFO – MGO spread in period t, e% additional scrubber energy consumption, and additional cost add_cost_t. Under the condition of NPV > 0 the investment is usually considered as favorable since it adds a positive value to the capital of the company [13].

Besides the capital budgeting calculation related to an investment, the paper also considers the risk aspects. The linkage between capital budgeting methods and their related risks will be done by using a VaR approach that describes the capital or percentage of capital loss to be surpassed with an assured probability or words confidence level over a certain period [14]. In the case of a scrubber investment, a 10% VaR (over a horizon of 2 years) of 500 t€ means, that there is a 0.1 probability that the value of the scrubber investment will fall to a value of more than 500,000 € (over a period of 2 years). Since the value of a scrubber, investment depends on the distribution of the price spread between HFO and MGO the spread distribution together with its quantiles have to be calculated by using historical data analysis [15]. The related value of the scrubber investment will be determined by the NPV of net fuel cost savings. The research identifies the VaR of scrubber investments enabling ship operators to quickly determine the risk level of their investments, i.e. the VaR is set equal to the loss on the scrubber investment at the hundreds of X percentile point of the distribution [16]. In theory, the investors seek to maximize the overall amount of return consistent with the rate of risk they feel appropriate [17] (Fig. 1).

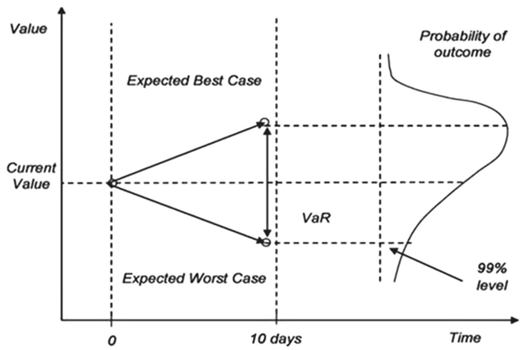


Fig. 1. Value at risk for investment appraisal [18].

3 Methodology

The research is based on empirical data from different sources. Expert interviews, focus group meetings and case studies have been carried out between March 2016 and May 2017 within the frame of the EU project "EnviSuM". These sources have been used to understand and assess the cost structures of different abatement and scrubber technologies as well as to identify trends and motivations of ship owners associated different scrubber investments decisions. The result of this research is important to the maritime industry investors to make a long-term investment strategy and with model minimize the risks during the process.

The research has been complemented by the analysis of statistical data and historical data over the last four years in order to determine the distribution of the spread between HFO and MGO used in shipping industry. The statistical analysis used data from one of the Estonian bunker fuel producer and conducted several statistical methods comprising correlation analysis, the calculation of empiric probability distribution of the spread value between HFO and MGO of the last four years as well statistical test theory.

The empiric probability distribution of the spread serves as input data for a Value-at-Risk analysis for scrubber investments by associating to each spread the NPV of a scrubber investment with the corresponding spread.

4 Data Analysis

The historical data analysis revealed a high correlation between HFO and MGO. The two main variable MGO and HFO of the current research in our study correlated with each other with a positive Pearson coefficient, which is, proves the results (Fig. 2).

Then, based on the sample data, the empiric probability distribution of the spread between HFO and MGO between 2013 and 2017 was calculated and tested. Both tests, the Kolmogorov–Smirnov test as well as the Shapiro-Wilk test confirmed a normality of the distribution of the spread between HFO and MGO (Fig. 3).

The statistical results have been combined with the other empirical data to model the VaR scenario for scrubber installations. The historical approach of VaR used in this research relies on a quantity, which already specified the period. Historical simulation approaches use the actual interval of observation period. The VaR with certain confidence level α is: $Prob(x \leq -VaR\alpha) = 1{-}\alpha$, If the distribution is bounded below then it means that if the probability density function of the distribution is $f(x)$ and the –L is lower bound of distribution so the model will be [19]:

$$\int_{-L}^{-\mathrm{VaR}\alpha} \mathrm{f(x)dx} = 1 - \alpha.$$

The model then was empirically validated by a case study for a ferry shuttling daily between Tallinn and Helsinki.

5 Case Study and Discussion

In the case study, the research focused on ferry ship – RoPax type which operates daily between Tallinn–Helsinki. The engine has a power of 48 MW and a maximal speed of 27 knots. Expert interviews revealed that for this ship a suitable scrubber system requires

Fig. 2. The spread of the HFO and MGO (USD) over the period (Calculations by Authors).

a power of 15 MW and the cost for an open loop scrubber is 4 984 000 million € plus installation costs of 0.7 million €. Scrubber installation, construction, testing, commissioning etc. will take about thirty days, which makes ship out of service in those days. In addition to that, annual maintenance costs of approximately 21 t€ p.a. are estimated as well as material costs of about 300€/ton fuel have to be added for chemicals and waste treatment of scrubber residuals.

The considered RoPax ferry has a daily fuel consumption of 60 t HFO, which yields an annual HFO bunkering volume of 60 t × 360 days = 21 600 tons. The data studied covered a period of 962 days where oil prices were observed to measure the spread between two major fuels HFO and MGO and to calculate VaR in the lower 10%, 5% and 1% quantiles of the related distribution of the spread. In theory, the VaR calculations focus on left tail of probability distributions. Consequently the corresponding value of a quantile is calculated by discounted value method, so the fuel cost saving which are set in the case study to $n = 15$ in the means of discounted value over the 15 years, and r = 11% as a risk free value. The results lead to the following Table 1.

Mean	273.4884376
Standard Error	2.011091842
Median	270.625
Mode	344.25
Standard Deviation	62.37627562
Sample Variance	3890.79976
Kurtosis	-1.147358631
Skewness	-0.129092029
Range	254.125
Minimum	134.5
Maximum	388.625
Sum	263095.877
Count	962

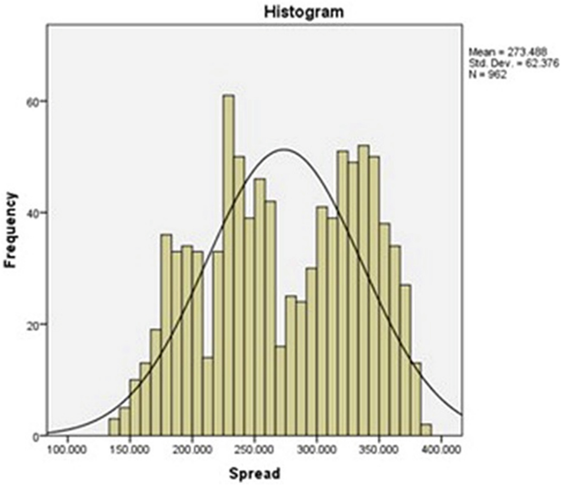


Fig. 3. Distribution of HFO/MGO spread (USD) (Calculations of Authors).

Table 1. Quantiles and PV of spread distributions earnings during project life time (Calculations by Authors).

Historical data	Days	Saving money from the fuel difference annually/Euro	Fuel spread	PV (Euro) in 15 years
10% days	96.2	3,356,640.00	155.4	19,768,031.74
5% days	48.1	3,143,448.00	145.53	18,512,494.59
1% days	9.62	2,748,816.00	127.26	16,188,415.18

The calculations show that the lower 5% quantile of the spread distribution leads to bunkering money of the scrubber investment of the ferry to minimum 3 143 448,00 Euro per year. The savings of bunkering by using HFO and scrubber technology for the ferry will bring with a probability of 95% a benefit of minimal about 18 million Euro over the scrubber lifetime of 15 years compared to using MGO. A further look into the table shows that a decrease of the spread to 127, 26 Euro leads to lower saving but this situation is also related to the lower 1% quantile of the spread distribution.

The crucial point in the calculation is related to the question if the spread distribution continues to be normal distributed with the same statistical parameters over the next year so the scrubber investment is not risky. Furthermore it has to be mentioned that the calculation of the NPV depends on the two variables spread and HFO price, i.e. NPV = NPV (spread, HFO). In order to be able to calculate the NPV only on the base of the spread a linear regression has been realized. The R square fits of the model was calculated to be 93% so the underlying model enjoys a high level of explanation.

6 Conclusion

The current scholar paper validates an academic approach for assessing risk for compliance to Sulphur regulations. The developed VaR model demonstrates that historical data analysis is a practical approach for simulating investment scenarios with different situation of oil prices of the future. As well as integrating and valuing the NPV and discounted cash flow of the MGO vs HFO. All these together with Scrubber options as a popular solution. The approach allows defining the best investment opportunity and making best decision with highest return.

The model tested on the real case and shows the higher risk time of the investment on scrubber with new VaR approach. It shows if the spread of fuel on a certain amount what will be the value at risk of investment. So the model proves if we assume the price for scrubber and services will constant and the spread prices are normally distributed then the model can work for measuring the value at risk of investment.

The results generally indicate the best investment strategies with a significant value and can considered at the early stages in a new ship orders as a new investment or investment on scrubber as retrofit devices for older ships. The model also can be used in different scenarios beside of scrubber and it can be the comparison of different fuel types and evaluates the risk of investment.

Future possibility of this research can be comparison of LNG, methanol, ethanol, CNG and other types of fuel or air purification technologies as a solution to each other to find the VaR of investment with analysis of historical data of each fuel spread with other solution or technologies.

References

1. IMO. Guidance on the application of regulation 13 of MARPOL annex VI tier III requirements to dual fuel and gas-fuelled engines. MEPC.1/Circ.854. London (2015)
2. Notteboom, T.: The impact of low sulphur fuel requirements in shipping on the competitiveness of roro shipping in Northern Europe. WMU J. Marit. Aff. **10**(1), 63–95 (2010)
3. Panagakos, G.P., Stamatopoulou, E.V., Psaraftis, H.N.: The possible designation of the Mediterranean Sea as a SECA: a case study. Transp. Res. Part D Transp. Environ. **28**, 74–90 (2014)
4. Olaniyi, E.O., Viirmäe, M.: The economic impact of environmental regulations on a maritime fuel production company. Res. Econ. Bus. Central Eastern Europe **8**(2) (2016)

5. Olaniyi, E.O., Prause, G., Boyesan, J.: Economic impact of SECA regulations on clean shipping in the BSR: First empiric results from EnviSuM project (2017)
6. Basak, S., Shapiro, A.: Value-at-risk-based risk management: optimal policies and asset prices. Rev. Fin. Stud. **14**(2), 371–405 (2001)
7. Acciaro, M.: Real option analysis for environmental compliance: LNG and emission control areas. Transp. Res. Part D Transp. Environ. **28**, 41–50 (2014)
8. Daduna, J.R., Prause, G.: The Baltic Sea as a maritime highway in international multimodal transport. In: Operations Research Proceedings 2015, pp. 189–194. Springer, Cham (2017)
9. Patricksson, Ø., Erikstad, S.O.: A two-stage optimization approach for sulphur emission regulation compliance. Marit. Policy Manag. **44**(1), 94–111 (2017)
10. Bergqvist, R., Turesson, M., Weddmark, A.: Sulphur emission control areas and transport strategies-the case of Sweden and the forest industry. Eur. Transp. Res. Rev. **7**(2), 10 (2015)
11. Hämäläinen, E., Hilmola, O.P., Prause, G., Tolli, A.: Forecasting maritime logistics costs – Will the low oil price remain? In: Serry, A., Lévêque, L. (eds.) Le transport maritime à courte distance (Short-Sea Shipping), Devport 2016 Conference, pp. 19–34. EMS Geodif, Le Havre, France (2016)
12. Herbst, A.F.: Capital Asset Investment: Strategy, Tactics and Tools. Wiley, New York (2003)
13. Hull, J., White, A.: Incorporating volatility updating into the historical simulation method for value-at-risk. J. Risk **1**(1), 5–19 (1998)
14. Angelidis, T., Skiadopoulos, G.: Measuring the market risk of freight rates: a value-at-risk approach. Int. J. Theor. Appl. Finan. **11**(05), 447–469 (2008)
15. Dowd, K.: Measuring Market Risk. Wiley, Hoboken (2007)
16. Jorion, P.: Value at Risk (2000)
17. Baker, H.K., Haslem, J.A.: The impact of investor socioeconomic characteristics on risk and return preferences. J. Bus. Res. **2**(4), 469–476 (1974)
18. Farid, J.A.: Risk Frameworks and Applications. Alchemy Technologies (2010)
19. Hendricks, D.: Evaluation of Value-At-Risk Models Using Historical Data (1996)

Strategic Energy Partnership in Shipping

Eunice O. Olaniyi[1(✉)], Patrick Gerber[2], and Gunnar Prause[1]

[1] Tallinn School of Business and Governance, Tallinn University of Technology University, Tallinn, Estonia
{eunice.olaniyi,gunnar.prause}@ttu.ee
[2] EP Consulting OÜ, Jaama 12-4 11621, Tallinn, Estonia
info@epconsulting.ee

Abstract. The International Maritime Organisation (IMO) is employing a global clean shipping approach to reduce shipping emissions and to improve the Maritime's carbon footprint. One of the measures was the establishment of Sulphur Emission Control Area (SECA) in special parts of the world including Baltic Sea Region (BSR). Since 2015, ships are allowed only to use fuel with a maximal Sulphur content of 0.1% forcing ship owners to use special bunker fuel like LNG or to invest in expensive abatement technologies like the scrubbers. These are more expensive than the usual heavy fuel oil (HFO). Predictions are that oil prices may increase in which case ship-owners who have started using the LNG or the scrubber's technologies will enjoy a competitive advantage over others due to the higher margins that can further increase with additional investments into energy efficiency.

In the context of SECA, this paper tackles the research objective of how strategic energy partnerships can be adapted by the maritime sector. The research focused on the adaptation of the scrubber technology for the Maritime Energy Contracting model (MEC) using the Energy Service Contracting concept. Since the authors currently participate in the EnviSuM project, which assesses the technical efficiency and the socio-economic impact of clean shipping solutions of the SECA regulations in BSR, the research was empirically validated by expert interviews, survey results and case studies.

Results illustrated how the MEC model can be a market mechanism for the delivery of emission reduction in the maritime sector.

Keywords: Business model · Emission reduction · EnviSuM project
SECA regulation · Maritime Energy Contract

1 Introduction

The perception of maritime transport is generally seen as one of the most environmentally friendly especially when measured by weight. The conclusion is based on the premise that ships move large volumes of goods, making its emissions low when distributed per unit weight. Although much of the pollution emitted by international shipping is deposited over the sea, it is the largest single source of acidifying and atrophying fallout on land in many countries in Europe [12]. Emissions from shipping in the form

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_10

of sulphur oxides, SOx, nitrogen oxides, NOx, carbon dioxide, COx and particulate matter (PM) are significantly detrimental to the environment [2]. Furthermore, these emissions can travel long distances and Sulphur dioxide emissions especially cause acid rain and generate fine dust known as particulate matter, which is dangerous for human health as they cause respiratory and cardiovascular diseases and reduce life expectancy up to two years [11]. This is why IMO engaged a global clean shipping approach to improve the Maritime's carbon footprint, emissions of Sulphur Oxides (SOx) and particulate matter from ships by technical and operational reduction measures. One of the measures to achieve these objectives was the establishment of Sulphur Emission Control Area (SECA) in special parts of the world including Baltic Sea Region (BSR). Since 2015, ships in the BSR are only allowed to use fuel with a maximal sulphur content of 0.1% (1,000 parts per million – ppm. In addition, in 2016, the global SOx for bunker fuel was reduced to 0.5% (5,000 ppm) from 2020 [9]. This ensures that all ships globally must pay attention to the sulphur content of the fuel they use.

It has been two years of 0.1% sulphur emission regulation in selected parts of the Sulphur Emission Control Area (SECA), most speculations on the negative effect of the regulation have been wrong due to the fuel price crash and low freight rates [14]. Reliance and demand for abatement technologies such as the scrubber technology have decreased drastically because of the low costs of bunkering. Ship owners are not willing to make the risk of the investments associated with the scrubber technology. With the 2020 sulphur global, cap in view, traditional fuel companies might no longer be able to cope with the decreasing demand for HFO because their major product will no longer be marketable. Due to sulphur emission directives, maritime stakeholders have been forced to look for innovative ways of adhering to the stipulation of emission reductions from ships [21]. They are faced with 2 types of challenges that are long and short-term solutions. The long term effect is borne on strategic solutions that carry along every aspect of the shipping industry while the short term only requires meeting the SECA regulations at the stipulated time of January 2015 and 2020. There is the need for a new-market based model implementation for energy efficiency and supply especially one that is also suitable for the ship emission directive.

Already in previous work [15, 17], authors proposed activities for sulphur emissions regulation compliance that are both conventional and traditional. They suggested that companies need to use the strategic fit between their internal characteristics (strengths and weaknesses) and their external environment (opportunities and threats).

This study proposes a business model consideration for high Sulphur bunker oil producers using an Energy Saving Contract (ESC) model commonly practised by the energy saving company (ESCO) for adaptation into the maritime sector – mostly the shipping companies and medium sized fuel company using the scrubber technology. With the high uncertainties attached to the fuel prices, high edged strategies must be about great insight, experimental and evolutionary undertaking as much as the traditional skills of planning and uncompromising performance. The new model pools two goals: to ensure the compliance of the Sulphur emission directive and to lower the transaction costs that emanate from the compliance measure. The core objectives of this work are summarised as follows: To adapt the Energy Service Contract model into a suitable model for maritime industry and to demonstrate its implementation. The questions the author seek to answer are what are features of the ESC that can be modified into the maritime sector? What are the constructs to be considered in the contract? All empiric activities were executed in the frame of the “EnviSuM” project in 2016. This article is arranged in the following way: the next talks about business models, its importance and how it is the ECS can be adapted into the marine time sector. The next section discusses the MEC contract and its intricacies while the last section gives the conclusion.

2 Literature Review

2.1 Emission Abatement Technologies

The introduction of 0.1% Sulphur content regulation in SECA in 2015 requires that ship owners consider the use of bunker oil also called HFO (heavy fuel oil). To achieve the desired level of Sulphur emission, low Sulphur distillate oils such as MGO (marine gas oil) or MDO (Marine Diesel Oil) which are cleaner and more expensive with a Sulphur content of 0.1% have become the popular option [14]. Another effective way to fulfil the SECA regulation is the use of alternative source of fuel such as the Liquefied Natural Gas (LNG), methanol, and other biofuels that ensures that the ship emits very little waste after of combustion. The LNG is mostly accepted as a promising energy source for shipping because LNG ensure the compliance of the anticipated NOx emissions regulation. However, even though it is less costly when compared to distillates and HFO, the costs of distributing it to ports and ships is very high and also depends on the distance of the port from the LNG import terminals [2].

In their efforts to reduce compliance costs, the European Commission has also given support for the promotion of new technologies [9]. An example is the scrubber technology, an abatement method that maintains the Sulphur level in the exhaust fume from the ship [21] so that the ship owner is able to use the HFO and still be SECA complaint. The scrubbers are said to reduce the SO2 emission by at least 99% alongside with the PM emission. There are two major types of scrubber technology - the dry scrubber and the wet scrubber. The wet scrubber is further classified as the open loop, the closed loop and the hybrid scrubber.

The cost of scrubber installation on ships varies depending on technology and the state of the ship and ranges from 2 to 4 million € for a ship [2]. To install the scrubber on a ship there is a need for the additional stabilising fortifications because of its weight and because it has to be kept in an elevated position, all of which are additional costs [11]. Another cost related to the scrubber installation is the opportunity cost of off hiring days (4–8 weeks) during which the ship owner loses revenue. In addition, operating the scrubbers increases the rate the engine consumes fuel and is estimated to increase to about 3% or more [7]. It is more expensive to install the scrubber system on an old vessel rather than on a new vessel, and the closed system scrubber is more expensive than the open system scrubber [21]. All these factors pose a great challenge and discouragement to ship owners when considering the use of scrubber as an abatement solution for sulphur emissions.

The scrubber has a lifespan of 15 years and a payback period of 3–5 years [1]. Ross, et al. explained that the payback period of an investment determines the time required to regain the capital expended for such investment [18]. The payback period is calculated by summing up the discounted cash inflows of a number of periods in order to get the net present value (NPV) of the investment. The minimum number of periods necessary to reach an NPV greater than zero is called the payback time. If the payback period of an asset is greater than its lifespan, it means the increased purchase price will not be regained.

The calculation of the spread between the MGO and the HFO is another important factor that is used to evaluate a scrubber investment. The MGO usually has a higher price value when compared to the HFO and the difference between them is referred to as the spread value [12]. The higher the cost of the MGO the higher the spread and the higher the ship owner savings will be if HFO is used [13]. The decrease in HFO demand could also mean a drastic increase of the spread in 2020 due to reduced HFO price as speculated by [1].

2.2 Energy Service Contracting

According to Chesbrough, technology by itself has no single objective value, the economic value of a technology will remain dormant until it is commercialised in some way through a business model [6]. In some instances, companies need to use business model to expand their perspectives in order to capture value from a new technology.

This work uses the common energy contracting models popular for the energy efficiency and supply. Energy Contracting (also called ESCO) is defined as an inclusive energy service model that is used to achieve energy efficiency in a bid to optimised cycle cost [5]. The popular basic ESCO business model either provide the needed energy – Energy Supply Contracting (ESC) or ensures energy savings – Energy Performance Contracting (EPC) to the end users [20]. The ESCO takes on the responsibility of the overall delivery of the needed energy from planning, installations, distribution, operations and maintenance as well as buying of needed fuel [3].

The authors will be relating to the Energy supply contracting which focuses on energy supply. A standard ESC is measured towards supply to reduce costs of operation [5]. ESC delivers energy solutions to the need of a customer who is not interest or not knowledgeable on the technical solutions especially when it comes to maintenance. Thus, energy is provided at a reduced price [3]. Usually, the initial capital investment is free for the customers and contract period of 10–20% energy savings can run up to 10 to 15 years. Financing an ESC is a matching process that is customised for individual customers to fit their needs. Fees and other elements are tailored or adjusted for occasions such as increased risks or length of contract [8]. When the initial installation is huge or has a larger scale of risk then the customer might be asked to make an upfront investment although when this happens, such upfront costs are limited to secondary items such as parts and usually referred to as “in-house” elements [3].

3 Method

As described by Siggelkow, a case study provides a grounded real-life scenario for the audience persuasion [19]. This work is a case study that highlights how the energy contract can be adapted to the maritime business activity. With many energy servicing companies in the market, the approach to this concept started with a literature review to understand energy contracting and its success factors. For this research, interviews with ESCO practitioners with building retrofit experience were made to identify salient factors to contract and to probe their opinions on the execution of the MEC projects/contracts. The interviews were recorded and transcribed. Holistic coding was used and was based on theoretical constructions to arrange the data. In vivo coding was used to better understand the things “through the eyes of the ESC practitioners” and process coding to describe and explore the actions [10]. Osterwalder’s nine-point decomposition of a business model [16] was used to change the fuel company’s traditional business and to the new MEC business model.

4 Results

4.1 Maritime Energy Contracting Case

Producing SECA complaint fuels require high investments, so also is the investment costs for abatement technologies. Current figures indicate a decrease in scrubber installations due to low bunkering prices and low freight rates. Most ship owners do not have the financial means to embark on such huge investments nor are willing to take on the associated risks, they would rather buy the low sulphur fuel even though its price is significantly higher than the HFO commonly used for marine bunkering. If this trend continues, some traditional fuel companies will not be able to cope with the decreasing demand for HFO because their major product will no longer be competitive in the market. In order to diminish this additional business risk, the implementation of a new business model is required for high Sulphur bunker oil producers.

A radical and promising new business model for maritime fuel producers is a change from being just fuel producers towards becoming an energy service company. In this regard, the Energy Supply Contract (ESC) concept is transferred to the maritime sector to create what the authors referred to as Maritime Energy Contracting (MEC). The contextual idea is to supply the HFO to contracted ships, pre-finance the project, and run the scrubber installation in order to protect the SECA compliance. The major motivation is to lower the transaction (compliance) costs from SECA regulation compliance of both shipping company and the fuel company. The new business model of the fuel company will become “energy solutions” using the scrubber installations on ships. The fuel company implements the energy service package at its own expenses according to the project specific requirements set by the customer. For its own profit, it will receive payment for the energy (fuel) delivered, depending on the actual consumption of the ship together with the flat rate costs for service & maintenance as well as the quality assurance. The cost savings of the construction will be shared between the fuel producer and the ship owner in the course of the contract lifetime. MEC guarantees energy costs savings so that the payback from the cost savings from the supplied energy throughout the contract period will cover the investment costs and the cost of risks made by the fuel producing company.

The components for the implementation of maritime energy package outcome will be as follows: Detailed planning (Project development, rough planning, agreement, contract, Scrubber installation and start up pre-financing, operation and maintenance, troubleshooting, optimisation and user motivation. Other activities are fuel supply, scrubber monitoring and controlling, quality assurance and other benefit through outsourcing of function i.e. price guaranty and outsourcing of commercial and technical risks.

4.2 Contract, Pricing and Contract Conditions

The typical duration of energy service contract for buildings is 10 years because stationary objects like buildings do not need shorter agreements and are similar to district heat contract durations. In case of ships, this situation changes drastically, as ships are mobile assets and are easily moved around the world to other jurisdictional areas. It is therefore recommended to have periods of 3 to maximum 5 years contract with each contract customer based and adjusted periodically. Using the Energy Supply Contracting as a prototype, the Marine Energy Contracting will consist of two price components. The needed energy (fuel) supply part and the asset financing including the additional services for the agreed service time. Both components are related to formulas where the influencing factors were considered and have adjusted prices consequently. The formulas are typically updated every month. Thus, using a typical ESC [4] and from conducted interviews, the following contract calculations and assumptions are made for MEC:

Energy Supply Calculated as:

$$\mathrm{AP_{HFO}}\left[€/\mathrm{mt}\right] = \mathrm{AP_{0HFO}}\left[€/\mathrm{mt}\right] + \mathrm{FS}\left[€/\mathrm{mt}\right] - \mathrm{FS_0}\left[€/\mathrm{mt}\right], \tag{1}$$

where: $\mathrm{AP_{HFO}}$ – Working price during contract time per metric tonne of fuel (€/mt); $\mathrm{AP_{0\_HFO}}$ – Baseline price according to official statistics in the certain period €/mt; FS – Price for fuel supply per metric tonne €/mt; $\mathrm{FS_0}$ – Fuel supply baseline in a particular period (i.e. 01-06/2017) €/mt.

Non-Energy (Assets) Calculated as:

$$\mathrm{LP}\left[€/\mathrm{a}\right] = \mathrm{LP_0}\left[€/\mathrm{a}\right] \times (0.5 + 0.3 + \frac{I}{I_0} + 0.2 \times \frac{L}{L_0}), \tag{2}$$

where: LP – New price during contract time per annum [€/a]; $\mathrm{LP_0}$ – Base price according to official statistics in certain [€/a]; I_0: – Current price index for consumer goods taken as the baseline (i.e. the consumer index of common goods of the year 2015 is set as 100); I – Current price index for consumer goods comparable to the I_0 (e.g., September

2017 = 103); L – Average salary index at a certain time during the contract time; L_0 – Average salary index for setting as starting point for the contract.

This equation takes in the original costs for the non-energy related part plus inflation legalisation (e.g. higher material costs) and changes in the personnel salary. In the proposed formula, 50% of the yearly price is fixed, whereas 30% are depending on the development of general inflation (consumer good index) with 20% depending on the development of salary costs, which have a strong influence on the provided services (maintenance, monitoring) during the contract time.

Contract Terms and Conditions

I. Owner of the scrubber asset during the contractual period is the fuel company.

II. Definitions:

- Maritime Energy Contracting Price: offered comprehensive competitive technical solutions and prices regarding the functional description of the energy services. It includes the fuel price, scrubber costs and adjustment costs.
- Scrubber costs: the capital cost of the scrubber installation spread into an amortisation over the years.
- Energy price: covering the consumption of HFO at current price level. This includes consumption related only to marginal costs defined exclusively in the service contract. It will also include the risk of price surge or decline.
- Adjustments: An additional margin for running the scrubber comprising of all operational costs for the scrubber usage such as administration, maintenance, personnel, insurance and management together with entrepreneurial risk, including a profit margin. *Adjustments open a negotiation space with the contractual customer.

III. The interests' rate is determined by the market (best available offer) and will stay stable over the contractual period to avoid sudden changes within the agreement.

IV. The contractual fuel company will supply a list of "partner network" where the ship owner can bunker fuel on events where ship is not within the jurisdictions of contractual fuel company. The contractual fuel company will work out rebates or compensation with affiliated partner company.

V. In the event where energy (fuel) is gotten from another supplier other than the partner network. To protect the purpose of selling own HFO, the amounts, which had been taken out of the "partner network", have to be compensated by the ship owner. The ship owner will give access to the bunkering book or float sensor for bunker measurements to confirm compliance or deviation.

VI. The non-energy related part remains stable over the contractual period, except in the event of the aforementioned influencing factors.

VII. At the end of the contractual time, directions on continuance have these option:

- The scrubber asset is taken into the ship owner's asset list.
- A new tender is organised for the operation of the scrubber, including all services like maintenance, monitoring, optimisation, etc. and handing the ownership of the asset over to the new partner.

VIII. In the case of sudden end of contract, the following procedure will come into effect: For the starting point of the contract, the financial value of the scrubber asset (scrubber plus installation) is defined and in case of a sudden end of the contract, the shipping company will pay a financial compensation. The amount of compensation is calculated by the linear depreciation over the full contract time. Example: If for a total contracted period of 36 months, the defined value of the scrubber asset is 3.6 M €. If the ship owner after 16 months decides to end the contract and to take over the scrubbers themselves, the ship owner will pay a compensation as follows: 36 months full contract time minus 16 months of contract, which equals a remaining 20 months. For each month, the compensation is 100.000 €. Therefore, for a 20 month deviation period the sum 2.000.000 € will be paid as compensation.

IX. Other issues arising.

- Border of property: It is very important to define the borders of property, as this will have a significant impact on the responsibility. The Scrubber and connected parts will be in the responsibility of the fuel company (for the contracted period) but all other components like engines, etc. will remain in the responsibility of the ship owner.
- Space for scrubber and retrofit is given free of charge: The extra space for the scrubber and the time for the retrofit have to be free of charge to remove the complexity involved in the calculation of the non-energy related part.
- Quality of supply: The quality of the Scrubber efficiency (percentage of SOx reduction) will be defined, monitored and guaranteed.
- Additional energy consumption is business of the ship owner: The additional energy consumption with the scrubber will be added to the total costs of the scrubber in the contract calculations.

The MEC model can be described as a focal concept for strategy that creates unprecedented opportunities in the maritime sector because it is actionable and offers feasible fresh way to innovation in an uncertain, fast-moving and unpredictable environment such as the maritime sector.

The cooperative structure will ensure the following:

Environmental benefits (SOx emissions reduction). (2) Money savings on initial investment costs (scrubber installation), utility costs and maintenance cost that is taken up by the fuel company i.e. reduction in investment risks, technical risks, market risks, and performance risks, leaving only "zero risks" to the ship owner. (3) Jobs and career creation. (4) Reduced operational costs through using the much cheaper HFO, thus, the shipping company can concentrate on its core function, which is transportation, and do away with the hassle of energy efficiency through the MEC third party contracting. (5) Free technology and expertise support for the ship owner. (6) A scalable investment for the fuel company. (7) Promises to have a higher margin for the fuel production company compared to the traditional HFO supply approach. (8) Customer fitted model i.e. customised contracts with the ship owners.

So far, even though the BSR has witnessed commendable compliance activities, the success level is far from satisfactory, especially as it relates to the heavy and risky investments the maritime stakeholders are subjected to. Another challenge is the uncertainties that surround the use of the each compliance method. Unfortunately, the reduced fuel prices have made many investments meaningless and wasteful. With the upcoming 2020 global restrictions, the energy consumption will increase as the demand level for the low fuel oil. The maritime industry may yet experience another game changer, which might force an increment in the price of the fuel. However, beyond that, there might be a shortage in the low sulphur fuel supply.

5 Conclusion

The authors put forward the MEC model as a market mechanism for the delivery of emission reduction in the maritime sector by using the scrubber technology to cushion the economic effect of the SECA regulations and illustrated modified features of the ESC model and construct to be considered in the Maritime Energy Contract model. Unlike the usual conventional strategies, business model concepts accentuate analysis and strategies embedded in experimentation as projected by the authors. With appropriate state enabling policies, the energy contracting model – a proven and resilient structure can boost the maritime industry by shifting the focus of the scrubber technology towards selling “energy solutions”, and jump-start a viable private-sector that targets the maritime stakeholders. In this light, the scrubber technology can provide a much-needed technology-push-effect for the European technology. The main disruptive advantages are based on the shift from CAPEX to OPEX (no direct investment for ship owners is needed) and the establishment of new services, with additional value in the maritime sector.

The work is limited to the use of only HFO/MGO and not to other sources of fuel like the LNG. The LNG approach can be an interesting angle for further research. Also, due to the scope of this work, the authors were unable to show other contracts that will involve the “partnering network” such as the scrubber manufacturer, other fuel company, maintenance and/or a financing house. This can also be a consideration for future research.

References

1. Atari, S., Prause, G.: Risk assessment of emission abatement technologies for clean shipping. In: 17th International Multi-Conference ‘‘Reliability and Statistics in Transportation and Communication’’ Riga (2017, forthcoming)
2. Bergqvist, R., Turesson, M., Weddmark, A.: Sulphur emission control areas and transport strategies – The case of Sweden and the forest industry. Eur. Transp. Res. Rev. **7**(2) (2015)
3. Bertoldi, P., Rezessy, S., Vine, E.: Energy service companies in European countries: current status and a strategy to foster their development. Energy Policy **34**(14), 1818–1832 (2006)
4. BFEE Homepage on Energy Contracting. http://bfeeonline.de/bfee/informationsangebote/publikationen/sonstige_publikationen/vfw_mustervertrag_waermelieferung_2012.pdf. Accessed 10 May 2017
5. Bleyl, J.W.: Conservation First! The New Integrated Energy-Contracting Model to Combine Energy Efficiency and Renewable Supply in Large Buildings and Industry. ECEEE summer Studies, Paper ID, 485 (2011)

6. Chesbrough, H.: Business model innovation: opportunities and barriers. Long Range Plann. **43**(2), 354–363 (2011)
7. EMSA Homepage: European Maritime safety Agency Annual 2010 Report. http://www.emsa.europa.eu/emsa-documents/download/1421/143/23.html. Accessed 10 May 2016
8. Goldman, C.A., Hopper, N.C., Osborn, J.G.: Review of US ESCO industry market trends: an empirical analysis of project data. Energy policy **33**(3), 387–405 (2005)
9. IMO Home page: IMO sets 2020 date for ships to comply with low sulphur fuel oil requirement. Press briefing release 28/10/2016. http://www.imo.org/en/MediaCentre/PressBriefings/Pages/MEPC-70-2020sulphur.aspx. Accessed 30 Nov 2016
10. Miles, M.B., Huberman, A.M., Saldana, J.: Qualitative data analysis: A sourcebook. Beverly Hills, Thousand Oaks (1984)
11. Notteboom, T.: The impact of low sulphur fuel requirements in shipping on the competitiveness of roro shipping in Northern Europe. WMU J. Marit. Aff. **10**(1), 63–95 (2010)
12. Nugraha, F.: Effective implementation of emission control area towards cleaner shipping operations: focusing on sulphur oxides (SOx) emission reduction. World Maritime University Dissertations (2009)
13. OECD/ITF Home page on Reducing Sulphur Emissions from Ships: The Impact of International Regulation. http://www.itf-oecd.org/reducing-sulphur-emissions-ships-impact-international-regulation. Accessed 10 June 2016
14. Olaniyi, E., Prause, G., Boyesen, J.: Economic impact of SECA regulations on clean shipping in the BSR- first empiric results from EnviSuM project. In: 2017 International Conference on Maritime Energy Management Proceedings. World Maritime University, Malmo, Sweden (2017)
15. Olaniyi, E.O., Viirmäe, M.: The economic impact of environmental regulations on a maritime fuel production company. Res. Econ. Bus. Central Eastern Europe **8**(2), 58–84 (2016)
16. Osterwalder, A.: The Business Model Ontology-a proposition in a design science approach. University of Lausanne (Doctoral Thesis) (2004)
17. Prause, G., Olaniyi, E.O.: The impact of environmental regulations on the regional development in Eastern Estonia. In: Muravska, T. (ed.) New Challenges of Economic and Business Development Conference, Riga, 2017, Riga. Latvian State University (2017)
18. Ross, S.A., Westerfield, R., Jaffe, J.F., Roberts, G.S.: Corporate Finance, vol. 7. McGraw-Hill/Irwin, New York (2002)
19. Siggelkow, N.: Persuasion with case studies. Acad. Manag. J. **50**(1), 20–24 (2007)
20. Sorrell, S.: The economics of energy service contracts. Energy Policy **35**(1), 507–521 (2007)
21. Wiśnicki, B., Czermański, E., Droździecki, S., Matczak, M., Spangenberg, E.: Sulphur Regulation–technology solutions and economic consequences for the BSR. Shipping market Institute of Maritime Transport and Seaborne Trade, University of Gdansk (2004)

Smart and Sustainable Cross-Sectoral Stakeholder Integration into Macro-Regional LNG Value Chain

Laima Gerlitz[1,2], Robert Philipp[1,2(✉)], and Anatoli Beifert[1]

[1] Department of Business Administration, Wismar Business School, Hochschule Wismar, University of Applied Sciences: Technology, Business and Design, Philipp-Müller-Str. 14, Wismar, Germany
{laima.gerlitz, robert.philipp, anatoli.beifert}@hs-wismar.de

[2] Tallinn University of Technology, Ehitajate tee 5, Tallinn, Estonia

Abstract. Since the introduction of lower sulphur content with no more than 0.10% in the Sulphur Emission Control Areas (SECAs) from 1 January 2015, the Baltic Sea Region (BSR) is increasingly becoming subject to search for new economically and environmentally competitive and survival strategies. The BSR stands for a flagship maritime region in Europe in terms of good economic, social and environmental performance. In order to sustain, and much more important, to improve the overall eco-system performance of the BSR, Liquefied Natural Gas (LNG) has been recognised as one of transitional measures, alternative strategies and business opportunities in maritime shipping and the entire transportation and energy system. LNG might become a viable stepping-stone alternative solution for business, and be considered as a regulation-driven demand to comply with environmental regulations that aim to achieve the goals set by 2020–2050. As a response, 'Go LNG' ERDF part-financed INTERREG V project aims at reducing technological, knowledge and business gaps by providing operational and strategic approach. In line with project, the present study aims to answer the research question how to integrate LNG stakeholders involved in diverse LNG activities, with different needs, capacities and capabilities into one macro-regional transport and energy supply and value chain.

Keywords: LNG · Value chain · Alternative fuel · Strategy · Cluster

1 Introduction

Affected by the establishment of the SECAs by the International Maritime Organisation, the BSR is once again pushed into the scene of economic discourses with need for ecologically competitive and survival strategies [15, 36]. The BSR itself represents a modal maritime region in Europe, which is associated with sound economic, social and environmental achievements. Especially, the transport sector is one of the most important engines of the BSRs' economy [8]. The business structures of the BSR are primarily characterised by SMEs, playing a decisive role in logistics and business innovation networks [7, 20]. Paradoxically, an economic wellbeing and good environmental

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_11

performance has been jeopardised due to intense shipping practices, ship operations and services, which, in turn, resulted from intensifying globalisation, trade and transport interactions. Echoing the International Transport Forum at the OECD [24], waterborne transport will grow with 327% by 2050, thus producing 238% more CO_2 emissions. In Europe, freight volumes will increase by 2050 by 216% with 174% CO_2 emissions, respectively. In this context, LNG is regarded as one of the most promising alternative strategies and business opportunities in maritime shipping as well as the entire transportation and energy system. Accordingly, LNG issues had been recently researched from the clean shipping [37] and strategic development perspective for a LNG powered transport corridor in the BSR [31].

Yet, no recent studies record approaches on LNG use as opportunity and value proposition leading towards emerging value chains that integrate different transport modes and business sectors, and where LNG is recognised as value proposition for all stakeholders involved. This bears a clear research gap that needs to be closed. Therefore, the 'Go LNG' ERDF part-financed INTERREG V project has set the ambitious goal to diminish technological, knowledge and business gaps by making operational and strategic approach available. In accordance with the project, the aim of this study is by setting up a macro-regional transport and energy value chain that combines different modes of transport, various technological options and customers with different necessities in a shrewd and sustainable manner, to create an integrated and sustainable macro-regional LNG transport and energy value chain.

The present study is organised as follows: The subsequent chapter deals with the usage of LNG as an alternative fuel, whereby the common LNG value chain is described and the potentials of inherent business opportunities are mentioned. Section three presents the methodology, followed by chapter four, which reveal the concept for the cross-sectoral LNG value chain integration. One step further (fifth part), a first snapshot on strategic cross-sectoral LNG value chain integration in the BSR is outlined. In the penultimate section, the results are discussed, whereby the paper ends with a conclusion inclusive some implications.

2 Literature Review

The usage of LNG as an alternative fuel for ships has been investigated by a number of researchers. This is deeply rooted in the introduction of SECAs from the 1 January of 2015. In this context, especially the environmental advantages have been stressed [47, 48, 51]. In comparison to conventional fuels for ships (e.g. HFO and MDO), the switchover to LNG can decrease the emissions of nitrogen oxide (NOx) up to 85 to 90%, carbon dioxide (CO_2) up to 15 to 20% and sulphur oxide (SOx) as well as particle matter (PM) almost completely [4, 38]. Besides such considerations, LNG is also proposed as an alternative fuel solution for road freight transport in order to reduce greenhouse gas (GHG) emissions. Related to this are examinations of LNG in comparison to conventional and alternative fuels. For example, comparative studies were conducted by Beer et al. [6]; Cheenkachorn et al. [11]; Kumar et al. [30]. They all arrived to more or less similar results, namely that the use of LNG needs to be promoted even more intensively as it has various environmentally friendly properties. Nevertheless, it can be stated that the usage of LNG as a clean fuel is still on a moderate level, whereby demand and thus the international trade is likely to increase in the future [22, 29, 50]. This is also the reason why LNG "is often considered the form of energy that will be the 'bridging fuel' to a sustainable energy system, sometime after 2050" [28, p. 4097].

In order to keep pace with the rapid development of international trade of LNG, stakeholder have to invest in the LNG value chain, which consist in general and simplified of five large-scale, complex and different elements with diverse inherent operations, which are very strongly interlinked and interdependent [19]. The LNG value chain is often described through the segments of exploration and production, liquefaction, transport, regasification and distribution [9, 19, 44]. In general, it can be pointed out that the amount of academic studies with the specific perspective on the holistic LNG value chain(s) is rather rare, whereas the following selected publications should provide a brief overview, and for this reason, just like before, makes no claim to completeness. Worth mentioning is the report of Foss [19], who gave a review on the reduction of the overall LNG value chain costs in the past primarily related to the US market, which is generally based on the technology advantages in each segment, while the increased demand for LNG paved the way for this development. Promising and extensive academic studies of LNG value chains worldwide based on 85 projects were carried out by Rüster and Neumann [44]. The authors conducted empirical analysis of companies' driving factors through vertical integration. In general, they claim that high transaction costs along the LNG value chain are the main reasons for a higher degree of vertical integration. However, it might be stated that all these studies have been focusing mostly on LNG as an alternative fuel, in which the specific perspective on the entire LNG value chain has been studied less thoroughly. Further, the authors argue that the relevant academic literature paid not enough attention to the possibilities of an efficient integration of the stakeholders of the LNG value chain and the resulting economic benefit of all.

The visualisation and identification of LNG related business opportunities and the transfer to real business activities generally call for new infrastructure investments on demand and supply-side. Linked to this are challenges which primarily can be seen in the uncertainty aspects of stakeholders' decision making (DM) [1]. DM in the case of LNG projects is a very difficult procedure. This is deeply rooted in the potential high number of players, who have to be considered, natural high investment needs of LNG projects and the inherent long term capital return [10]. Castillo and Dorao [10] identified this challenge and developed a DM concept for LNG related projects. Thereby, multiple parts of the LNG value chain are considered. The authors of the present paper argue that this is essential as stakeholders of the LNG value chain have to work together or at least pay regard to the other participants during their own DM in terms of exploration of business opportunities, as the usage of LNG or the investment in related projects by a stakeholder, in turn, creates new business opportunities for the stakeholder and other related (potential) stakeholders of the LNG value chain. In the face of this premise, Gerlitz and Paulauskas [22] assume that the value and advantage of the use of LNG, provided that the infrastructure is established in the port environment, would be added to the stakeholders in the immediate vicinity of the port in question, as a multiplier effect for new business opportunities. In their opinion, the most promising resulting scenarios would be the usage of LNG for additional port activities with regard to vehicles, cranes and machinery, as well as the construction of linkages to filling stations for trucks, busses and municipal service vehicles. Furthermore, the usage of LNG for power supply generates additional business opportunities. Therefore, the construction of a LNG terminal in a port could be seen as one of the major measure for the development of well-needed LNG infrastructure, as it will have a positive influence on LNG related business opportunities as well as on the creation of new job opportunities [34]. To exemplify, the building up of the LNG terminal in 2014 in Klaipeda (Lithuania) had a decreasing effect on the NG delivery price. This remarkable advantage fostered the demand of new LNG related projects and allowed new LNG stakeholders to come into the LNG-market to use LNG as a fuel in diverse transport modes as well as in the energy sector, which, in turns, had a positive influence on the NG prices in the neighbouring regions. As a result, LNG has to be emphasised economically, as a rise of LNG related operations and thus closely linked new stakeholders facilitate the access of the SBSR into the international LNG market and reciprocal lead to an increase of growth and innovations as well as employability, which on the other hand contribute to the EU Blue and Green Growth as well as anyway to the Clean Fuel Strategy [22]. Thus, it can be expected, that, due to the establishment of further LNG terminals, numerous business opportunities arise which are connected with alternative gas supply possibilities, decrease of LNG prices for the end users, synergy effects between the involved players and the development of new services (Ibid).

To conclude, considerable LNG related investment projects by all stakeholders are needed to establish an integrated LNG infrastructure and in order to generate a benefit for all participants of the LNG value chain, these stakeholders have to cooperate during DM, for the issue, how to change to a long-term market; reciprocal, the LNG market will profit by additional stakeholders [46]. To foster this promising development, the authors of the present paper advocate the development of a concept to facilitate a better integration and market entrance of stakeholders, and thus to create a greater value for all.

3 Methodology

In the framework of the given study, the theory based and practice related research have been applied built upon comprising qualitative expert interviews, surveys and practical findings that have been originally collected and produced in the framework of the following projects: (1) The "MarTech LNG – Marine Competence, Technology and Knowledge Transfer for LNG in the South Baltic Sea Region"; and (2) "Go LNG". Both projects are being implemented in the framework of the INTERREG IVA South Baltic and INTERREG VB Baltic Sea Region programmes. Among other things the projects focus on the implementation of the EU Clean Fuel Strategy and the EU

Directive on Deployment of Alternative Fuel Infrastructure through technology and knowledge transfer in LNG related business activities and improvement of the LNG value chain in the Baltic Sea Region.

Apart from the systematic literature review, analysis and study of relevant theories and concepts, relevant policy regulations and guidelines, the research findings demonstrated here have been mainly based on primary and qualitative data collected directly by the projects' partners, associated organisations and project interest groups. The primary empirical data sources were gained in form of evidence-based observations (here: case studies), empirical data from quantitative surveys and qualitative expert interviews with the involved project experts, researchers and relevant stakeholders, observations gathered from respective project activities such as workshops, conferences, round table discussions and open LNG thematic conferences with relevant stakeholders. Furthermore, gained research findings have been validated and verified by the main project target groups during practical workshops and targeted seminars. The main target groups include: policy makers that are responsible for the ports' and energy infrastructure development; ports' and terminals' operators, incl. cargo handling companies; international associations and corporations involved in the LNG value chain; shipping companies, ship building yards; relevant academic and research institutions as well as regional industries that might benefit from governmental investments and higher energy security, herewith increasing LNG and clean fuel usability.

4 Conceptualisation of Strategic Cross-Sectoral LNG Value Chain Integration

The current research aimed at answering the question on how to integrate LNG stakeholders involved in diverse LNG activities, with different needs, capacities and capabilities into one macro-regional transport and energy supply and value chain. A result thereof is a first-step grassroots conceptual and strategic approach that shall facilitate an integration of regional LNG value chains in the BSR into one macro-regional and multidimensional LNG value chain. The research claims to reduce the research gap by adding a qualitative theoretical contribution to the applicable research province. By using both deductive and inductive reasoning and reflecting upon diverse LNG practices in the BSR, which are discussed below using systematic ecosystem view, the authors propose a holistic ecosystem-based strategic approach embodied in the regional performance domain of LNG actors and stakeholders. This approach, which touches upon infrastructural (technological), economic, environmental, social, cultural and policy dimensions, becomes essential when taking into account multiple needs to respond to current and future challenges as well as to reduce the rapid pace of uncertainty.

In particular, the authors claim that an efficient integration of LNG activities on a macro-regional scale, i.e. activities that merge resources, capabilities, industry interactions, company performance and organisational operations from individual BSR regions and thus the Member States surrounding the Baltic Sea require a two-prong strategic approach. On the one hand, it is argued that an efficient an effective integration presupposes a common challenge that can be solved once it is shared by the entire ecosystem, which encapsulates diverse industry and social actors and activities. This, in turn, enables to achieve a shared value, once it is reduced or eliminated on the market, i.e. in the entire community. On the other hand, integration can be facilitated by efficiently merging business innovation and governance dimensions from individual regions into one macro-regional strategic approach rooted in key treatise of regional integration [32] and regional innovation [14]. Based on Cooke [13] and Holbrook and Salazar [23], the business innovation dimension addresses behaviour of market and industry players and entrepreneurial conditions, such as (a) presence of enterprises and industry actors on the market, incl. market and industry structure and domination; (b) research reach; (c) public and private R&D landscape; and (d) level of collaboration, associationalism among market actors. The governance dimension encapsulates such indicators, as (e) source and level of initiation; (f) funding schemes; (g) research landscape; (h) technical specialisation, and (i) scope and scale of coordination).

As a result, a systemic approach towards regional setting in the nexus of innovation and integration can be associated with the theory of location and clusters, as coined by Delgado et al. [16]; Krugman [27, 28] and Porter [41, 42] and that shape the strategy based on competitive forces. Regional Innovation Systems (RIS) or Regional Systems of Innovation (RSI) and increasing role on location has slowly replaced the National Innovation System (NIS) [2, 3, 13, 17]. Such an approach let us emphasise the regional dimension of innovation, which encapsulates both top-down (external) and bottom-up (internal) sets of characteristics that shape innovation emergence and its management. Externally, it refers to the role of institutions, crucial for the knowledge creation on the local and regional level and governance of innovations, whereas internally it includes internal characteristics of interaction and collaboration among different actors of an ecosystem. In aggregate, within this perspective, focus is given to issues, components, and processes in a system of innovation that operates at a localised and regional level. Thus, the perception RIS underpins the emphasis of spatial proximity and agglomeration in the dynamics of innovation and economic growth [16, p. 257]. This, in turn, allows us bridging bridge innovation, competitiveness and growth within the regional context.

As a result, two perspectives are twins of strategic orientation and build the ground for the value chain integration. The theoretical mainstay of the two-fold strategic approach refers on the one hand to the external perception – the market-based view, which has its roots in the Organisational Theory (OI) and particularly Porter's Five Forces Model [39], accompanied by related concepts that share the strategic manner rooted in the environment, e.g. competitive advantage and business strategy perception [40]. On the other hand, the theoretical foundation is underpinned by incorporation of internal strategic strengths associated with internal resources [5], dynamism and dynamic capabilities [18, 45], learning and tenets of tacit knowledge [12, 26].

Bearing in mind the theoretical considerations and partnerships among adopted concepts that share the same or similar meaning and sense-making, the proposed strategic framework consists of four key building blocks that might be referred to as regional performance domains: (1) Knowledge & Skills; (2) Infrastructure & Technology; (3) Economy & Business; (4) Environment & Governance (rf. Fig. 1). These ones cover both external and internal organisational and thus regional dimensions and merges them. In addition, these four need to intertwine in the frame of regional interactions in order to achieve a certain degree of integration. Here, the supply side, which is expressed by existence and number of knowledge and skills related to LNG as well as existing organisational and governance structures, availability of applicable policies and framework conditions, compliance with diverse dimensions of sustainability and organisational learning, availability of infrastructure and technology, financial regimes and schemes, innovation platforms need to be merged with the demand side – the environmental, social and economic compliance of the region with national and international regulations, entrepreneurial needs and demand drivers – business growth perspectives, opportunity recognition, uncertainly reduction, capacity building.

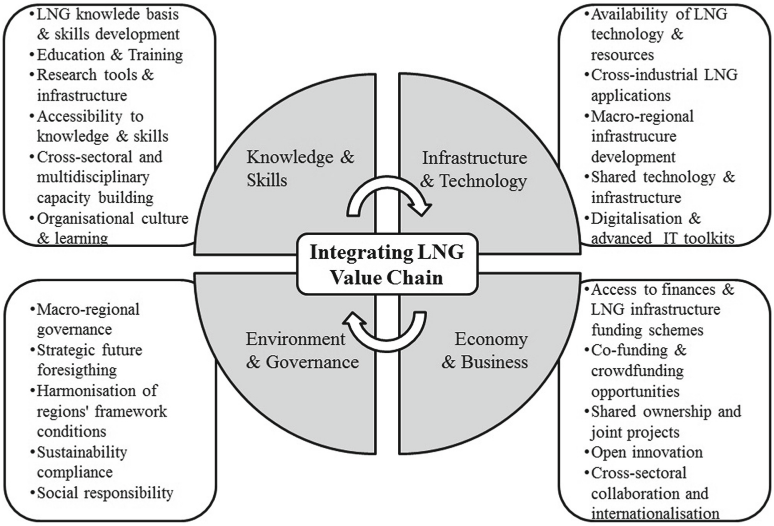


Fig. 1. Strategic ecosystem approach for macro-regional and cross-sectoral LNG value chain integration (source: own illustration, [21]).

This, in turn, can lead towards increased innovation potential, and, respectively, to better competitive edge and growth perspectives, since innovation is key to both – competitiveness and growth. In this nexus, the market structure and behaviours of market participants need to be studied at a first glance. Next to this, what is increasingly becoming important when taking into account new market trends and growing demand to respond to these market trends by industry actors and companies is associated with the focus on internal organisational resources and capabilities, organisational culture and learning and adoption of knowledge and skills to new changing paradigm of social and economic performance. In this sense, we claim that it is essential to merge external and internal provinces of performance that both affect interactions in organisations and regions. Both are subject to impact by externalities as well as internalities. As coined by Wernerfelt [49], it is worth to recall the duality between markets and resources (p. 172). In addition, it is also crucial to merge bottom-up and top-down approaches and involve both – governance actors, suppliers and buyers – for a common macro-regional approach. This is essential in today's regional setting. Also the region is well performing today, it must strategically prepare for the future and be able to adopt to rapid changes. As stated by Mohapatra [35], "individual forces and their collective impact will change as the government policies and macroeconomic and environment conditions change" (p. 274). Indeed, the process of integration can be facilitated by fulfilling or providing with conditions that are plotted in the scoreboard of each individual performance domain. These indicators might serve as guiding parameters for all regional stakeholders that are involved in LNG activities to a different degree.

5 First Snapshot on Strategic Cross-Sectoral LNG Value Chain Integration in the BSR

According to the 2016 State of the Region Report 'The Top of Europe – Doing Well Today, Feeling Worried About Tomorrow', the BSR stands for a flagship maritime region in Europe in terms of good economic, social and environmental performance indicators [25, 33]. Yet, the social and economic integration of individual regions surrounding the Baltic Sea, as well as integration of LNG markets into one macro-regional LNG market, and thus, supply and value chain is still being hampered. The same applies for the cross-sectoral industry collaboration and joint involvement concerning LNG utilisation in the region along the horizontal industry supply and value chains. Although the region demonstrates outstanding performance in specific LNG related areas, e.g. knowledge and skills as well as LNG applications in diverse markets, (e.g. Norway, Sweden), the level of performance is still not fully dispersed to other involved individual regions. As a result, the region lacks cohesion in terms of sustainable and balanced development and integration of LNG. The rationale behind this situation can be traced back once applying the proposed strategic approach and using it systematically throughout the regional LNG supply and value chain performance analysis. The BSR LNG supply and value chain includes such segments, as LNG liquefaction, import terminals, LNG shipping, LNG shipbuilding, LNG onshore infrastructure (bunkering, filling and distribution) and LNG end-user solutions (e.g. LNG powered trucks, public transportation, containers for railways, etc.). The discourse below briefly overviews recent key development patterns within all domains along the BSR LNG value chain by shortly pinpointing key strengths and weak points. These are also summarised in the Table 1 below.

Table 1. Mapped integration of LNG value chain integration in the BSR as of 2017 (source: own illustration, [21])

Domain of LNG value chain integration								
Segment of BSR LNG supply and value chain	Segment specification	Knowledge & Skills	Infrastructure & Technology	Economy & Business	Environment & Governance	Location	Evaluation criteria	
Shipping	LNG feeder vessels	+++	+	0	0	SE	Existing	+++
	LNG bunker vessels	+++	++	++	++	SE	Developing	++
	Ship-to-ship bunkering (STS)	+++	++	++	++	BSR	Projected	+
LNG terminals	LNG import terminal	+++	++	++	++	LT, NO, PL, SE	Planned	0
LNG onshore infrastructure	Small-scale export/bunker facilities	+++	+	+	+	DK, LT, PL, SE	Missing	–
	LNG bunker stations	++	0	–	0	DK, NO, SE		
	LNG filling stations	++	++	–	0	DE		
	LNG fuel tank containers	+++	+++	++	++	BSR		
	LNG trucks	+++	+	+	+	DE, LT, PL, SE		
End-users	Tank & bunkering solutions	+++	++	++	++	BSR		
	LNG trucks for roads	+++	++	–	++	LT		
	LNG-fuelled buses for public transport	+++	++	–	++	NO, PL		
	LNG power supply	++	–	–	–	DE, NO, SE		
	LNG ferries	+++	+++	++	++	DE, DK, NO, PL, SE		
	LNG tank containers for railways	++	0	–	–	LT		

Taking into account *Domain 1 – Knowledge and Skills*, the macro-region records diversified and distributed knowledge and skills portfolio provided by the supplying institutions (education, training and research) as well as differing levels of knowledge and skills that are absorbed by companies, entrepreneurs and other demanding actors. All member states and therefore regions of the BSR that are involved in the 'Go LNG' INTERREG project aiming at developing integrated LNG value chain approach – Denmark, Estonia, Germany, Lithuania, Poland, Sweden and Norway – demonstrate strong or substantial developing LNG related knowledge and skills. More specifically, LNG relevant skills exist in all participating regions, thus enabling cohesive development of LNG relevant capacity. Yet, the degree and quality of knowledge do vary, what, in turn, increased knowledge transfer and absorption opportunities among the involved actors. This is especially true with the knowledge and skills transfer from Scandinavian regions, especially Norway, to the Baltics and Poland. The reason behind this might be traced back to the fact of differing organisational culture bound to utilised top-down governance approach for decades and thus subjected to path dependency and thereof resulting development and performance patterns. Furthermore, the knowledge and skills from Norway, Germany and Scandinavia shows high degree of concentration and clustering, what might be the reason for state-of-the-art knowledge accumulation in these individual regions and higher demand for this knowledge and skills outside them. Yet, the regions, especially in the Baltics and Poland should increase organisational learning capabilities and knowledge sharing among involved LNG actors and stakeholders, thus increasing multiplier effects off knowledge diffusion and balanced development.

With regard to *Domain 2 – Infrastructure & Technology*, the macro-region shows rather unbalanced development of LNG infrastructure. The majority of the existing large-scale LNG technology and infrastructure concentrates in Norway, Lithuania, Poland and Sweden. Other regions possess rather flexible infrastructure (e.g. Samsø in Denmark) and LNG equipment or have not yet declared any LNG infrastructure development projects. The existing LNG infrastructure is limited to LNG ship demonstrations (Rostock and Tallinn) or small-scale LNG bunker ships (Denmark, Lithuania, Germany, e.g. hybrid LNG barges and ferries). This is rather a paradoxical development taking into account the EU directives and regulations that make all the players bound to their compliance in the region. This might imply that infrastructure projects will be mushrooming from 2020 onwards in order to comply with the regulation setting or induced by technological transformation (e.g. digitalisation, rf. [21]), or the infrastructure will be shared among the individual regions by benefiting from the new technological trends or stronger development and usage of rather mobile and small-scale infrastructure, such as trucks, public transportation, small-scale ships and bunker and filling stations. Interesting to note is also a rather limited utilisation of potential synergy effects from the not equally distributed LNG infrastructure on the macro-regional scale, which, once utilised, could lead towards greater economic interactions among different actors, what, in turn, would increase demand for LNG, strengthen competitiveness, open up new growth perspectives and contribute to the macro-regional integration. As the examples of Lithuania and Poland show, establishment of LNG infrastructure in regional sea or inland ports (e.g. LNG import or distribution terminals) would allow a breakthrough of LNG development, by attracting other stakeholders, projects, initiatives to the established infrastructure and circled LNG clusters. Beyond this, individual regions do further advocate the development of storage and bunkering stations as well as gas pipeline systems and possibility of onshore power supply.

Taking into account *Domain 3 – Economy & Business*, the macro-regional landscape can be treated as being rather scattered in terms of efficient and effective entrepreneurial discoveries and intensive business interactions. The reason behind a lower level of economic operations with LNG might be referred to higher investments costs, missing funding and support schemes. Here, the state support, and thus, the governance level involvement, plays a crucial role, as this has been the case in, e.g. Lithuania and Poland, where the governance has supported the development of LNG infrastructure. In other regions, e.g. Denmark or Germany, companies and entrepreneurs are claiming about the missing support schemes for LNG evolvement. From the supply and competence side, the macro-region is represented by the highest proportion of the stakeholders and players, where most of them are involved into maritime-related activities, i.e. shipping, ship repair and construction as well as gas and oil supply. Furthermore, when it comes to the assessment of the stakeholders and players, it can be argued that also the BSR in general has available stakeholders and players relevant for the LNG supply and value chain, the bottlenecks tend to lie in LNG itself as a primary resource and its supply or export. On the contrary, when it comes to those components of the LNG supply and value chain that refer to LNG shipping, potential locations for LNG importing (terminals), regasification, its storage, distribution and marketing, the region seems to reveal a sound potential for the utilisation of LNG in the future. From the data gathered it is apparent that there are actors capable of taking over the particular LNG activities along the entire LNG supply and value chain. It is claimed here that businesses and entrepreneurs, also with support of governance institutions, should start utilising diverse business models that appear to be very successful in other industry applications, e.g. open innovation, crowdfunding schemes or similar. Shared ownership and joint investment into the infrastructure crossing regional boundaries could bear also a feasible option and business opportunity.

Finally, looping the *Domain 4 – Environment & Governance* to the overall LNG value chain integration within the BSR, the researchers claim that this province needs improvement and a more common approach. According to the data gathered, it can be observed that governance in terms of LNG macro-regional development is still not interconnected from the individual regions' perspective. Authorities and public institutions act rather isolated what, in turn, causes additional time, investment resources, lower flexibility ability to adopt to changes, lower trust and inter-institutional collaboration. This, in turn, implies less integration with the compliance of EU regulations crossing the borders and lower achievements of better environmental status, social, economic, financial and environmental sustainability. Hence, there is a need for a macro-regional governance approach and stronger focus on future foresight. The Go LNG project that can be regarded itself as a macro-regional initiative addresses this challenges and points to the missing links. Yet, stronger involvement and presence of authorities and other public institutions that govern LNG development need to be gathered together. Authorities and institutions should also facilitate not only the regulation-driven response of introducing LNG as a need to comply with regulations, but should rather facilitate recognition of LNG as a business opportunity, entrepreneurial discovery and facilitate real-life creation of market conditions that would enable diversification and multiplication of LNG activities. In the BSR, which is bound to specific compliance with environmental and transport regulations, LNG enables reduction of negative environmental footprint and provides transport, mobility and energy stakeholders with a resource that enables to comply with regulations. This contributes also to economic, environmental and social sustainability to continue regional transformation towards a more sustainable, innovative smartly developing and growing region.

6 Discussion

The proposed approach builds upon the conceptual partnerships that share a common denominator – value generation. Value delivery to and capturing by diverse stakeholders on the regional scale facilitate social and economic integration and strengthens strategic positioning of both organisations and the region itself. Respectively, taking the theoretical concepts of innovation and integration as two grounding conceptual approaches adopted to the domain of geographically bound and proximate interactions – the Baltic Sea Region – the existing conceptual partnership of shared value creation and value chains [43]. In sum, these all implying value proposition should be wrapped up within the regional integration, innovation and ecosystem perspective, where individual regional interactions, especially those emerging not only along vertical, but rather along horizontal (cross-sectoral) supply and value chains, might lead to macro-regional value creation allowing stronger innovations and more crucial competitive edge of the BSR. Rüster and Neumann [44] claim that vertical integration and strategic partnerships lead to a LNG market development, where large players by their current dominating role in the LNG value chain may limit competition at the horizontal level and therefore liberalization efforts in downstream markets may be hindered. Thus, apart from the quantitative validation and verification of the gained results, in the future course of the research a stronger attention may be paid to the efficient integration also of the smaller stakeholders (here: SMEs) into LNG value chain.

7 Concluding Observations and Implications

A macro-regional integration, development and innovation require comprehensive understanding of economic, environmental, social, cultural and policy (governance) systems. Therefore, solving a particular challenge, problem or turning opportunity into business, value generation and value capturing is subject to combination of diverse theoretical approaches and concepts. Yet, similarly, as with the strategy formulation and strategic management, a problem or challenge oriented solution requires also understanding of ecosystem's performance. In this regard, economic interactions are reflected in line with environmental and social arrays, which, in turn, are affected by certain policy regulations or outcomes thereof. For this and in order to integrate all affected stakeholders from the BSR into one LNG value chain, the ecosystem perspective, which merges economic, environmental, social, cultural and governance perspective is likely to be feasible. Integration of stakeholders from involved sectors and clusters in terms of LNG is supported by embodied strategic perception: how value is generated in the chain and can be captured from using LNG. LNG is exploited as a resource, capability and capacity, competitive advantage, value creator and thus innovation enabler and output. A regional setting (BSR) places special framework conditions to be addressed in terms of integration and innovation, thus understanding the BSR as Regional Innovation System and cluster with emerging clustered and smart specialising industries of water transport, land-based mobility, energy and power generation and environmental compliance. For this, it is necessary to know under which economic, environmental, social and governance conditions the integration takes places and LNG value chain is emerging.

References

1. Alhosalo, M., Karppinen, A.: Kaasua Satakunta, Selvitys nesteytetyn maakaasun (LNG) liiketoimintamahdollisuuksista Satakunnassa. A report on the business possibilities of liquified natural gas in Satakunta region. Turun yliopiston kauppakorkeakoulun Porin yksikön julkaisusarja A, nro A43/2013 (2013)
2. Andersson, G.: Rethinking regional innovation. Syst. Pract. Action Res. **26**(1), 99–110 (2013)
3. Asheim, B.T., Coenen, L.: Knowledge bases and regional innovation systems: comparing Nordic clusters. Res. Policy **34**(8), 1173–1190 (2005)
4. Bagniewski, M.: LNG Fuel for Ships: A Chance or a Must?. Det Norske Veritas. Gdynia, Poland (2010)
5. Barney, J.: Firm resources and sustained competitive advantage. J. Manag. **17**(1), 99–120 (1991)
6. Beer, T., Grant, T., Williams, D., Watson, H.: Fuel-cycle greenhouse gas emissions from alternative fuels in Australian heavy vehicles. Atmos. Environ. **36**(4), 753–763 (2002)
7. Beifert, A., Prause, G.: Logistic innovation networks in the Southern Baltic Sea Region. In: Baltic Business and Socio-Economic Development 2009, 5th International Conference-Kalmar, Sweden, September 14–15, 2009, vol. 3, p. 3. BWV Verlag (2011)
8. Beifert, A., Gerlitz, L., Prause, G.: Sustainable business development models for regional airports. In: Kersten, W., Blecker, T., Ringle, C. (eds.) Innovations and Strategies for Logistics and Supply Chains (Proceedings of the Hamburg International Conference of Logistics (HICL)), pp. 256–284. epubli GmbH, Berlin (2015)
9. Cameron, D., With, H.M., Skjerven, Ø., Hyllseth, M., MannsåKer, P.: Integrated distributed dynamic simulation of the LNG value chain: concepts and experience. Chem. Eng. Res. Des. **83**(6), 686–691 (2005)
10. Castillo, L., Dorao, C.A.: Consensual decision-making model based on game theory for LNG processes. Energy Convers. Manag. **64**, 387–396 (2012)
11. Cheenkachorn, K., Poompipatpong, C., Ho, C.G.: Performance and emissions of a heavy-duty diesel engine fuelled with diesel and LNG (liquid natural gas). Energy **53**, 52–57 (2013)
12. Cohen, W.M., Levinthal, D.A.: Absorptive capacity: a new perspective on learning and innovation. Adm. Sci. Q. 128–152 (1990)
13. Cooke, P.: Introduction: origins of the concept. In: Teoksessa Braczyk, H.-J., Cooke, P., Heidenreich, M. (eds.) Regional Innovation Systems (1998)
14. Cooke, P.: Regional innovation systems, clusters, and the knowledge economy. Ind. Corp. Change **10**(4), 945–974 (2001)
15. Daduna, J., Prause, G.: The Baltic Sea as a maritime highway in international multimodal transport. In: Doerner, K., Ljubic, I., Pflug, G., Tragler, G. (eds.) Operations Research Proceedings 2015, pp. 189–194, Springer, Berlin (2017). https://doi.org/10.1007/978-3-319-42902-1
16. Delgado, M., Porter, M.E., Stern, S.: Clusters, convergence, and economic performance. Res. Policy **43**(10), 1785–1799 (2014)

17. Doloreux, D.: What we should know about regional systems of innovation. Technol. Soc. **24** (3), 243–263 (2002)
18. Eisenhardt, K.M., Martin, J.A.: Dynamic capabilities: what are they? Strateg. Manag. J. 1105–1121 (2000)
19. Foss, M.M.: An overview on liquefied natural gas (LNG), its properties, the LNG industry, and safety considerations (2012). http://www.beg.utexas.edu/energyecon/INTRODUCTION%20TO%20LNG%20Update%202012.pdf. Accessed 23 June 2017
20. Gerlitz, L.: Design for product and service innovation in industry 4.0 and emerging smart society. J. Secur. Sustain. Issues **5**(2) (2015)
21. Gerlitz, L.: Design-driven innovation in SMEs: smart and sustainable organisation within industrial transformation. Int. J. Environ. Policy Decis. Mak. **2**(2), 98–124 (2017)
22. Gerlitz, L., Paulauskas, V.: Emerging LNG Value Chain in the South Baltic Sea Region (2015). http://martech.golng.eu/files/Main/PDF/2015.06.17_Joint_LNG_Study_consolidated.pdf. Accessed 23 June 2017
23. Holbrook, A., Salazar, M.: Regional innovation systems within a federation: do national policies affect all regions equally? Innovation **6**(1), 50–64 (2004)
24. International Transport Forum at the OECD: Global Trade: International Freight Transport to Quadruple by 2050. ITF, Paris (2015). https://www.itf-oecd.org/sites/default/files/docs/2015-01-27-outlook2015.pdf. Accessed 23 June 2017
25. Ketels, C., Pedersen, H.: State of the Region Report: The Top of Europe – Doing Well Today, Feeling Worried About Tomorrow, Baltic Development Forum, Copenhagen (2016). https://www.researchgate.net/profile/Christian_Ketels/publication/310793400_The_Top_of_Europe_-_Doing_Well_Today_Feeling_Worried_about_the_Future_State_of_the_Region_Report_2016/links/5836ee5808ae503ddbb5529b/The-Top-of-Europe-Doing-Well-Today-Feeling-Worried-about-the-Future-State-of-the-Region-Report-2016.pdf. Accessed 23 June 2017
26. Kogut, B., Zander, U.: Knowledge of the firm, combinative capabilities, and the replication of technology. Organ. Sci. **3**(3), 383–397 (1992)
27. Krugman, P.: Increasing returns and economic geography. J. Polit. Econ. **99**(3), 483–499 (1991)
28. Krugman, P.: Space: the final frontier. J. Econ. Perspect. **12**(2), 161–174 (1998)
29. Kumar, S., Kwon, H.T., Choi, K.H., Cho, J.H., Lim, W., Moon, I.: Current status and future projections of LNG demand and supplies: a global prospective. Energy Policy **39**(7), 4097–4104 (2011)
30. Kumar, S., Kwon, H.T., Choi, K.H., Lim, W., Cho, J.H., Tak, K., Moon, I.: LNG: an eco-friendly cryogenic fuel for sustainable development. Appl. Energy **88**(12), 4264–4273 (2011)
31. Madjidian, J., Gerlitz, L., Paulauskas, V., Henesey, L., Ölcer, A., Dalaklis, D., Ballini, F., Kitada, M.: Developing a strategy for LNG powered transport corridors in the Baltic Sea Region. In: Proceedings of 2017 International Conference on Maritime Energy Management, Malmö, Sweden, World Maritime University, Springer Verlag (2017)
32. Mattli, W.: The Logic of Regional Integration: Europe and Beyond. Cambridge University Press, Cambridge (1999)
33. Metzger, J., Schmitt, P.: When soft spaces harden: the EU strategy for the Baltic Sea Region. Environ. Plan. A **44**(2), 263–280 (2012)
34. Ministry of Trade and Industry Singapore. Singapore's LNG Terminal Starts Commercial Operations (2013). https://www.slng.com.sg/website/binarystream_processor.aspx?T=oHrZIePy%2F%2FIsmOw4n0HfkA%3D%3D&C=3nZ502kqfj3GSxT9GjBmJQ%3D%3D&PK=dt3HyIwl7xY%3D&K=01C351&SC=1. Accessed 23 June 2017

35. Mohapatra, S.: IT and porter's competitive forces model and strategies. In: Information Systems Theory, pp. 265–281. Springer, New York (2012)
36. Olaniyi, E.O., Prause, G.: The impact of environmental regulations on regional development in Eastern Estonia. In: Muravska, T. (ed.) Proceedings of the International Conference New Challenges of Economic and Business Development – 2017. Latvian State University, Riga (2017)
37. Olaniyi, E., Prause, G., Boyesen, J.: The impact of SECA regulations on clean shipping in the BSR: first empiric results from EnviSuM project. In: Proceedings of 2017 International Conference on Maritime Energy Management, Malmö, Sweden, World Maritime University, Springer Verlag (2017)
38. Pitt, M.E., Eide, M.S., Brathagen, S., Dimopoulos, G.: Natural gas versus diesel power for ships–a technical review, also considering economic aspects. DNV, Norway (2010)
39. Porter, M.E.: How competitive forces shape strategy. Harvard Bus. Rev. **57**(2), 137–145 (1979)
40. Porter, M.E.: What is strategy? Harvard Bus. Rev. **6**, 61–78 (1996)
41. Porter, M.E.: Location, competition, and economic development: local clusters in a global economy. Econ. Dev. Q. **14**(1), 15–34 (2000)
42. Porter, M.E.: The value chain and competitive advantage. Underst. Bus. Process. 50–66 (2001)
43. Porter, M., Kramer, M.R.: Creating shared value. Harvard Bus. Rev. **89**(1/2), 62–77 (2011)
44. Rüster, S., Neumann, A.: Corporate strategies along the LNG value added chain-an empirical analysis of the determinants of vertical integration. In: Globalization of Natural Gas Markets Working Papers, WP-GG-17. German Institute for Economic Research (2006)
45. Teece, D.J., Pisano, G., Shuen, A.: Dynamic capabilities and strategic management. Strateg. Manag. J. 509–533 (1997)
46. TIAX: US and Canadian Natural Gas Vehicle Market Analysis: Liquefied Natural Gas Infrastructure. Lexington, MA (2012). http://louisianacleanfuels.org/images/downloads/Reports/lng_infrastructure.pdf. Accessed 23 June 2017
47. Tzannatos, E., Nikitakos, N.: Natural gas as a fuel alternative for sustainable domestic passenger shipping in Greece. Int. J. Sustain. Energ. **32**(6), 724–734 (2013)
48. Verbeek, R.P., Kadijk, G., van Mensch, P., Wulffers, C., van den Beemt, B., Fraga, F., Aalbers, A.D.A.: Environmental and economic aspects of using LNG as a fuel for shipping in The Netherlands. TNO, Delft (2011)
49. Wernerfelt, B.: The resource-based view of the firm: Ten years after. Strateg. Manag. J. **16**(3), 171–174 (1995)
50. Wood, D.A.: A review and outlook for the global LNG trade. J. Nat. Gas Sci. Eng. **9**, 16–27 (2012)
51. Yoo, B.Y.: Economic assessment of liquefied natural gas (LNG) as a marine fuel for CO_2 carriers compared to marine gas oil (MGO). Energy **121**, 772–780 (2017)

Time Series Analysis and Prediction Statistical Models for the Duration of the Ship Handling at an Oil Terminal

Julia Rudnitckaia(✉) and Tomáš Hruška

Faculty of Information Technology Brno, Department of Information Systems, BUT, Božetěchova 1/2, 61200 Brno, Czech Republic
{irudnickaia,hruska}@fit.vutbr.cz

Abstract. The main points of this paper are researching of time series and, then, building statistical prediction models based on obtained characteristics. Such investigation is often conducted in medicine and economics. The practice shows, that the studying of such issues in the marine transport logistics area is in demand. Time series analysis comprises methods for analyzing time series data in order to extract meaningful characteristics of the data and forecast future, based on knowledge of the past. The following models are applied to ship handling duration data: exponential smoothing (single, double and triple), and two more sophisticated models – auto-regressive moving average (ARMA) and ARIMA with seasonal component (SARIMA). The choice of a suitable model depends both on the distribution of the data and on the useful information that it will bring. Experiments for every model are carried out with different values of the model parameters to find the best fitting one. For all models Mean Square Error (MSE) is calculated on the training and test data. The best results were achieved by ARMA model with weekly dataset frequency. Nevertheless, there are some other ways to improve prediction models and to obtain more accurate results, which are also proposed in the article.

Keywords: Time series forecasting · Statistical models · ARMA · SARIMA Exponential smoothing · Time prediction · Ship handling · Oil terminal

1 Introduction

A time series (TS) is an ordered sequence of values of a variable at equally spaced time intervals [1]. Examples can be found in a variety of fields ranging from medicine to economics. The investigation of the time series makes it possible to control the process that generates time series, to clarify the mechanism, situated in the basis of the process, to clear a number of outliers, and also to make predictions for the future based on knowledge of the past. The main characteristics of TS, which make it different from a regular regression problem, are time dependent and seasonality trends (i.e. variations specific to a particular time frame). The basic assumption underlying the analysis of time series is that the factors, affecting the object, in the present and in the past, will

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_12

affect it in the future. The goals of TS analysis are finding dependencies or patterns and express them in the form of mathematical expressions. To achieve this, many mathematical models have been developed, designed to study the TS components. Thus, it makes possible to forecast future values according to the historical data that can help in the development of plans and development strategy of the organization. As a rule, an analysis includes the following steps: building, identifying, fitting, and checking models [2]. All these steps were done to choose prediction models and implemented to analyzed dataset. Also, the most important components of TS (trend, seasonal variation, cyclic changes, and irregular factors) are considered and models are built according to them.

A time series analysis and search statistic prediction models are held in the port and reflected in the works [3, 4]. As a rule, the object of research is a container terminal, and the target forecast variable is terminal throughput volumes. The obtained models are effective in forecasting and applicable in practice. Nevertheless, using of such models at oil terminals is not so widespread. In connection with the specifics of the oil terminal, forecasting cargo volumes is not of great importance. Undoubtedly, the market conjuncture affects on the throughput volumes, but the choice of the transport corridor remains with the oil companies, which in turn make up a plan for the transshipped cargo distribution in advance for a year. In this paper we propose to draw attention to the parameter that determines the service level provided by the port – ship case handling duration. The multiplicity, dynamism and interdependence of the factors influencing this parameter, does not allow the forecast to be created by traditional methods with a sufficient degree of reliability. The ship case handling can be divided into cargo and auxiliary operations. And if there is no problem with forecasting of cargo operations duration, then the

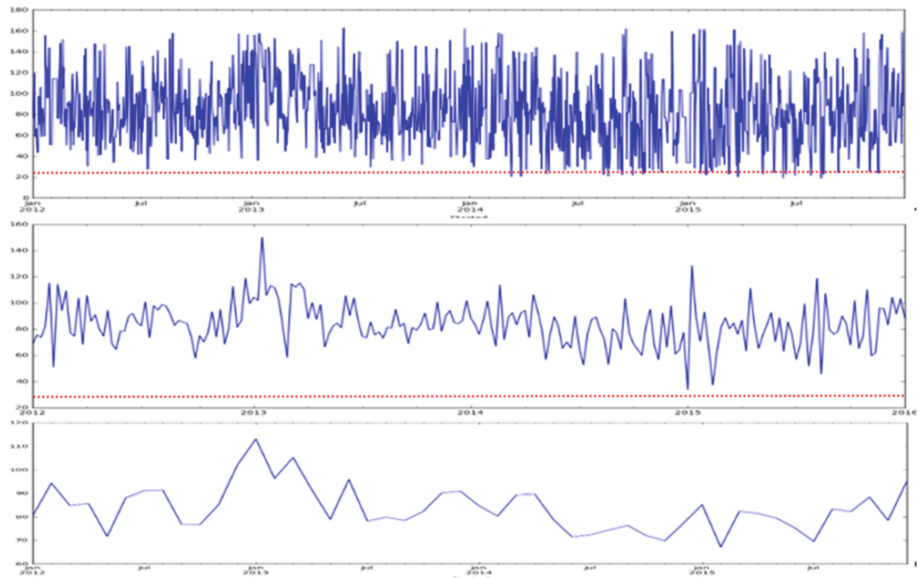


Fig. 1. Ship handling duration time plot with different frequencies (from top to bottom –TSd (days), TSw (weeks), TSm (months)).

auxiliary operations cause a number of difficulties. According to the analyzed port standards, cargo operations are forecasted for a day (regardless of 60 tn or 180 tn). Thus, this part of the ship case handling is taken as a constant (24 h), therefore Quantity of cargo parameter is not considered in the article. Auxiliary operations are of greater interest. The Fig. 1 shows the duration of the overall ship handling. Red dotted line specifies the cargo operations standard (24 h). The graph above this line represents the auxiliary operations duration. The presence of extremes is caused by delays in auxiliary operations and requires additional analysis to identify the reasons and factors contributing to this.

So, the purpose of the article is to determine the best statistical prediction and to find and predict possible delays in the ship handling with its help.

The paper is organized as follows. The first part comprises means and methods, using in the subsequent analysis. First of all, the core of work, the analyzed dataset, is described. This part discusses the way in which the data set was filtered and how it was modified before statistical models were applied. Further, the description of the TS, which shows the distribution of the ship handling duration for the last 3 years, and its main characteristics follow. Finally, two different approaches: exponential smoothing methods and auto-regressive (ARMA, SARIMA) methods are defined. The second part presents the results of prediction models application to the port data. According to MSE the best model is chosen. The last part of the paper is conclusion. As parameters of the above models are not fully measured and so some solutions to improve these statistical models (as well as further research) are proposed.

2 Means and Methods

2.1 Describing of the Analysed Dataset

The analyzed data were obtained from timesheets documents that comprise necessary information about full ship handling operations at an oil terminal for subsequent financial settlement between ship and a port. For time series analysis two values are needed. There are Duration of each ship-case handling (expressed in hours) and Start timepoint. According to results of my previous article [5], the data was pre-processed in the following way: ship cases with start activity "Ship arrived" and end activity "Pilotage" were defined as a process with a normal behavior; all ship cases that do not belong to confident interval {the median value ± standard deviation} were referred to data noise and outliers. Before pre-processing the analyzed data structure didn't have normal distribution and specified confident interval was more suitable. It was able to identify short duration cases as well as long-term ones. In total, 33% of all cases were filtered out. Also, it could be useful to analyze such cases and to find patterns or dependencies in data, which can explain the reasons for their appearance. However, this kind of analyses is beyond the scope of this article.

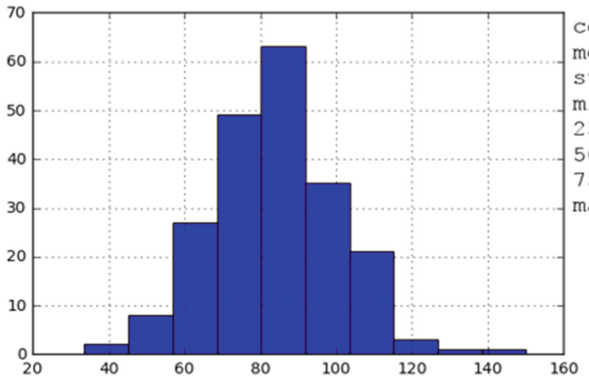
Ultimately, after all transformations, the data set has 2121 instances and two attributes: Started point of ship case handling and Duration (h.) of full handling (not just cargo operations). It includes period from 2012-01-02 to 2015-12-28. Frequency for time series analysis can be adjusted as daily, weekly or monthly. The one case takes an average of 83 h (more than 3 days), so it is logical to determine weekly frequency. So, although the daily one can provide more detailed model with slight fluctuations, it was decided to compare results of models, based on time series with weekly frequency (see Fig. 1 TSw (weeks)).

2.2 Analysing of Time Series

In this work for time series analyzing and other investigations language Python, in particular librarystatsmodels, was used. First, to get a general idea of the data, time series should be visualized (see Fig. 1).

Figure 1 shows two peaks for TSw (January 2013 and 2015) and one peak for TSm (January 2015), but this figure is not specific enough to discuss whether the data has trend or seasonality. As it is explained above, all prediction models are based on TSw, which contains 210 instances and belongs to time interval [2012-01-01; 2016-01-03].

According to TSw indicators and histogram, depicted on Fig. 2, TSw more homogeneous and has a relatively small dispersion. It is also evidenced by the coefficient of variation (the ratio of standard deviation to average value) that equals to 0.387. Moreover, to determine the normal distribution, Jarque-Bera test was conducted and the null hypothesis of the normality is rejected with p-value = 0.11 (as a rule, a threshold equals to 0.05) that means series has a normal distribution [6]. Such results were expected, because of preprocessing input dataset.



count	210.000000
mean	83.826460
std	16.702945
min	33.670000
25%	73.929038
50%	82.635076
75%	94.124196
max	150.102500

Fig. 2. TSw bar chat and its main indicators.

Stationarity. The explored time series (TSw) is stochastic and discrete, don't have a pronounced trend. Nevertheless, in the first place, time series stationarity should be checked to find out if it is suitable for statistical techniques. Dickey-Fuller testing [7] showed that mean value appears to be varying with time slightly and test statistic is much smaller than 1% critical values so we can say with 99% confidence that this is a stationary series.

In order to ensure TSw stationarity another way, ACF (Autocorrelation function that presents correlation between different lag functions) and PACF (Partial autocorrelation

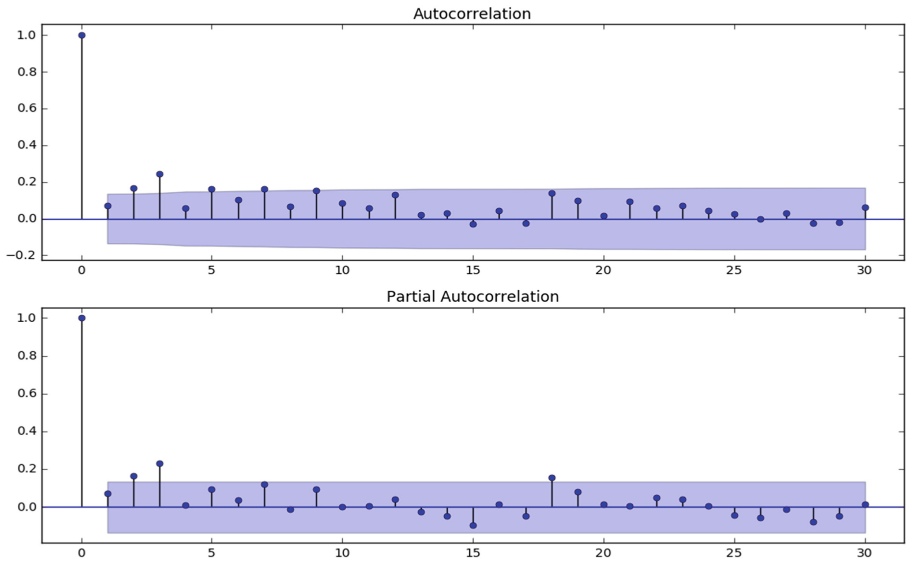


Fig. 3. ACF and PACF of TSw.

function) are showed on the figure below. For stationary time series both functions should decrease to the zero as the lag increases, as evidenced of the presented functions.

Time Series Components. On the Fig. 4 there are the seasonal variation, which will be processed in SARIMA model, and trend, which can be caused by changes in the organization. Trend component indicates improvements and reduction of the ship handling after 2014 year. As regard to seasonality, in the first half of the year, there are difficulties in the ship cases handling.

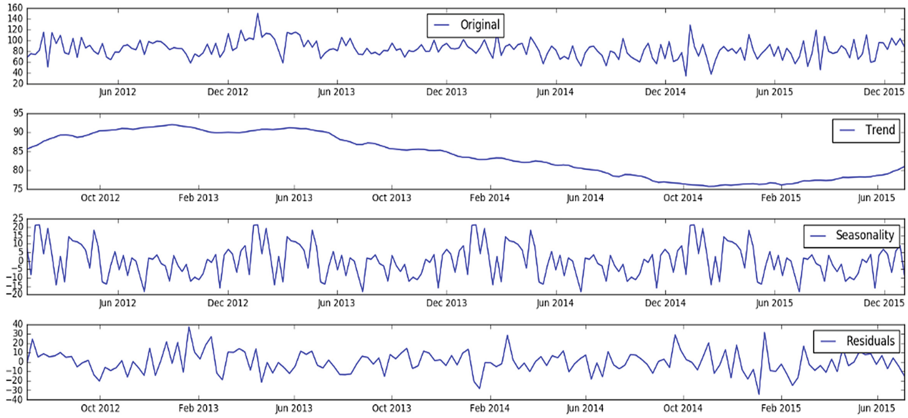


Fig. 4. Decomposition of TSw.

2.3 Time Series Prediction Models

Splitting Data. The model performance is measured on the test data. So, before model is built, it is necessary to divide the data into two parts: for training and testing. Since we work with TS, which has time structure and dependencies, cross-validation method was omitted. Data were separated manually. Training data take the values from start of TSw to end of 2014 year, and test data take all remaining values.

That is training data belong to interval [2012-01-01; 2014-12-28] and test data – [2015-01-04; 2016-01-03]. After splitting the data, we took the training data to analyze and find the best fitting model.

For model comparison AIC (Akaike Information Criterion) and MSE (Mean Square Error) criterion are used.

The models that have the lowest AICs values tend to give slightly better results than the other models.

Exponential Smoothing. Exponential smoothing of time series data assigns exponentially decreasing weights for newest to oldest observations. There are three kinds of exponential smoothing: simple, double, and triple (more information about each type in the online source [8]). All three methods were built for our TSw and are depicted on figures below. On the Fig. 5 it is clever to define more suitable coefficient (for Single it is 0.3, for Double (0.9; 0.9)). Also, we have obtained low values of MSEs with these coefficients (for Single MSE = 139.5, for Double MSE = 179.59).

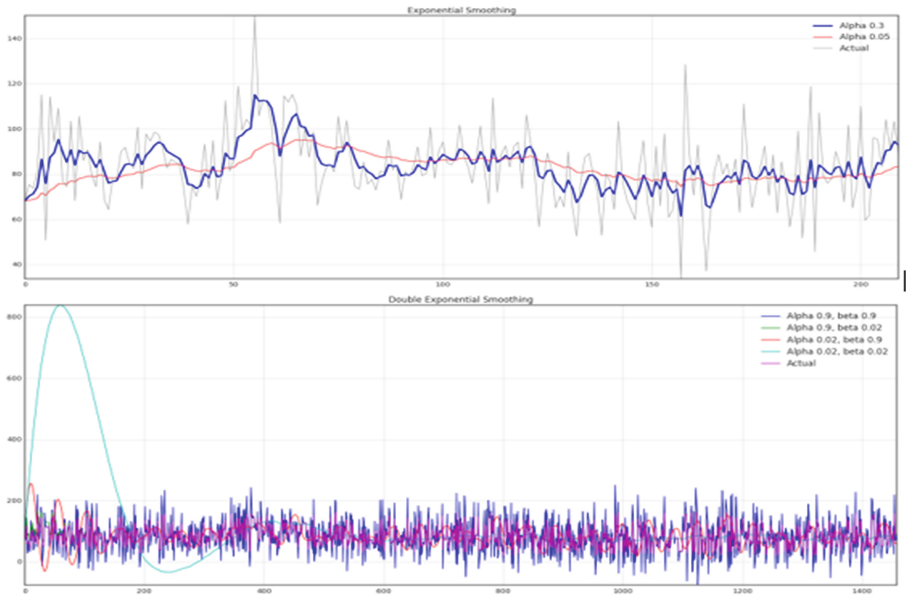


Fig. 5. Single and double exponential smoothing.

However, as previously discussed, analyzed TSw has trend and seasonality. Therefore, we use triple exponential smoothing for further prediction (see Fig. 6).

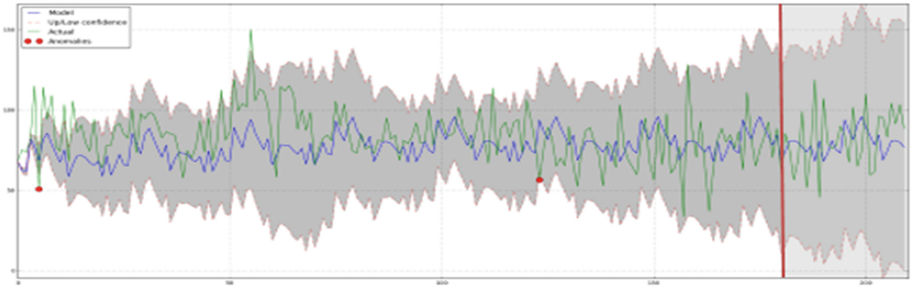


Fig. 6. Triple exponential smoothing or Holt-Winters model.

ARMA, SARIMA. ARMA stands for Auto-Regressive Moving Averages and provides a description of the stationary stochastic process in terms of two polynomials, one for the autoregression and the second for the moving average [2]. Finding appropriate values of p and q in the ARMA (p, q) model can be facilitated by using ACF and PACF plots (Fig. 3).

SARIMA is Seasonal autoregressive integrated moving average model. It is extension of ARMA model (with seasonal component). Often time series possess a seasonal component that repeats every s observations. For monthly observations s = 12 (12 in 1 year), for quarterly observations s = 4 (4 in 1 year).

3 Results

In this part of the paper three prediction models will be built, based on discussions in the previous part. There are Holt-Winters, ARMA, and SARIMA.

Holt-Winters. For this model, the main problem is to determine the best coefficients. Also, the resulted model should not be overfitting or generating. That is why the following coefficients were chosen: $\alpha = 0.0066$, $\beta = 0.0$, $\gamma = 0.0467$. Figure 6 demonstrates results of forecasting with confidential interval, with Brutlag's algorithm of anomalies detection [9]. Training and test data are separated by red line. In Table 1 values of MSEs for both data set are pointed.

Table 1. Comparative table for prediction models.

Model	MSE	
	Train data	Test data
Triple smoothing/Holt-Winters ($\alpha = 0.0066$, $\beta = 0.0$, $\gamma = 0.0467$)	299.24	387.35
ARMA (3,3)	211.34	350.16
SARIMA (313) $(110)_{12}$	392.01	422.3
SARIMA (303) $(113)_{12}$	399.01	496.82

ARMA. As discussed earlier, important parameters (p, q) can be derived by ACF and PACF plots. On the Fig. 3 it can be observed that both functions tend to zero values after lag 3. It means parameters p = q = 3 and our model looks like ARMA (3,3) or ARIMA(3,0,3). Figure 7 shows actual time series, which is indicated by a blue line, and forecast by a red dot line.

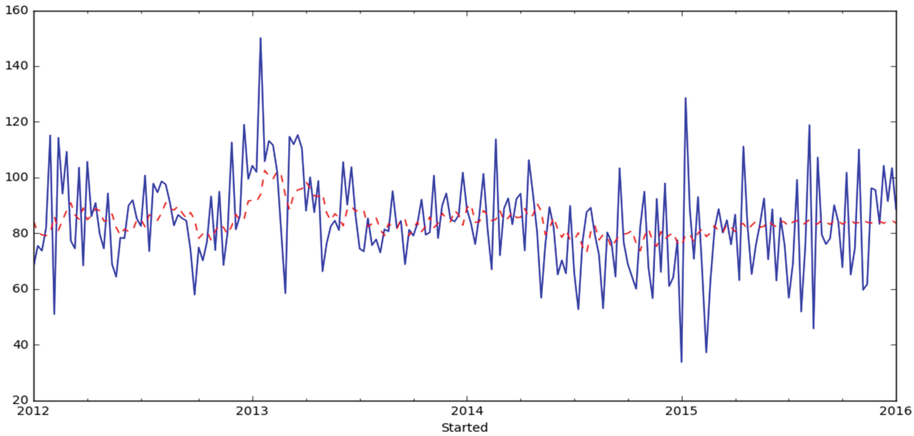


Fig. 7. ARMA (3,3) prediction model.

SARIMA. To choose parameters for this model, additional investigations were conducted. The grid search method was used for searching of the best parameters. We tried a suite lag values (P) and just a few difference iterations (D) and residual error lag values (Q). From the Fig. 4 we can conclude that s = 12 (a year). To reduce the number of calculations (p, d, q) remains the same as in ARMA, i.e. (3,0,3). The best parameters are reported as ARIMA $(3,1,3)(1,1,0)_{12}$ and are corresponding to the lowest MSE value and to quite low AIC value. Also, the model ARIMA $(3,0,3)$ $(1,1,3)_{12}$ has a good results and it will be added to comparison table (Table 1). It is necessary to note, that the best parameters cannot be found based just on AIC. There were iterations with fairly low AIC, but model couldn't be built. Moreover, it is important to run the model results diagnostics to assure that none of the assumptions made by the model was violated.

Forecasts from the ARIMA $(3,1,3)(1,1,0)_{12}$ model (which has the lowest MSE value on the test set) are shown in the Fig. 8.

To compare results of all developed models, comparative table is presented below. The most accurate model is ARMA (3,3).

Although, it shows the lowest MSE with train data as well as with test data, it still needs furthermore investigation. To check whether this model can be further improved, other forecast errors should be examined. For instance, to get more appropriate model, one can change seasonal component or choose a model not with the lowest AIC (despite the theory, such models can be more accurate).

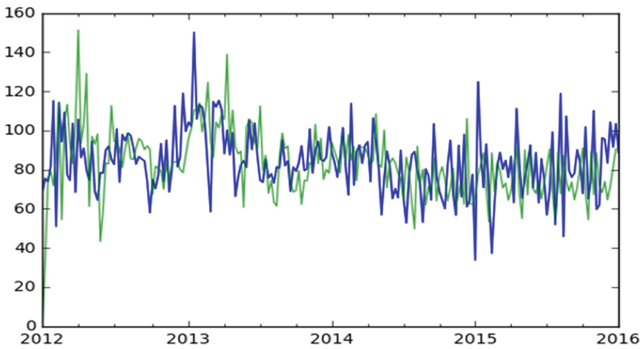


Fig. 8. ARIMA(3,1,3)$(1,1,0)_{12}$ prediction model.

4 Discussion

Based on historical data, we reveal the delay in the ship handling duration and, taking them into account, make forecasts for the future. The port works continuously, however, third-party organizations are involved in the auxiliary operations, and therefore the time series shows seasonality.

Occurrence of these delays can be caused by both external (meteorological) and internal factors (lack of resources, reorganization processes, etc.). External factors can be tracked with the weather archive help (since temperature and precipitation do not matter, wind force is the defining parameter, but mainly it is vital to note storm warnings, because there is a chance of a lightning strike even in sunny weather). Internal factors are more difficult to identify, and in this article we consider the processing ship handling as a black box with a start and end point, i.e. without analyzing internal processes.

The results of this article also can be used as a comparative characteristic for the further researches. Statistical methods do not support multicriteria analysis and can identify and predict a problem period. In order to determine the causes of delays and to make a more accurate prediction, it is necessary to use the process model (the creation of which is described in the article [10]) and more complex prediction methods such as machine learning methods (ANN, GA, SVM) and process mining method (transition system). Moreover, adding of the new process attributes can improve the results of the model.

5 Conclusions

In this article, time series, which shows the distribution of the ship handling duration for the last 3 years, was analyzed. After filtering and preprocessing data, we visualized times series and defined its main characteristics. There is seasonal component (s = 12) and small trend (overall ship handling duration is reduced, that says about some changes in the organization). Taking into account obtained characteristics of the time series, three prediction models (Holt-Winters, ARMA, SARIMA) were built and compared. The best results, according to MSE for application model on test data, showed ARMA(3,3).

However, it is necessary to note, that models strongly depend on the choice of parameters. In this article, just the range {0…3} for p, q, P, Q was considered.

For the port, the applying of prediction models is important. Correct forecast will help reduce the idle time of port resources, and determining the causes of delays will improve the service of the ship handling. All this, in turn, can lead to a reduction of material cost on the part of vessels and increase the productivity of the port, the utilization of its capacities, more realistic planning time-frame and thereby the increase of competitiveness among other transport channels.

Acknowledgements. This work was supported by The Ministry of Education, Youth and Sports of the Czech Republic from the National Programme of Sustainability (NPU II); project IT4Innovations excellence in science – LQ1602.

References

1. Rausch, P., Sheta, A.F., Ayesh, A. (eds.): Business Intelligence and Performance Management: Theory, Systems and Industrial Applications. Springer, London (2013)
2. Box, G.E.P., et al.: Time Series Analysis: Forecasting and Control. Wiley, Hoboken (2015)
3. Chen, S.-H., Chen, J.-N.: Forecasting container throughputs at ports using genetic programming. Expert Syst. Appl. **37**, 2054–2058 (2009). https://doi.org/10.1016/j.eswa.2009.06.054
4. Pen, W.-Y., Chu, C.-W.: A comparison of univariate methods for forecasting container throughput volumes. Math. Comput. Model. **50**, 1045–1057 (2009). https://doi.org/10.1016/j.mcm.2009.05.027
5. Rudnitckaia, J.: Methods of detection of noises and outliers in data structure on the example of ship handling duration at an oil terminal. Vestnik Gosudarstvennogo Universiteta Morskogo I Rechnogo Flota Imeni Admirala Makarova **4**(44), 866–873 (2017). https://doi.org/10.21821/2309-5180-2017-9-4-866-873
6. Ruppert, D., Matteson, D.S.: Statistics and Data Analysis for Financial Engineering. Springer, New York (2015)
7. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting. Springer, Cham (2016)
8. NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/
9. Brutlag, J.D.: Aberrant behavior detection in time series for network monitoring. In: Proceedings of the 14th USENIX Conference on System Administration, New Orleans, Louisiana (2000)
10. Rudnitckaia, J.: Study of descriptive process model of ship handling at an oil terminal based on timesheets. Vestnik Astrakhan State Tech. Univ. **1**, 7–15 (2017)

Advanced Vibration Diagnostics for Perspectives of Helicopter Technical Maintenance

Aleksey Mironov(✉), Pavel Doronkin, and Alexander Priklonsky

D un D centrs, 7 Margrietas street, Riga 1046, Latvia
info@ddcentrs.lv

Abstract. This paper considers different aspects of advanced vibration diagnostic system development, validation, promotion and application. Authors discuss the list of tasks for such system from the helicopter technical maintenance and repair point of interest. Main problems and possible solutions are considered, including uniform diagnostic platform for board and ground systems, life cycle traceability for helicopter aggregates, single survey diagnostics etc. The general view on advanced system is presented, based on the state-of-the-art and advanced techniques of vibration diagnostics and SHM. Resolution levels for helicopter power unit aggregates are considered based on solutions of the high-resolution diagnostics collected under *Vibropassport™* brand. Authors discuss operating demonstrators of diagnostic techniques as well as application cases as the main tools for advanced system promotion and market entering.

Keywords: Vibration diagnostics · Helicopter · Technical maintenance

1 The Tasks and Solutions for Advanced Diagnostic System

1.1 The Tasks

Perspectives of helicopter technical operations is the basis of condition based maintenance for all helicopter life cycle, including manufacturing, overhaul and even utilization. Taking above into account, the following tasks for advanced diagnostic system [1] to be considered:

- Monitoring of power unit and transmission aggregates of a helicopter;
- structural health monitoring (SHM) of most damageable parts of a helicopter, like blades, tail boom, landing gears;
- High resolution diagnostics of aggregates on production and overhaul stages;
- Capability to manage technical state of main helicopter aggregates in field conditions;
- Common diagnostic, data and information platform for on-board as part of Health and Usage Monitoring System (HUMS), portable ground-based and stationary expended vibration diagnostic systems for each helicopter type;

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_13

- Applicability of portable vibration diagnostic terminals (VDT) on non-equipped by HUMS helicopters, including:

– quick and easy operation on board in field conditions for engine survey,
– no screws locking or dismantling of aggregates and violation of documentation,
– transportability of VDT set by air or other way.

- Fleet vibration data development based on monitoring center that maintains interactive vibration data base (IVDB), including data concentration from operation (both *HUMS* and *VDT*) and stationary systems, links to both operators and manufacturers, regular update of thresholds;
- Monitoring and diagnostic services for helicopter operators, including data concentration, thresholds update, monitoring of helicopter aggregates parameters, single survey diagnostics on call, full monitoring and diagnostics of operator's helicopter fleet, fleet tendencies analysis and development of recommendations for manufacturers and maintenance and repair operator (MRO).

The monitoring and diagnostic system conforming to above requirements is not limited by flight safety tasks, but becomes part of helicopter technical life cycle including production, operation and utilization. What is the current situation and what is lacking?

1.2 The Problems and Solutions

Problems of existing diagnostic systems are typical for all helicopter aggregates so, here below it is considered on an engine as an example.

Uniform diagnostic platform and life cycle traceability. The existing vibration diagnostic systems, both stationary and board ones carry out limited set of tasks. Most systems at manufacturer test rig are limited and may indicate one of rotors as the sources of high vibration only.

Aiming at different tasks, existing stationary and board systems apply different concepts and diagnostic techniques and are not related directly and mutual data exchange is complicated. To conform to advanced diagnostic requirements the system needs unified platform of measurement, data collection and development techniques for any stage of engine's life cycle.

Common techniques and parameters would allow smooth data exchange between all diagnostic systems and provide traceability of engine's condition during its life cycle starting from manufacture till utilization.

High resolution diagnostics. The main approach of state-of-the-art HUMS is finding abnormality in operation of aggregates that is determined using vibration parameters trend. Algorithms of abnormality identification use multiple observations during long operation therefore it may not detect a failure immediately after its appearance. So, there is a time lag between a failure and its indication by ground station of existing HUMS.

Advanced diagnostic system must have a high resolution of each aggregate state that has to be experimentally validated. The high resolution diagnostics may not be substantiated on mathematics only and needs extensive research study. Based on experimental and theoretical researches, the appropriate vibration models were built up and then its adaptation to specific aggregate type was provided. The new techniques of data development and diagnostic parameters computation follow and then they are experimentally verified. These verified parameters allow evaluate each engine or helicopter aggregate and even its specific units, like bearing, compressor stage, oil pump, gear etc.

Single survey diagnostics. It is important, that advanced diagnostic system may allow state identification using single observation. It is necessary both for an engine tested at engine shop and for helicopter engines at a maintenance check even if it is a first test. Single survey diagnostics could become reality if any diagnostic parameter has common relative scale of state for all aggregates of the same type. In other case a scatter of individual vibration parameters would reduce resolution of a diagnostic technique because of expanded tolerances of thresholds.

Automation. A human role in any diagnostic system has to be left for final solution about engine's state for authorized and skilled persons.

The advanced diagnostic system registers all data automatically and does not require any skilled staff for that. Depending on system version, the collected data may be sent to IVDB or be developed in situ using VDT if an operator has skilled and authorized personnel for providing diagnostic solutions. One of the main privileges of an advanced system is application of those vibration diagnostic techniques that allow automatic detection of latent failures signs and its indication without attraction of diagnostic specialists. The only solution to be taken by specialist is how long and on what terms the detected fault could be tolerated. An important role of monitoring and diagnostic center will remain for fleet data analysis, thresholds update and recommendations development, but will even expand as it will concentrate data from VDT servicing helicopters without HUMS.

1.3 Common Solution of an Advanced System

In order to eliminate written above limitations and to monitor helicopter aggregates during all life cycle an advanced diagnostic system shall combine all above discussed solutions:

- The advanced system has terminal and central levels;
- The both board and ground components may represent terminal level of the advanced system, as well as stationary system on a test rig;
- The common methodology and unified software is the core of both system levels;
- The relative scale of diagnostic parameters and its unitary boundaries for whole fleet, providing both high resolution diagnostics and single survey diagnostic capability;
- The automatic data processing up to decision making;

- The united vibration data base maintained by the monitoring and diagnostic center that interacts with operators, manufacturers and MRO as well as provides boundaries for diagnostic parameters based on data from all participants;
- The both autonomous and centralized variants of the system application are available depending on customer needs and capabilities.

So, at helicopter production or overhaul stage the terminal level is represented typically as an equipment set fixed to the test facility of a main gearbox, of an engine or other aggregates. In field conditions the onboard and ground parts of HUMS equipped helicopters will be on the terminal level, whereas other helicopters need autonomous portable measurement set that will be capable for application within technical maintenance. The terminal level will be related to central level as monitoring and diagnostic center provided with data base.

2 Advanced System – How It Looks

2.1 System Configuration

In order to achieve desired capabilities the advanced helicopter monitoring & diagnostic system must obtain essential improvements in comparison with actual systems, including advanced measurement equipment and perspective vibration diagnostic and SHM techniques. For light and unmanned helicopters the specific portable diagnostic system should be created. Rotating HUM part and SHM part with wireless data transfer will upgrade main system to become advanced one. Helicopter SHM system including sensor network will become the new part of HUMS and will include both rotating and stationary parts. The rotating part will cover main rotor aggregates diagnostics, but the stationary one will allow monitoring of a tail boom and landing gears.

Stationary part. State-of-the-art HUMS to be modified dramatically (Fig. 1). As actual HUMS an advanced system have stationary data acquisition unit (DAU) (1) in avionic bay receiving signals from stationary sensor network and from rotating data unit (2). Three-axial accelerometers replace single ones on a fuselage bottom (3), engines (4), oil fan (5), main gearbox (6) and tail gearbox (7). Only drive train of tail shaft (8) has two-axial accelerometers. Also tachometers (phase indicators) of engines, main and tail rotors relate to stationary DAU. Through embedded wireless receiver stationary DAU obtains data from rotating data unit.

Replacement of single accelerometers to tri-axial ones provides spatial vibration vector measurement in wide frequency and amplitude range allowing deep diagnostics of all aggregates of helicopter power unit using advanced vibration diagnostic techniques:

- The accelerometer mounted on a fuselage bottom (3) at vertical axis of a helicopter being used together with other accelerometers provides monitoring of landing gear functionality and fuselage integrity [2];
- The accelerometers and tachometers of engines supply data for monitoring of:

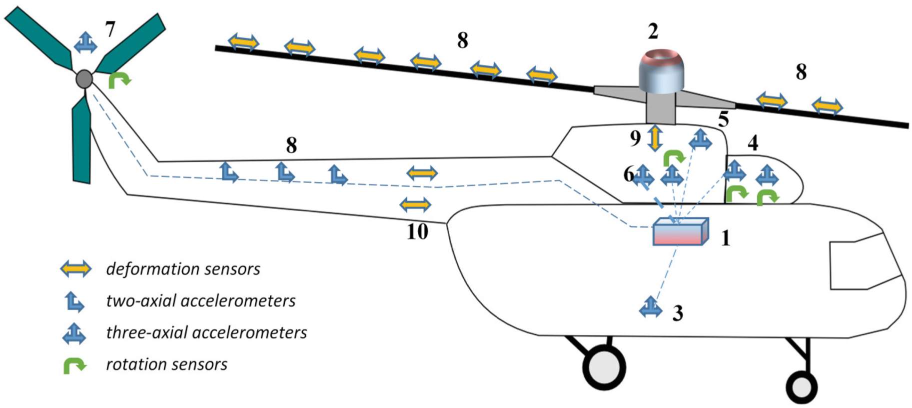


Fig. 1. Allocation of advanced diagnostic system components.

– engine mounts serviceability;
– aero & mass unbalance of a compressor and a turbine;
– each compressor & turbine stage aerodynamic uniformity (blades aerodynamic capabilities failure caused by wear or damage) and compressor instability (surge margin lose);
– hot gas path uniformity (gas temperature unevenness downstream of combustion chamber);
– bearings & gears;

• The accelerometers of the oil fan (5) and main gearbox (6) using main rotor phase indicator allow its state monitoring as well as its balancing considering aero imbalance;
• The accelerometer and a phase indicator on a tail gearbox (7) provide gearbox and its bearings monitoring as well as tuning of the tail rotor;
• The accelerometers of the drive train (8) provides monitoring of shaft and bearings.

SHM part of the advanced HUMS includes a sensor network (10), that is allocated around a helicopter. Depending on system's purposes, sensors could be allocated in each part of a helicopter, but they should be placed at least along the tail boom as the most critical structural part. Data from the sensor network provides a wide dynamic range signals allowing application of perspective Operational Modal Analysis (OMA) approach application for integrity monitoring of tail boom structure. Above approach uses multiple responses of allocated sensors for structure's modal properties estimation.

Rotating part. The rotating part of HUMS provides monitoring of main rotor and has both sensing and measurement parts. As shown on Fig. 1, rotating sensing network consists of embedded dynamic deformation sensors (8) on each blade and same sensors on main rotor shaft (9) as well. Rotating DAU (2) provides signal multiplexing, conditioning, analogue-to-digital conversion and wireless data transmission to the stationary part (1).

The measurement controller manages serial interrogation of dynamic deformation sensors of each blade and a shaft cyclically. Wireless data transmitter sends DAU measured data to principal measurement unit located in avionic bay of a helicopter.

In order to supply power for the rotating measurement system, an autonomous harvesting unit is to be included in the system. The rotating and stationary parts of DC generator are located on main rotor shaft and pitch control systems. The harvesting system supplies power to DAU, wireless transmitter and preamplifiers of sensors allocated along blades and shaft. The rotating HUMS also has own phase indicators mounted on each blade.

Signals management. The stationary DAU also applies serial interrogation of accelerometers groups aiming to reduce the number of expensive measurement channels. The principal data controller of HUMS manages data from both rotating part (received wireless) and stationary part. The data is going for in-flight processing to check alarm parameters and permanently for data recording.

Thanks to serial interrogation both stationary and rotating DAU have essentially less measurement channels than sensors quantity. Table 1 considers sensors allocation and measurement channels quantity of advanced system around a typical helicopter with 3–5 blades depending on its size. For instance, the systems configuration on a medium size helicopter has 13 accelerometers, 108 deformation sensors and 10 phase indicators. However, minimum number of required channels is 15 in stationary part and 13 in rotating one. So, total number of measurement channels in advance system would not increase in comparison to the actual ones.

Table 1. Sensors in advanced HUMS.

		Part	Stationary									Rotating				Stationary						
Sensor type			accelerometers								deformation sensors				revolution/ phase indicator							
		kind	3-ax						2-ax													
	type	location	engine 1	engine 2	MGB	oil fan	tail gearbox	fuselage	drive train	total	tail boom	rotor shaft	blade	total	each blade	engine 1 GG	engine 1 PT	engine 2 GG	engine 2 PT	main rotor	tail rotor	total
Helicopter	light	3	1	1	1	1	1	1	2	8	12	2	14	56	3	1	1	1	1	1	1	9
	medium	4	2	2	2	1	1	1	4	13	20	4	21	108	4	1	1	1	1	1	1	10
	heavy	5	2	2	2	1	1	1	6	15	24	4	21	133	5	1	1	1	1	1	1	11

Portable diagnostic set. The portable diagnostic set may autonomously provide monitoring and diagnostics for such types of helicopters that do not have HUMS.

Portable set includes a set of equipment, including easy mounted vibration and additional rotation sensors, T-connectors for board tachometers taping, data acquisition unit (DAU), ruggedized laptop, tool set for easy mounting and software providing signals conditioning and transformation as well as data development, diagnostic parameters computation and exchange with monitoring and diagnostic center. Portable autonomous system applies techniques for monitoring and diagnostics of aggregates of helicopter power unit and transmission. A portable set represents a terminal level of advanced system for non-HUMS helicopters, just like a board system (HUMS) is a terminal level of actual HUMS.

Small preparatory works and vibration data obtained during engine test allow quick survey of helicopter aggregates diagnostics. Regular use of the portable system allows monitoring of above aggregates even on non-HUMS helicopters. Such portable system may be used for diagnostics of power unit and monitoring of helicopter structural parts. One portable system and skilled specialist may provide data collection for monitoring and diagnostic services for many helicopters.

Unification of an advanced system. Uniform methodical, software and information basis provide compatibility of both board and portable ground versions of the advanced system. The both systems use the same algorithms, partly software and most part of diagnostic parameters. That means three components of the advanced system: board, portable ground and diagnostic center could freely interact between each other using the same typical diagnostic parameters. By this way an advance system could cover all kind of helicopters whether it has the board system or not. Typically, only few percent of helicopters are equipped with board systems (HUMS) which operations may cause higher risks. The majority of helicopters do not have HUMS, due to economic reasons, so availability to perform its monitoring by portable systems becomes an important advantage for most of helicopter fleet.

2.2 Methodical Solutions

In order to realize the ambitious tasks, an advance diagnostic helicopter system must have principal novelties, based on sophisticated vibration models and breakthrough data development techniques. As the most important some novelties could be outlined, like high-resolution diagnostics of power unit aggregates and helicopter transmission, multi-patching approach of OMA application, data pre-processing transformation for its condition monitoring and other modal analysis techniques application for portable ground SHM system.

Solutions for high-resolution diagnostics. The experimentally verified physical models and advanced data development techniques serve as the set of tools for diagnostic parameters computation. Such set of tools developed by *D un D centrs* was called *Vibropassport*^{TM} [3]. There are three groups of physical models in the basis of Vibropassport$^{\text{TM}}$:

- spatial impulse models for bladed aggregates and epicyclical gears,
- dedicated models of bearing dynamic forces and vibration, and

- models of mechanical properties relation to modal parameters of each helicopter structure.

The advanced diagnostic parameters allow estimation of current technical state of each particular aggregate or even its separate component. However, even very accurate computation of parameter is not enough for exact diagnosis if thresholds of above parameters are scattered. The advanced diagnostic techniques must use relative domain for parameters scaling that could be common for all individual engines of the same type and sometimes for other types. In relative scale thresholds of diagnostic parameters are not measured in vibration units, but rather transformative features of the aggregate.

With regards to an engine, the advanced system is capable to detect latent faults even if vibration levels did not change. Aero- and mass unbalance, compressor stability, technical state of each bearing and gear, compressor and turbine stages could be controlled. This approach differs from actual HUMS, allowing fault detection of gas generator or power turbine entirely when vibration levels exceed threshold. High-resolution techniques expand diagnostics abilities also for main gearbox, where on top to typical techniques of actual HUMS, the advanced system would monitoring of epicyclical stages and its bearings as well. A tail rotor to be balanced without additional tests using parameters computed based on flight data only thanks to advanced balancing technique.

Multi-patching in OMA application. OMA techniques allow SHM of aircraft main structures, including stationary and rotating ones (like blades). Both monitoring of its integrity and early identification of developing failures are based on the relation between mechanical and modal properties of structural components. To obtain these data two main tasks of modal analysis are to be carried out: efficient excitation and adequate measurement of structural response. OMA allows determination of modal properties of aircraft's structure using random excitation by airflow in flight. But this technique application requires the response to be measured in multiple points of the structure. So called multi-patching approach (serial measurements of sensor groups or patches) allows OMA realization using limited channels of the acquisition system.

3 The Way to Advanced System

3.1 The Demonstrator of Advanced System

A promotion of the advanced system reminds a road, where two trains must meet on a halfway. From one side a helicopter manufacturer needs to move with further validation and successful applications of his product, from another side a potential customer like MRO needs confidence in new techniques capabilities, its practical applications and his future benefits. Only practical approval in natural conditions or similar may convince the helicopter operator and support producer's efforts. The operating demonstrator of advanced vibration diagnostic techniques is one of the most effective tools for both of the above mentioned tasks. Specially built facilities conforming to requirements and solutions mentioned in p.2 may be the basis for research of advanced techniques as well as for certification tests and may act as techniques demonstrators for customers.

The demonstrators briefly described below does not need to look like helicopter, but have to functionally include most of helicopter operating units, as engine, drive train, main gearbox, main and tail rotors and some other aggregates (Figs. 2, 3, 4, 5 and 6). Each of these units has own electric drive and are controlled in operation modes similar to actual ones. To conform to main tasks – demonstration and experimental research – the construction of each unit is tuned to provide easy state modification of the aggregate. Such preparation allows demonstration of seeded faults influence on VibropassportTM parameters response. The demonstrator may also work for validation of any vibration diagnostic techniques that supposed to be applied on helicopter operating units.

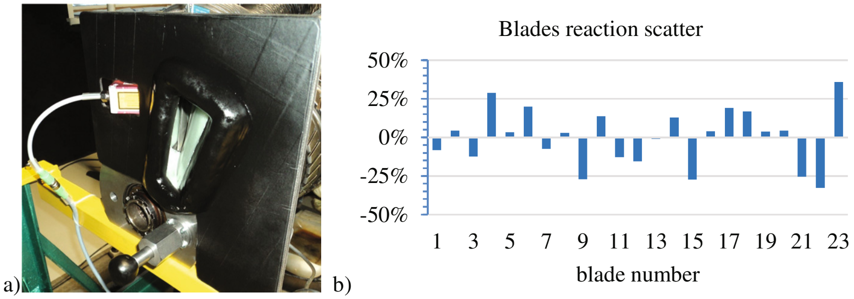


Fig. 2. Blades aerodynamic features estimation: (a) – the test rig, (b) – scatter of blades reaction to aerodynamic loads.

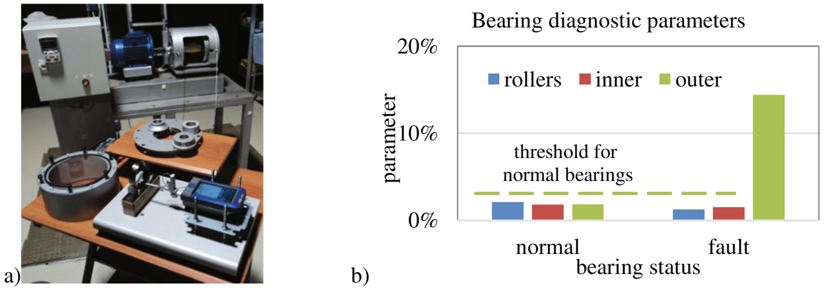


Fig. 3. Bearing testing: (a) – testing bench, (b) – diagnostic parameter relation to bearing status.

Engine diagnostic system demonstrator. Preparation to the Vibropassport™ demonstration takes from few minutes till days depending on the engines part considered. For instance, diagnostic ability for unevenness of turbine flow duct could be arranged by prompt switching of air supply system through combustion chamber however, test fault implementation into the bearing could take a day or even more.

a) b) c)

Fig. 4. Demonstrators: (a) – balancing rig, (b) – drive train and tail rotor, (c) – main gearbox and main rotor diagnostic demonstrator.

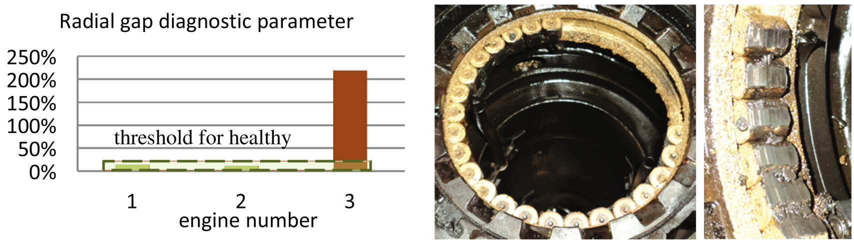
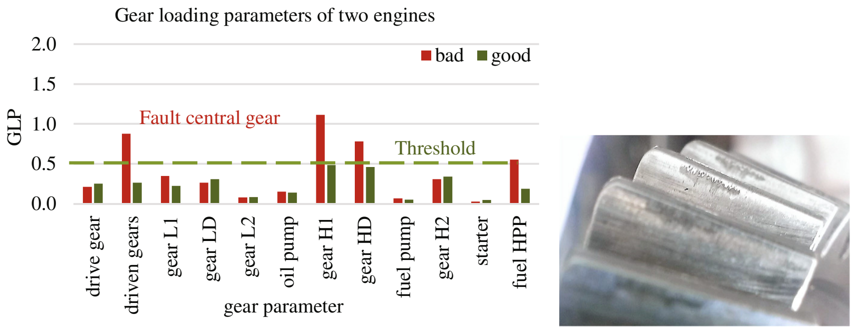


Fig. 5. Faulted bearing: histogram of bearing radial gap parameter (left) and photo of the collapsed bearing (right).



a) b)

Fig. 6. Engine gears diagnostics: (a) – diagnostic diagram, (b) – photo of teeth damage.

The engine demonstrator is equipped by 3-axial acellerometers, the phase indicator, dynamic pressure and deformation sensors. There are also unique actuation system simulating impulse interactions between blades and vanes. The list of optional failures that may be simulated in the engine prototype is:

- Compressor or turbine unbalance,
- Shadowed intake,
- Damaged blades or unevenness of blade rows of compressor stages,
- Reduced stall margin of separate compressor stages,
- Unevenness of flow turbine upstream,
- Overlapping of turbine blade shrouds,
- Damage of inner, outer ring or rollers/balls,
- Damaged teeth of gears,
- Misaligned gears.

For calibration of testing damages some auxiliary facilities are included to the demonstrator.

The test rig (a portable wind tunnel) for scatter estimation of blades aerodynamic features (Fig. 2a) may be applied to a compressor blade row (assembled on a disc).

This experimental facility allows estimation of each blade deflection in response to identical flow actuation. Difference of aerodynamic and mechanical properties causes variance of blade tip deflection. The test rig may also demonstrate to a customer the necessity to consider unevenness of components of engine flow duct. Figure 2b illustrates the scatter of blades mechanical reaction (deflection of blade tip) to identical aerodynamic loads of the worn blade row (1st compressor stage). The histogram shows – 45% … 35% variation of blades deflection caused by its properties scatter. The specific bearing testing bench (Fig. 3a) provides tuning of bearing diagnostic technique. Implementing failures of different types and sizes it is possible to determine thresholds for diagnostic parameters for each bearing component: inner and outer rings, rollers/balls and radial air. As VibropassportTM parameters have relative scale irrespective to the bearing size, experimentally determined thresholds could be applied for other bearing types. Figure 4b illustrates thresholds and actual values of diagnostic parameters measured within demonstrational experiment with testing faults implementation.

The balancing rig (Fig. 4a) plays its role in experiments and demonstrations related to aero- and mass imbalance modification.

Demonstrator for drive train and tail rotor diagnostics. The drive train and tail rotor diagnostic demonstrator (Fig. 4b) includes the actual drive train with supporting structure and the tail rotor. Supporting structure is able to simulate flight deformations of helicopter tail boom and its influence on operation conditions of the drive train. Depending on demonstration task the shaft or bearings could be defected or supporting structure could be modified to simulate operational conditions. All bearing supports of drive train are equipped by accelerometers and the tail gearbox is equipped by 3-axial accelerometer and indicator of rotation. Demonstration case of tail gearbox diagnostics could be carried out if artificial failure would be implemented. Also the demonstration of aero- and mass balancing of the tail rotor is available.

Demonstrator for main gear box and main rotor diagnostics. The most complex demonstrator (Fig. 4c) includes the main gearbox and the main rotor for on-line diagnostics of both. Electrically driven gearbox allows demonstration of ordinary and epicyclical gears diagnostics. The demonstrator has own harvesting system for power supply of data measurement and acquisition unit as well as sensor preamplifiers.

The sensors embedded in the rotating blades provide comprehensive data about blades oscillations. Operational Modal Analysis techniques applied to above data allow identification of blades dynamic properties and detection of its modification.

3.2 Success Stories of Vibropassport^{TM} Application

The successful stories of advanced system application play important role for potential customer to realize the potential and benefits when such cases occur in natural circumstances.

For instance, aiming to monitor bearings condition VibropassportTM technique was applied to turboprop engines within maintenance routines. Once, the diagnostic parameter surveying radial gap of bearings showed tenfold growth in the turbine bearing in comparison with threshold and parameters of other engines (Fig. 5). Seven months later the engine collapsed (after landing).

One more case is the drive gear misalignment occurred at engine assembly in the engine shop, where the system estimates operation quality of all engine gears during acceptance test. Diagnostic gear loading parameter (GLP) indicates how both interacting gears influence on dynamic loads of each gear.

On the diagram (Fig. 6) the green dotted line shows GLP threshold as average value around a fleet of this engine type. For normally operating gears threshold is about 0.6–0.7 and if GLP of some gear exceeds 1.0 it indicates the problem in this gear. Figure 6 shows GLP of all gears driven by gas generator for two particular engines. The “good” one has GLP of all gears below threshold (green bars), whereas the “bad” one (red bars) had exceedance of three central gears (driven gears, gear H1 and gear HD). Subsequent disassembly of “bad” engine allows discovering some violations of requirements to central drive assembly, like non-roundness, clearances etc. Actual tracks of diagnosed gears abnormal operation are shown on Fig. 6b.

4 Conclusions

Based on the advanced system tasks, authors considered the set of main requirements for health and structural monitoring system of a helicopter. Based on above the architecture of new system is developed, which covers all critical helicopter units. Analysis of available hardware and diagnostic techniques shows shortage of latter. However, there are new techniques allowing detection latent faults in operating engines and monitoring even rotating structural parts of a helicopter. Some samples of practical application of advanced diagnostic techniques illustrate its capabilities and high resolution. It was concluded about necessity to promote new diagnostic techniques to manufactures and MRO using special demonstrator.

Acknowledgements. The paper uses materials related to research 1.2.1.1/16/A/008 “The Study of Dynamic Properties of Aviation Engine Aggregates for Vibration Diagnostic Technology Research and Demonstrator Development” of the project “Establishment of Engineering Systems, Transport and Energy Competence Centre” in cooperation with Central Finance and Contracting Agency of Latvia. This work contains the results of the research and the development of the technical requirements for engine aggregates testing rigs.

References

1. HUMS Extension to Rotor Health Monitoring: CAA Homepage https://publicapps.caa.co.uk/modalapplication.aspx. Accessed 15 Aug 2017
2. Greaves, M.: Towards the next generation of HUMS sensor. In: ISASI 2014 Seminar, Adelaide, Australia (2014)
3. Mironov, A., et al.: Advanced vibration diagnostic techniques for overhaul costs saving of helicopter engines. In: Proceedings of the 42nd European Rotorcraft Forum 2016, 3AF, Lille, France (2016)

Evaluation of the Risk Management Issues in the Seaports of Latvia and Lithuania

Darius Bazaras(✉) and Ramūnas Palšaitis

Transport Engineering Faculty, Vilnius Gediminas Technical University,
Plytinės 27, 10105 Vilnius, Lithuania
{darius.bazaras,ramunas.palsaitis}@vgtu.lt

Abstract. This paper reviews the of transportation risk assessment in the Latvia and Lithuania ports. The essential steps of assessment are the identification of primary criterions, the determining the underlying criterion groups and quantitative assessment of international environment indicators and nature, infrastructure indicators and organizational indicators. Designed system of criteria creates a possibility for objective evaluation of risk management processes and allows planning objectively long-term risk management strategy in the Latvia and Lithuania ports according to certain economic development circumstances.

Keywords: Ports · Risk analysis · Planning · Risk management · Logistics
Hazard · Project · Incidents

1 Introduction

Enlargement of international business and economic relations between countries are directly connected with the growths of international cargo transportation. Ports located in the Latvia and Lithuania act as a mediator in the expansion of trade connections of the most important transportation corridors between East and West.

Significance of the biggest Latvia and Lithuania ports as transport system hubs can be extended if it will be generated as safe and more favourable environment for business transforming ports as transit points to the logistics services [4]. Transportation security, cargo safety and cost are the main factors influencing transit flows territorial distribution and stability. In pursuance to enhance the Baltic states importance in the European business context it must be created more liberal conditions for the transit cargo transportation through the Latvia and Lithuania ports and the shaping of public policies and plans that either modify the causes of disasters or mitigate their effects on people, property, and infrastructure.

Risk assessments are very important to both logistics operators and Rescue Services and other Civil Protection, such as Customs, Border Guard, Coast Guard and Police and environmental agencies. Effective transportation and the transportation safety risk management in transport hubs can increase interoperability in transporting goods and persons in North–South and East–West connections based on increased capacity of transport and logistics actors.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_14

For these reasons there is necessity to make the Latvia and Lithuania ports with hinterland infrastructure complex activities and service quality analysis in parallel with the analysis of the improvement of the interoperability of resources in case of emergencies and created the model for an all involved actors resource and risk management in emergency situations.

2 International Project HAZARD as the Foundation for Scientific Research and Data

The problem aspects analysed in this publication cover diverse types of risks and factors pertaining to accidents in Latvia and Lithuania seaports. The ongoing international project HAZARD of the BSR INTERREG program provide favourable conditions to assess situations under analysis, data collection and formulation of respective scientific assumptions [8]. To briefly present the project, it is worth noting that ports, terminals and storage facilities are often located close to residential areas, thus potentially exposing a large number of people to the consequences of accidents. The HAZARD project deals with these concerns by bringing together rescue services, other authorities, logistics operators and established knowledge partners.

HAZARD project aims at mitigating the effects of emergencies in major seaports in the Baltic Sea Region. The types of safety and security emergency include, for example, leakages of hazardous materials, fires on ships at port, oil spills in port areas as well as explosions of gases or chemicals.

HAZARD brings together Rescue Services, other authorities, logistics operators and established knowledge partners. HAZARD enables better preparedness, coordination and communication, more efficient actions to reduce damages and loss of life in emergencies, and handling of post-emergency situations by improving:

- harmonization and implementation of safety and security codes, standards and regulations;
- interoperability of resources through joint exercises;
- communication between key actors and towards the public;
- the use of risk analysis methods;
- adoption of new technologies.

The project duration is 36 months (spring 2016–spring 2019) and the total budget is 4.3 M€, which is partly funded by the EU’s INTERREG Baltic Sea Region Programme.

3 Conceptual Provisions and Formalized Foreground Multiple Criteria Analysis for the Potential Emergency Situations Risks Management

The complete risk analysis and assessment methods in mitigating accident risks in seaports and areas adjacent to these management life cycle includes the shaping of public policies and plans that either modify the causes of accident risks or mitigate their effects on people, property, and infrastructure.

The most common and relevant phases of emergency situations are mitigation, preparedness, response and recovery. The total accident risk in the port and ports hinterland management cycle includes prevention mitigation, preparedness, response/ recovery and rehabilitation/ reconstruction phases. These phases are described in flood risk management cycle [6] and disaster management cycle [5].

The mitigation and preparedness phases in the ports occur as risk management improvements are made in anticipation of a risk event. Developmental considerations play a key role in contributing to the mitigation and preparation of a port liable structures and port municipality government to effectively confront a disaster risk. As a disaster risk occurs, disaster management actors, in particular humanitarian organizations become involved in the immediate response and long-term recovery phases.

There are now quite many EU policy areas on built environment contributing to disaster risk management. EU Regulations and Directives take the major part of these EU strategies, legislation or programmes that are related to the disaster risk management and Member States (one of them is Lithuania) must comply with these legislative provisions [7]. The programs will support multimodal transport safety issues including protection from emergencies and accidents (including hazardous substances) associated with transport to reduce risk to human life and environment [1, 8].

During the analysis of transit transport flows distribution through ports it must be considered the protection from emergencies and accidents and quality of transportation services in the TEN-T ports, whosoever particularly influence transit flows territorial distribution.

At the macro level transit cargo transportation risks in the East–West transport corridor must be analyzed from the dimension of ports global competition.

On the basis of the formulization emergency risks components and algorimization of the quantitative evaluation process it is possible to solve optimization problems. During the formulization emergency situations risks factors it must be determined dependence from the whole influencing indicators, including their action directions such as port and hinterland infrastructure, human resources activities, environment and natural factors, global economic transformations. Multi-criteria system for choosing effective risk management in the ports could be used.

Evaluation of emergency situations risk management must be connected with the criteria system, which allows to identify optimal risk management strategy [2].

For practical purposes it is often expedient to use digital characteristics instead of the random factors' distribution laws. Although the digital characteristics give insufficient information on random factors, they fully suffice for the solution of risk management issues, and their determination is far easier. A complete analysis and synthesis of the characteristics of the risk management processes is carried out according to general characteristics, i.e. according to the conditional and unconditional distribution laws. These laws may be employed for the definition of different characteristics of the technological port operational processes.

4 Analysis or Emergency Incidents in the Latvia and Lithuania Ports

Investigation of transport flows through the BSR ports indicated that is essential conditionally resolve primary emergency risk indicators representing internal and external macro environment component into three groups: environment and nature, infrastructure and economic-organizational indicators [3].

Vilnius Gediminas Technical University in 2017 performed survey of incidents in the BSR ports. The questionnaires consisted of two parts: one part with questions about incidents in the port (for the ports of HAZARD counties project partners), and another part with specific questions concerning the infrastructure indicators. The questionnaire was twice distributed between Hazard project partners but filled questionnaires were got only from Latvia and Lithuania. Gain data are enough informative but they are underused for quantitative evaluation and to formulate respective evaluation models for emergency situations risks.

Preliminary data analysis showed that level of organization and coordination of transport flows in the ports and risk management in the Latvia and Lithuania is organized enough good (Table 1).

Table 1. Number of incidents in the Latvia and Lithuania ports 2011–2016.

	Latvia	Lithuania
Fire/explosion	1	0
Collision between commercial vessel and leisure or fishing vessel	1	0
Tug girding	2	0
Collision between leisure vessels	0	1
Bunker barge damaged on berthing	0	12
Commercial vessel in collision in approach channel	1	0
Vessel drags anchor and collides with another	1	0
Oil split from bunker operations	2	1
Petrol fire in small craft	1	0
Divers run over whilst surfacing	0	1
Tanker grounds within port limits	1	0
Commercial vessel drags anchor and grounds	0	1

Klaipeda port during the year is serving about 7 000 ships with total cargo turnover 40.14 mln. tones. Riga, Venspils and Liepaja ports are serving from 9 000 to 10 000 ships with total cargo turnover 63.8 mln. tones. The biggest part of goods transported through the ports is transit goods: Klaipėda port – 41%, Latvia ports –57.8% (oil products, fertilizers, coal.) which belong to dangerous goods categories.

In the Latvia and Lithuania ports direct responsibility for the cargo safe loading, reloading and transportation security are responsible governmental Fire and Rescue institutions and Seaport Authorities. During the last 5 years there were absence incidents in the ports with dangerous goods loading operations and transportation. It shows good safe loading and transportation control in the ports.

As it was mentioned previously, the qualitative analysis on the obtained results is hardly possible due to lack of respondents' activeness and adequate preparedness to take part in the survey. By conducting additional interviews with respondents and analysing the survey process and the selected methodology, it was observed that key issues are related to the following circumstances:

- availability of statistical data and its proper systematization;
- competencies of the personnel, job description as well as possibilities to dispose the systematized statistical information related to incidents in the port;
- access to information is restrained and blocked for commercial secrecy or security reasons.

While conducting the survey it was once again observed that seaports are objects of strategic significance, not always willing to disclose information pertaining to incidents and its causes. Nevertheless, it is possible to analyse the obtained data presented in Table 1.

Firstly, it is important that whilst considering extremes such as "Bunker barge damaged on berthing" mentioned in the port of Klaipeda for 12 times, it is possible to make an assumption that the total number of incidents in the ports of Latvia and Lithuania is similar – 10 in Latvia and 16 in Lithuania, respectively. Additionally, it is important to consider the fact that statistical data used in the survey was obtained from Latvian seaports and Lithuania is represented by statistics date from Klaipeda seaport. This is a justifiable and explainable fact, as there are no other commercially operating ports except for Klaipeda State Seaport. It is possible to claim that a real threat associated with handling operations in Latvia and Lithuania is linked to oil split from bunker operations. The risk of oil and its products' spillage can be identified as one of the most realistic one. Therefore, it is needed to strengthen the appropriate preventive measures to minimize the level of the risks at sea ports. Other incidents are distributed equally, but are not mentioned in ports of other countries. As it was mentioned, it is thus purposeful to draw attention to the incidents of bunker barge damaged on berthing, which are particularly frequent in Klaipeda. Additionally, it is worth noting the cases of the tug girding which taking place in Latvia. To summarize, it is possible to claim, that distribution of incidents in the seaports of Latvia and Lithuania is unequal, reflecting the specifics of handling operations and freight processing.

5 Theoretical Background for the Transportation Risk Assessment in the Latvia and Lithuania Ports Criteria Evaluation

It is necessary to analyse and to validate methods for quantitative evaluation and to formulate respective evaluation models (practice equations) for emergency situations risks.

The detailed analysis systemic publications of various authors permits to take into attention those of the evaluation methods' which may be potentially used more widely for the determining the compound magnitude. First of all we selected the evaluation methods group, as the most adequate to the formulated tasks. The *Simple Additive*

Weighting (SAW) and *Complex Proportional Assessment (COPRAS)* methods attached to this group are mostly used for the determining of the compound magnitude. Their peculiarities may be revealed by the specifies in the formation of criteria system evaluation, the determination of their criteria significance and the evaluation of the research object on this basis.

This corresponds to an offered three-stage compound quantitative assessment system on the basis SAW method, which enable to convert indicators primary indicators expressions to their assessment in 10 points scale (decimally). It is realized in the following consequence:

- The identification and expertise (quantifiable) assessment of primary indicators determining the selected groups, assessment their significances as well as ranking of primary indicators;
- The determination compound indexes of indicator groups as partial criteria for assessment generalize criterion – international environment index;
- The determination index international environment (as a composition of indicator groups) as a consolidated measure.

The expertise of the primary indicators and their weights using the provided technique is treated as a first stage of quantitative assessment. After all, this evaluation of emergencies risk evaluation system in the ports (applying quantitative methods and evaluation process algorithms) may be incorporated into the general transportation management system.

6 Conclusions

1. Latvia and Lithuania ports are playing significant role in the serving transit transport flows in the East –West transport corridor. The biggest part of goods transported through the ports is transit goods: Klaipeda port – 41%, Latvia ports –57.8% which belong to dangerous goods categories.
2. Ports activities management system interfaces with the indicators connected with the transportation safety and security in the port and are analysed insufficiently.
3. The creation of the methodology of the assessment emergency situations risks in the ports will support multimodal transport safety issues including prevention and elimination of undesirable events (failure, accident, collision, disaster), minimization of their appearance risk and mitigation of their consequences.
4. A very low response rate was encountered whilst carrying-out the research on the incidents at the Baltic Sea ports in order to explain the availability of statistical data and its adequate systematization activities in corresponding seaports. Moreover, personnel competencies, job descriptions and the availability of systematic statistical information related to incidents in the port caused the low percentage of responses. In addition, in view of the recent increase in the risk of terrorist attacks, the access to information is restricted and blocked for commercial confidence or security reasons.

5. The total number of incidents in the ports of Latvia and Lithuania is similar – 10 in Latvia and 16 in Lithuania, respectively. It is possible to claim that a real threat associated with handling operations in Latvia and Lithuania is linked to Oil split from bunker operations. The risk of oil and its products' spillage can be identified as one of the most realistic one, thus port authorities must provide appropriate preventive, and liquidation measures to minimize the level of the risk. Other incidents are distributed equally, but rarely occur in ports of other countries; however, they do reflect the specifics of handling operations and freight processing. A separate issue lies in the risks associated with new cargo, such as Liquid Gas, which loading and storage operations have started to be performed at Klaipeda Seaport.

References

1. Batarlienė, N.: Risk analysis and assessment for transportation of dangerous freight. Transport **23**(2), 98–103 (2008). https://doi.org/10.3846/1648-4142.2008.23.98-103
2. Bazaras, D., Palsaitis, R.: The impact of the market structure on safety and security in BSR: Lithuania point of view. In: Proceedings of the 11th International Conference Reliability and Statistics in Transportation and Communication (RelStat 2011), 19–22 October 2011, pp. 173–175. Transport and Telecommunication Institute, Riga (2011). ISBN 9789984818467
3. Bazaras, D., Palšaitis, R.: Factors of emergency situations risks management in the BSR ports. In: RelStat-2016: Proceedings of the 16th International Scientific Conference Reliability and Statistics in Transportation and Communication, Procedia Engineering, vol. 178, pp. 501–506. Transport and Telecommunication Institute, Riga (2016). ISSN 1877-7058
4. Kabashkin, I.: Freight transport logistics in the Baltic Sea region. Regional aspects. Transp. Telecommun. **13**(1), 33–50 (2012). https://doi.org/10.2478/v10244-012-0004-x
5. Palukuri, S., Jain, P.: Operations in Disaster Management – Case Study Based Approach (2014). http://www.mbaskool.com/business-articles/operations/9351-operations-in-disaster-management-case-study-based-approach.html. Accessed 16 Dec 2015
6. Van Veen, B.A.D., Vatvani, D., Zijl, F.: Tsunami flood modelling for Aceh & West Sumatra and its application for an early warning system. Cont. Shelf Res. **79**, 46–53 (2014)
7. White Paper Roadmap to a Single European Transport Area - towards a competitive and resource efficient transport system. Commission of the European Communities, Brussels (2011). ISBN 978-92-79-18270-9
8. Project HAZARD. Mitigating the Effects of Emergencies in the Baltic Sea Region Ports. INTERREG Baltic Sea Region programme project 2016–2019 (2017). https://blogit.utu.fi/hazard/

Reliability, Safety and Risk Management

Advanced Structural Health Monitoring and Diagnostics of Transport, Industrial and Energy Facilities

Aleksey Mironov[1(✉)], Deniss Mironovs[1], and Igor Kabashkin[2]

[1] Vibroakustikas laboratorija, Riga, Latvia
info@viblab.lv
[2] Transport and Telecommunication Institute, Riga, Latvia
kiv@tsi.lv

Abstract. The condition-based maintenance (CBM) strategy utilizing continuous monitoring with proposed Combined Sensor Network (CSN) technology is discussed in this paper. The paper introduces a decision-making approach for design of optimal integrated monitoring and diagnostic architecture that combines maintenance decisions with CSN-based diagnostic system and its integration with Structural Health Monitoring (SHM) approaches for achieving efficient system for maintenance and lowering life-cycle cost of industrial structures. The technology prototype is tested in laboratory and real-life conditions on a 20 m tall airport radar tower in Riga, Latvia.

Keywords: Multi-patch operational modal analysis
Structural health monitoring · Industrial structures · Structural diagnostics

1 Introduction

The goal of structural health monitoring (SHM) is to control the condition of a structure. Understandingly, SHM should not influence normal functioning of the monitored structures, which means that control, monitoring, inspection and other activities must be performed at the structures location, without removing it from its position, e.g. performing measurements directly on an airport radar tower, while in operation, which is one of the subject of this research study.

Some SHM methods are based on fibre-optics (Fibre-Bragg grating), electrical-resistance of materials, acoustic emission [1]. Vibration-based methods are also very promising thanks to their high sensitivity, versatility and reliability [2]. The focus of this paper is on the vibration-based method of SHM, namely Operational Modal Analysis (OMA) [3, 4].

As with any type of modal analysis, OMA aim is to determine systems modal characteristics or modal parameters – mode shape ψ, frequency f and damping ζ. These parameters are related to mechanical properties of the system and accordingly to its condition, and if somehow the system and its condition changes then modal characteristics are modified as well. Based on that, OMA is becoming more popular for SHM of massive civil objects, like bridges and towers. The monitoring technology presented in

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_15

this paper is called OMA based SHM. Theoretical foundation of SHM application feasibility is also presented in this paper.

2 Existing Problems and Possible Solutions

There are certain limitations in practical OMA that restrict its further expansion to SHM of industrial objects. These limitations are:

Environmental:

- Applied excitation force must not contain any periodic components,
- Energy level of the excitation must be adequate to excite lower order modes.

Technical (transducers allocation and protection, data collection), Financial (cost of transducers, data transfer and acquisition units).

Mentioned problem of SHM application for objects having definite periodic excitation requires certain solutions that are presented in this paper. For instance, periodic extraction tool is designed to separate measured deformation signals into modal response of the structure and periodic vibration of the rotating parts on the structure. A successful attempt to implement thin piezo film deformation sensors is made in order to simplify transducers mounting and cost, based on work in [5].

OMA approach requires a network of sensors distributed around the structure in order to perform effective analysis. The number of sensors depends on the structures complexity and can vary from tens to hundreds. Normally, one would need the same number of input channels on a data acquisition unit (DAQ). These units can be really expensive, especially with mentioned amount of input channels. In order to save costs, multi-patch OMA approach is suggested. This gives opportunity to sequentially measure all sensors with a limited amount of input channels, e.g. 8 input channels for 28 sensors.

Excitation in OMA is usually ambient and is not controlled or measured. Engineers have to rely on natural circumstances, like wind, rain etc., and unrelated human activity, like road traffic, construction works and so on, to excite the studied object well enough. With advanced methods in [6] it is possible to monitor excitation quality in spectral and energy terms to validate measurements and get satisfactory modal response of the studied object.

3 Piezo-Film Sensor Based Measurement Chain

3.1 Sensor

This section is devoted to piezo-film based transducers (Fig. 1), specially designed for OMA based SHM. The sensing part is piezo film element that reacts to extension or compression along its axis with a variable electric charge. The charge is then amplified and formed into a differential current signal in the preamplifier. Due to some production factors, as well as uncertainty of electronic components in the preamplifier, the sensors may have different output sensitivity. This could be a problem if one wants to measure

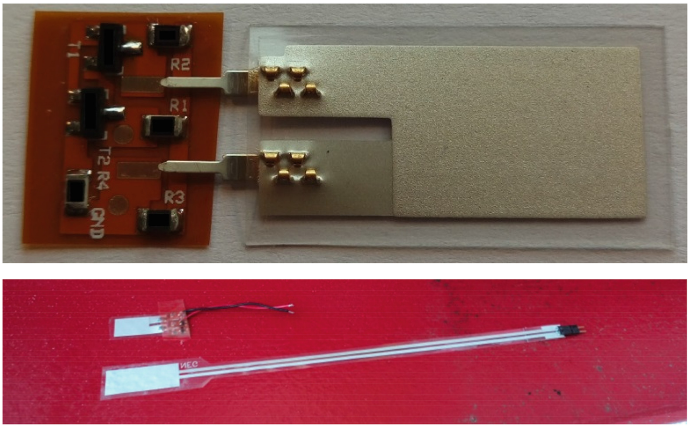


Fig. 1. Piezo-film sensors, short and long options. Long option has connector for the preamplifier.

absolute deformations, but when it comes to relative measurements, which is the case in SHM, different output sensitivity does not influence the results at all.

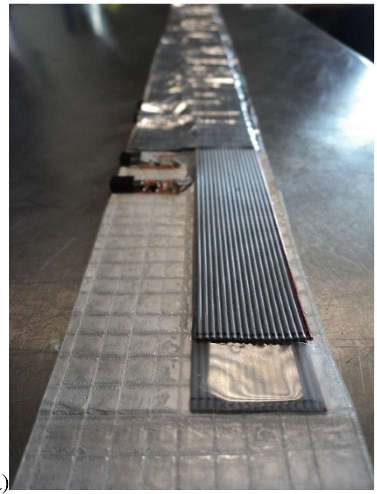
The sensors are placed and glued on a measured structure's surface via strong double sided tape with good elastic properties. Elasticity allows surfaces deformations to be transmitted to the sensing element without filtering out important frequency ranges.

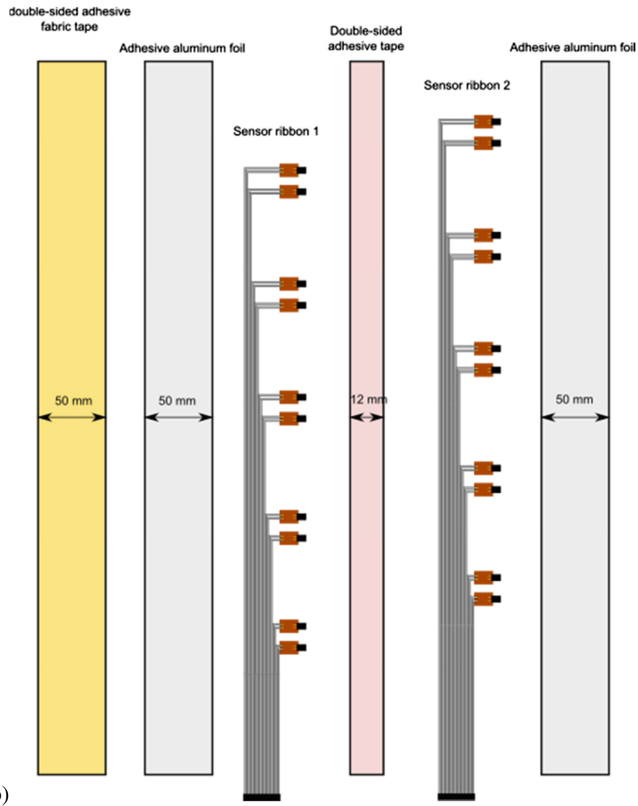
The sensor is Integrated Electronics Piezo Electric (IEPE) type transducer, similar to ICP sensor, require power supply for the preamplifier. Supply current of 4–20 mA is sent via the positive and negative leads and the sensors output voltage is around 6–15 V if paired with traditional DAQ like Bruel & Kjear LAN-XI module.

These sensors have serious advantage compared to traditional accelerometers – piezo film sensors are rather cheap and thin, which makes them a really practical tool in OMA based SHM.

3.2 Cable for Sensors Networking

Flat ribbon cables (FRC) are used to transmit signals and power supply between sensors and DAQ. For each structure to be examined cable dimensions are designed in such way, so it is well integrated in a presumably complex architecture of frames and barriers. Cable design considers planned sensors locations on the structure and preamplifiers are placed accordingly on the cable. A cable concept for application on actual radar tower (Fig. 2a) and an example of a real layout (Fig. 2b) are shown in Fig. 2.


a)



b)

Fig. 2. Measurement cable assembly. (a) Concept assembly; (b) layout for airport radar tower.

Cable is a multi-layer structure. The base is a double sided adhesive tape for connection to the structures surface. The second layer is metallic foil for shielding from powerful electro-magnetic interferences (EMI). On top of that rest one or two FRC, with wires cut precisely to match the position of preamplifiers. These wires ends are soldered to the preamplifiers output. In case of two FRC cables (if larger amount of sensors is used), they are glued together using double sided adhesive tape. The assembly is then covered with another layer of metallic foil to close the EMI shielding. The sensors are also covered with a shield.

The output of the cable is terminated using traditional D-sub connector. To establish multi-patching, a commutation switchboard is designed in such a manner, so that separate sensors or sensor groups could be connected to designated DAQ inputs. This switchboard allows connecting any configuration of sensors to a limited amount of input channels. For some purposes, if there are enough input channels, it is also possible to connect all available sensors into DAQ at the same time.

4 Periodic Component Extraction

4.1 Problem Formulation

It is important to eliminate periodic components (harmonics) before performing modal analysis, because these can be mistaken for structural modes during signal analysis. Also, periodic components can mask structural modes in frequency range where the former dominate. Of course, the best way to deal with harmonics is to block them from being measured. Generally it means shutting down rotating machinery on the structure. Obviously, this is undesirable, especially in energetics (power plants, electric motors) or navigation (airport and naval radars). So there must be a post-processing tool to suppress or extract harmonics before performing analysis of vibrational data.

4.2 Theoretical Basis

Let systems frequency response be presented as $H(\omega)$, in which all necessary information about a system (e.g. any type of structure) is stored. The stochastic part of a response can be shown as

$$X(\omega) = H(\omega)F(\omega), \tag{1}$$

where $F(\omega)$ is the excitation force and $X(\omega)$ is the measured signal, without presence of any deterministic components. Note that in case of traditional OMA $F(\omega)$ (and its inverse Fast Fourier transform $f(t)$) is not measured nor controlled and it is assumed that the excitation force is spectrally uniform (white noise) and evenly distributed around the system. Often in reality this is not the case, which constitutes one of the limitations mentioned in Sect. 2.

Deterministic part in a simplified way, can be presented as

$$D_n(\omega) = D_n \mathrm{e}^{\mathrm{j}\omega_n t + \phi_n}, \tag{2}$$

where n denotes one of N deterministic components ($n = 1, 2, \ldots, N$), and ω_n is the rotation speed of rotor or fundamental frequency of a periodic component. It is important to have periodic components frequency time function $\omega_n(t)$, e.g. recording of a periodic component, as it can fluctuate in time and influence further signal processing.

Further extension of Eq. 2 is

$$\begin{aligned} D_n(\omega) &= D_{n1}\mathrm{e}^{\mathrm{j}\omega_{n1}t} + D_{n2}\mathrm{e}^{\mathrm{j}\omega_{n2}t} + \ldots + D_{nK}\mathrm{e}^{\mathrm{j}\omega_{K}t}, \\ D_n(\omega) &= \sum\nolimits_{k=1}^{K} D_{nK}\mathrm{e}^{\mathrm{j}\omega_{K}t}, \end{aligned} \tag{3}$$

which basically means that periodic part is superposition of numerous harmonics. In case there is more than one periodic part then these parts can create intermodulation that results in extra harmonics,

$$D_{ne}(\omega) = \left(\sum\nolimits_{k=1}^{K} D_{nK}\mathrm{e}^{\mathrm{j}\omega_{K}t}\right)\mathrm{e}^{\mathrm{j}\omega_0 t}. \tag{4}$$

Altogether, the periodic component of a vibrational signal on a structure with different machinery and/or EMI noise consists of separate polyharmonical periodic parts and their intermodulated harmonics, as shown in Eq. 4.

The equation for harnessed signal with periodic contamination is given as

$$X(\omega) = H(\omega)F(\omega) + D_n(\omega). \tag{5}$$

Now assume there are some experimentally obtained $D_{na}(\omega) \approx D_n(\omega)$. Extracting $D_{na}(\omega)$ from (3) will result in

$$X_p(\omega) \approx H(\omega)F(\omega), \tag{6}$$

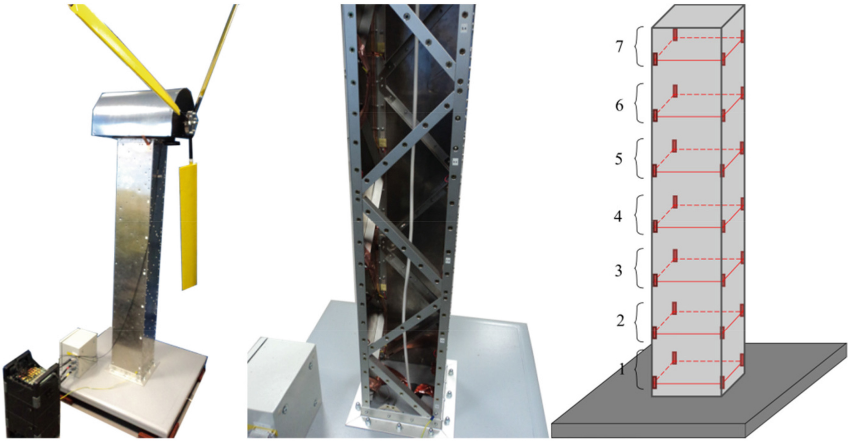
and

$$x_p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_p(\omega)\mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}\omega, \tag{7}$$

where index p means "processed". Most commonly the signal enhancement technique is applied to extract periodic components, i.e. to obtain $D_{na}(\omega)$. The enhancement process based on fundamental frequency like rotation speed or other allows detection of all related signal components. As typical machine has more than one fundamental frequency enhancement process is repeated for each one of it. The resulting accumulated vector $d_{na}(t) = d^1_{na} + d^2_{na} \ldots$, with the same length as $x(t)$, is subtracted from $x(t)$ and $x_p(t)$ is obtained.

4.3 Laboratory Model

The topic of periodic component extraction shall be explained on a real-life example, namely scaled wind generator model (Fig. 3). It is equipped with a fully-functioning rotor with blades, driven by an electric motor. This is placed on top of the tower made of aluminium. The tower is firmly fixed on a heavy concrete base, which is considered to be a rigid boundary condition. Experiments were performed using 2.7 Hz rotation speed.



a) b) c)

Fig. 3. Wind generator laboratory model 1.4 m tall; (a) common view, (b) sensor network installed inside with power cable for motor, (c) sensors placement.

The signals, picked up by transducers, are the towers modal response to stochastic and deterministic excitation. Stochastic source is ambient excitation but deterministic one are generated by

1. Rotor induced vibration (2.7 Hz and harmonics) – $D_a(\omega)$,
2. Electro-magnetic interferences (50 Hz and harmonics) – $D_b(\omega)$,
3. Electric motor vibration (100 Hz and harmonics) – $D_c(\omega)$,
4. Intermodulation between all of them.

Parts 2. and 3. are interconnected, but should be regarded separately for this task, which is separation of stochastic and deterministic parts. Stochastic component is then the subject of modal analysis techniques [4]. It is worth mentioning that deterministic part is also very useful for rotary machinery diagnostic purposes. Relating to Eq. 3 the full deterministic part can be shown as

$$D_n(\omega) = D_a(\omega_a) + D_b(\omega_b) + D_c(\omega_c) + D_a(\omega_a)\mathrm{e}^{\mathrm{j}\omega_b t} + D_a(\omega_a)\mathrm{e}^{\mathrm{j}\omega_c t}. \quad (8)$$

This paper only deals with first 3 terms of Eq. 7. Each accumulated vector of a periodic part $d_{na}(t)$ has to be obtained separately.

4.4 Experiment

Proposed periodic component extraction technique has been applied on experimentally obtained vibrational data from the wind generator model. The test excitation was automatic, meaning that no manual excitation like hammer impacts was used. Instead, two small dummy weights were fastened around the rotor shaft with 1 mm steel cable on top of the model. Weights rotate together with rotor, hit the surface of the towers head, drag along this surface and lift up. The cycle continues. The signal is formed from the impacts, which are supposed to invoke towers modal response, and all deterministic parts, mentioned in the previous section.

Empirically it was established that extraction of periodic parts should be performed starting from the part with the most energy and finishing with the part with the least energy. Energy levels of periodic components are easily evaluated from frequency spectra of the measured signals. For this case, first the 100 Hz with harmonics was extracted, then 50 Hz with harmonics (taken away what remained from 100 Hz as well) and lastly 2.7 Hz with harmonics. Again, inter-modulated components were not treated in this study.

The results of applied periodic extraction algorithm are presented in Fig. 4. Top left plot shows time series of a half-second piece of typical 120 min record. To the right is the averaged FFT spectrum of this recorded signal. Bottom line shows the same for the corrected signal, i.e. without deterministic parts. It can be observed that strong harmonics were removed at 2.7, 5.4, 10.8 Hz etc., 50 Hz and 100 Hz. This is seen both in time and frequency plot. Unfortunately, some deterministic components remained, e.g. at 97.3 Hz, 94.6 Hz or 102.7 Hz and 105.4 Hz. These spectral components are the products of 100 Hz modulated by 2.7 Hz rotating inertial forces. Extraction of those signal components needs more dedicated processing and will be dealt with in future studies.

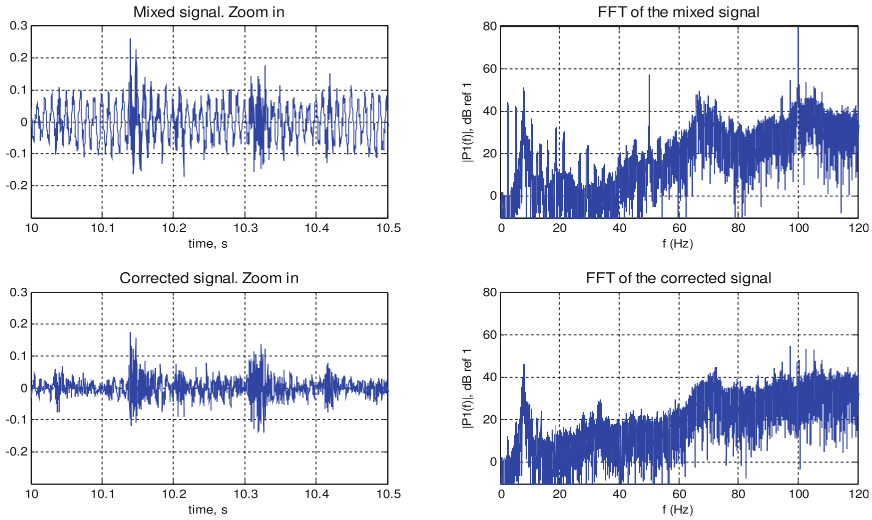


Fig. 4. Signal with and without deterministic parts.

Still, even with modulated harmonics, the signal became "cleaner" and is now suitable for modal analysis processing.

5 Airport Radar Tower

The technique, described in this Sect. 4.2, was designed to be used for recorded deformation signals from airport radar tower (Fig. 5 right). Primal investigation of radars operating conditions hints on the following possible undesirable deterministic parts: radar antenna rotation and EMI interferences.

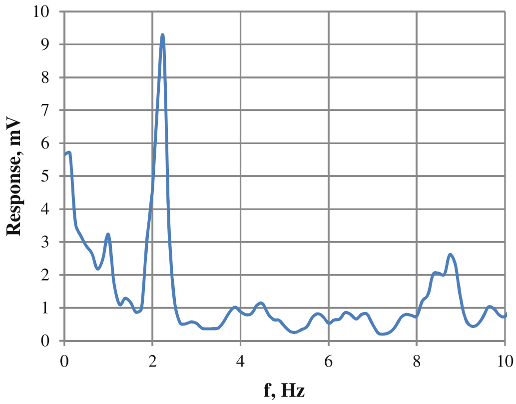


Fig. 5. On the left – frequency spectrum of the tower deformation signal from a single sensor. On the right – airport radar tower 20 m tall. Radar antenna on top. Steel frame.

From the presented frequency plot (Fig. 5 left), one can see a harmonic at 1 Hz, which is rotation frequency of the antenna. From measurements it was observed, that there is also a second harmonic at 2 Hz and presumably first structural mode at 2.25 Hz, which masks 2 Hz harmonic in the plot (Fig. 5 left).

Some amount of EMI noise was present, especially at 50 Hz, which is expected. In future studies, recorded deformation signals from radar tower shall be processed using the periodic extraction technique and used for modal analysis.

6 SHM in Automatic Excitation Conditions

OMA based SHM was successfully applied on the wind generator model in [5, 7]. As a continuation of that work for this paper, it was decided to check robustness of the method by drifting away from OMA technical requirements, namely uniform frequency spectrum of the excitation force and even distribution of the excitation around the studied structure.

A series of OMA tests were performed on the laboratory wind generator model. The excitation was provided in automatic mode, as described in Sect. 4.4. As was expected, the modal parameter estimation quality was poor, with low coherence between sets of modes. Consequentially, estimated MPVI parameters could not show any correlation between objects state (with or without defect). Again, this reminds about limitation mentioned in Sect. 2 – excitation energy should be enough to get response from important modes.

7 Cost Effectiveness of Condition-Based Maintenance with SHM

Maximizing the effectiveness of SHM requires it to be designed for industrial structures on early stage of life cycle and integrated into construction of the object. Keeping the proper focus on SHM is achieved by health management integration, whose ultimate goal is to ensure an optimal SHM balance across the system, resulting in improved system safety, improved reliability and reduced life cycle costs. One express approach to the development of above mentioned SHM to provide condition-based maintenance for more efficiency of industrial structures life cycle on criteria reliability/ cost is described in this section.

The efficiency of SHM is assessed by the total cost of life cycle

$$C(U,t) = C_p(U) + C_o(U,t), \tag{9}$$

where $U = \{\pi_i\}, i = \overline{0,n}$ – Architecture of SHM with π_i channels (sensors) for diagnosis; $C_p(U)$ – cost for the development and production of SHM with U architecture; $C_o(U,t)$ – cost for the identification of failures in system with U architecture of SHM during time t of system operation.

The problem of design for integrated diagnostics we formulate as development of SHM with optimal architecture U_{opt} with minimal total cost during life cycle T_{LC}:

$$C\left(U_{opt}, t\right) = min\left\{\left(C(U,t)|t = T_{LC}\right)\right\}. \tag{10}$$

Every component, equipment, or system has an intrinsic design capability (Fig. 6). The demand or the expected value may be below this level. Over time, the curve of capability will decrease due to fouling, wear, fatigue, or chemical attack. When this happens, some maintenance has to be done to bring the capability up to the acceptable design level.

Unfortunately, not all parts of industrial objects covered by SHM. In this case periodical test actions are required for monitoring of part of the object not covered by the SHM. We are using the approach proposed in [8] to determine the optimal monitoring frequency of part of the system not covered by the SHM.

Markov's state transition diagram for such a system is shown in Fig. 7a, where: $\lambda = 1/T_0$ – failure rate of the system with mean time between failures T_0; $\mu = 1/T_R$ – repair rate with mean time between repairs T_R; $T_m = 1/\lambda_m$ – mean of exponentially distributed monitoring time with parameter λ_m; $\tau_m = 1/\mu_m$ – mean of exponentially

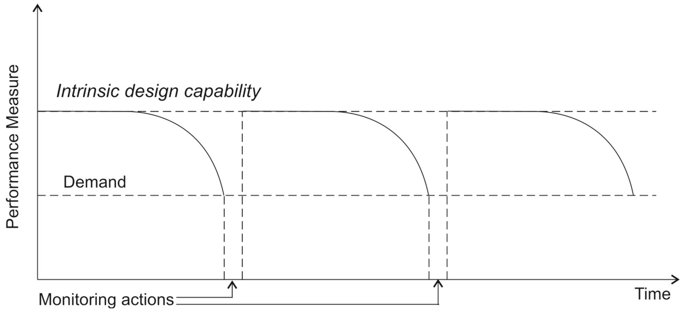


Fig. 6. Effect of demand fluctuations on maintenance timing.

distributed duration of maintenance with parameter μ_m; $a = \lambda_t / \lambda$ – part of system with failure rate λ_t tested by SHM ($0 \le a \le 1$).

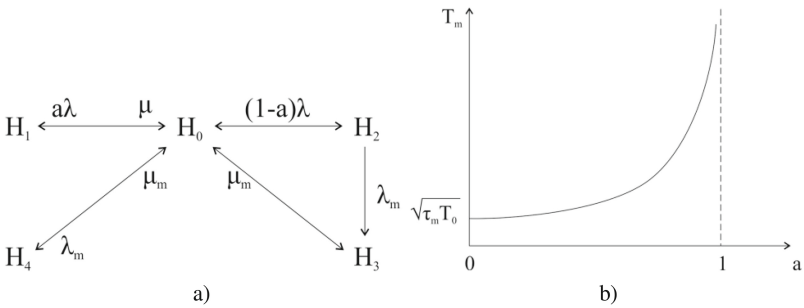


Fig. 7. (a) Markov's state transition diagram; (b) $T_m = f(a)$ function graph.

The behaviour of the examined system is described by the states transition (Fig. 7a): H_0 – state, in which system is operating and available for use; H_1 – failure is in the part of the system with monitoring of the built-in test equipment; H_2 – failure is in the non-monitoring part of the system; H_3 – the failure in the non-tested part of the equipment is being repaired during maintenance operations; H_4 – system is without failures and under maintenance.

On the base of the state transition diagram for a Markov's process shown on the Fig. 7, we can write the system of Chapman-Kolmogorov's equations according to the general rule for a stationary process [9]. After transformation the above-mentioned set of equations we can obtain the value for $P_i (i = 1 \ldots 4)$,

$$P_1 = \frac{a\lambda P_0}{\mu}, P_2 = \frac{(1-a)\lambda P_0}{\lambda_m}, P_3 = \frac{(1-a)\lambda P_0}{\mu_m}, P_4 = \frac{\lambda_m P_o}{\mu_m}. \tag{11}$$

Value of P_0 can be obtained by replacement $P_i (i = 1 \ldots 4)$ in the normalizing equation $\sum_{i=1}^{4} P_i = 1$:

$$P_0^{-1} = 1 + b, \tag{12}$$

where

$$b = \frac{a\lambda}{\mu} + \frac{1-a}{\lambda_m} + \frac{(1-a)\lambda}{\mu_m} + \frac{\lambda_m}{\mu_m}. \tag{13}$$

Let us define the unavailability U_a of the examined system. Unavailability can be defined as the probability that an item will not operate correctly at a given time and under specified conditions. Mathematically, unavailability is $U_A = 1 - A$.

For the examined system $A = P_0$ and

$$U_A = 1 - A = \sum\nolimits_1^4 P_i = \frac{P_0}{1+b}. \tag{14}$$

For a highly reliable system with $\lambda \ll \mu$, $\lambda_m \ll \mu_m$ approximately $P_0 \to 1$. In this case the system unavailability U_A is defined as follows:

$$U_A = (1+b)^{-1}. \tag{15}$$

By differentiating this equation with respect to λ_m, we get

$$\frac{dU_A}{d\lambda_m} = -\frac{(1-a)\lambda}{\lambda_m^2} + \frac{1}{\mu_m}. \tag{16}$$

Setting last equation equal to zero we get optimal value of monitoring time $T_{m\,opt} = \frac{1}{\lambda_{m\,opt}}$:

$$T_{m\,opt} = \sqrt{\frac{\tau_m T_0}{1-a}}. \tag{17}$$

The graph of function $T_{m\,opt} = f(a)$ is shown at the Fig. 7b. For the special case of a full monitoring system $a \to 1$ the monitoring time $T_{m\,opt} \to \infty$. It means that if the system has a hundred per cent built-in test equipment for SHM, the manual monitoring of such a system is not needed.

8 Conclusion

Presented work states main difficulties of OMA based SHM and gives adequate solutions to some of them.

First, the problem of periodic contamination of modal response is being solved by registration of rotation phase and extraction of mentioned periodic component from the measured signal. Further research has to be done to improve periodic extraction for intermodulated harmonics.

Second, technical and financial difficulties are overcome by introducing affordable and small piezo film sensors and multi-layer flat cables.

One of the factors that can bring uncertainty is the energy of excitation. One has to be sure that the force is suitable spectral-wise and there is enough of it. In any case, there are perspective methods [6] that can give information to engineer, when the excitation is good enough for OMA. It is of author's intuition that in field conditions, there will be enough energy (at least occasionally, when strong wind is blowing) to excite airport radar tower, as well as other possible objects.

A thorough rationale of SHM was also presented from the cost effectiveness point of view.

Acknowledgement. The paper uses materials related to research 1.2.1.1/16/A/008 "The Study of Dynamic Properties of Aviation Engine Aggregates for Vibration Diagnostic Technology Research and Demonstrator Development" of the project „Establishment of Engineering Systems, Transport and Energy Competence Center" in cooperation with Central Finance and Contracting Agency of Latvia. This work contains the results of the research and the development of the technical requirements for engine aggregates testing rigs.

References

1. Balageas, D., Fritzen, C.P., Güemes, A.: Structural Health Monitoring. Wiley, New York (2015)
2. Fritzen, C.P.: Vibration-Based Structural Health Monitoring – Concepts and Applications. Key Engineering Materials, vol. 293–294, pp. 3–20. Trans-Tech Publications (2005)
3. Parloo, E.: Application of frequency-domain system identification techniques in the field of operational modal analysis. Ph.D. thesis. Vrije Universiteit, Brussel (2003)
4. Chauhan, S.: Parameter estimation algorithms in operational modal analysis: a review. In: Proceedings of the 6th International Operational Modal Analysis Conference, pp. 277–287, Gijon, Spain (2015)
5. Mironov, A., Doronkin, P., Mironovs, D.: OMA techniques optimization for SHM of typical industrial structures. In: Proceedings of the 7th International Operational Modal Analysis Conference, Ingolstadt, Germany, pp. 124–127 (2017)
6. Mironovs, D., Chauhan, S.: Modal parameter estimation in multi-patch operational modal analysis: perspectives and approaches. In: Mains, M., Blough, J.R. (eds.) Topics in Modal Analysis & Testing, vol 10, pp. 107–126. Springer (2017)
7. Mironov, A., Mironovs, D.: Experimental application of OMA solutions on the model of industrial structure. In: Proceedings of 3rd International Conference Innovative Materials, Structures and Technologies, Riga, Latvia (2017)
8. Kabashkin, I.: Optimal monitoring strategies. In: Wiley Encyclopedia of Operations Research and Management Science, vol. 8. Wiley (2011)
9. Rubino, G., Sericola, B.: Markov Chains and Dependability Theory. Cambridge University Press, New York (2014)

Risk Management of Innovative Projects: New Aspects

Alina Konovalova(✉)

P.G. Demidov Yaroslavl State University, Sovetskaya, 14, Yaroslavl, Russia
alinakonovalova9@gmail.com

Abstract. Project-based approach is one of the most widespread approaches to promotion of innovations. Risk management in innovative projects is one of the most difficult tasks in project management due to high degree of uncertainty of innovative projects.

The mechanical connection of various methods of analysis and risk management described in risk management approaches often impossible due to their methodological disconnection.

The aim of the research is to improve the risk management system of innovative projects at all key stages (identification, assessment, risk mitigation) through the overcoming main disadvantages of current researches and interconnection of risk response procedures with the nature of risks and their quantification.

The improvements are mostly general and don't take into account sphere of innovations and stage of project life cycle. Moreover, the system of project risk management is usually unique because it depends on individual set of project risks. Generally the developed by author risk management system described in article operates the parameters, which are uniform in project planning; so it may be adapted to different areas of innovations.

The results of this study may be used for development of application software (in the form of an independent software product or in the form of an automated system module), which would implement a set of risk assessment procedures and determining the key measures to mitigate risks.

Keywords: Risk · Risk management · Innovative project

1 Introduction

In case of large-scale transformations in the Russian economy innovations are considered as the principal factor leading to quality changes in the business environment and economic growth.

Project-based approach is one of the most widespread approaches to promotion of innovations. And the task of innovative projects' risk management is one of the most difficult in project management due to significant degree of uncertainty of future result.

It is necessary to recognize insufficient illumination in the scientific literature of questions of risk management of innovative activity. In the approaches analyzed, risks are mainly identified by the risk object, so risks of a different nature are considered to be equivalent. The methods of risk estimation and risk mitigation are usually considered

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_16

without connection with each other what can lead to diversion of money from the company's turnover, serious losses and even to the curtailment of the project.

The mechanical connection of various methods of analysis and risk management is often impossible due to their methodological disconnection. In order to overcome the fragmented nature of scientific knowledge, the task is to form a methodological approach to assessing the risks of innovative projects.

So the aim of the research improve the system of innovative project risk management at all key stages (identification, assessment, risk mitigation) through the overcoming main disadvantages of current researches and interconnection of risk response procedures with the nature of risks and their quantification.

So the main objectives on every stage of risk management to reach the goal of research are:

– to improve risk identification and drawing up the risk map taking into account the heterogeneous nature of project risks and strong impact of hardly quantifiable specific risks;
– adaptation, enhancement and development of methodical tools for innovative projects risk assessment taking into account real axiological probability values;
– matching of treatment measures with risk assessment results.

Fullfilling of all these objectives help to simplify risk management and to avoid additional spending to cover risk losses.

2 Literature Review

Theoretical and methodological approaches to the study of specific features of innovative development were discussed by many economists. Modern issues of risk management were examined by Bernstein (Bernstein 1996), Balabanov (Balabanov 1996), Ivanov (Ivanov et al. 2008) etc., and, particularly, specific features of risk management in respect of innovations in general and innovative projects in particular were formulated by Gracheva and Lyapina, [6], Kulikova, Karzhaev, Popov [10], Vertakova and Simonenko [11], Valdajcev, Samovoleva, Taplin and Schymyck, Shapira, Johnstone-Bryden (Johnstone-Bryden 1996), Elahi et al.

Researchers in risk management of innovative projects usually faces with following difficulties in identification, assessment and mitigation of risks:
Risk identification:

– the innovative projects' risk system, risk assessment and risk treatment methods directly and closely depend on the area of innovations' application;
– on different phases of innovative project's implementation a set of risks and risk-generating factors, which, in turn, requires different methodological tools of risk assessment and risks analysis [7];
– risks, as an economic category, are heterogeneous, having the dynamic nature; as a result, the structure of risks, as well as their impact on the project's key parameters may change in the course of time. These risk features are enhanced by the degree of uncertainty intrinsic to innovative projects [1].

Risk assessment:

– most risks of innovative projects are not subject to quantitative analysis with the use of mathematical models (e.g., a group of technical, technological, legal risks, force-majeure risks etc.);
– risk assessment methods are usually discrete, and mechanical joining of commonly used methods of risk analysis and assessment is often impossible due to methodological separateness thereof;
– risk assessment operates static project data or the statistical sample which is insufficient to effectively apply statistical methods of risk assessment;
– non-applicability of the probability theory based on the classic standard distribution for consideration of subjective (axiological) probabilities, intrinsic to innovation processes [8].

Risk mitigation:

– risk mitigation actions usually don't match with identification and assessment results.

Hence, the task of overcoming the above issues and formation of a systemic methodological approach to risk management of innovative projects is quite relevant [2]. The above difficulties and deficiencies of approaches to risk management of innovative projects pointed out to the need for development and enhancement of approaches to identification, assessment and mitigation of risks.

3 Development of Innovative Project Risk Management

3.1 Risk Identification

The most important task at the identification stage is determining the number of risks and their homogeneity. In the approaches examined, risks of different nature are usually considered to be equivalent and homogeneous.

As a rule, risk is considered as a probability of occurrence of adverse conditions, which are related either to:

– failure to achieve the expected results from innovations;
– deficit of planned resources.

Therefore, risk situations may be divided into four types [6]:

– the result was not achieved at all due to impossibility to implement an innovative idea;
– technical or economic characteristics of innovation turned out to be worse than the characteristics expected by the company's management;
– innovative project was implemented, but its costs exceeded the original budget;
– innovations' objective was achieved, but later than expected (see Fig. 1).

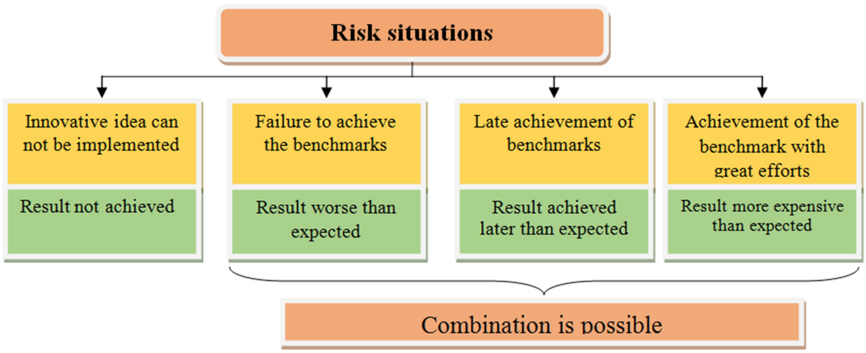


Fig. 1. Risk situations (made by authors).

The abovementioned risks are complex, being initiated by individual or mixed impact of specific risks of a particular project, the impact of which is difficult to express in quantitative terms. So these risks can be considered as core (key) risks caused by specific risks. Therefore, two levels of risks should be distinguished:

– first (upper) level risks include key risks of innovative projects related to occurrence of basic risk situations during project implementation (see Fig. 1);
– second (lower) level risks include innovative project-specific risks usually heterogeneous resulting in the launch of the key project risks.

The risk map is individual for each innovative project, especially concerning second level risks, but author's approach to project risk map can be applied to any innovative project. It allows to take into account the heterogeneous nature of project risks and to point out the project stages which experience the strongest impact of hardly quantifiable specific risks.

3.2 Adaptation, Enhancement and Development of Methodical Tools for Innovative Projects Risk Assessment

A great number of risk assessment methods described in the literature are discrete, and mechanical joining of commonly used methods of risk analysis and assessment is often impossible due to methodological separateness.

The main problem of the stage of an innovative project risk assessment is static project data. Not so often the project manager develops several scenarios for the project budget, but nevertheless it's insufficient to effectively apply statistical methods of risk assessment.

Project risk assessment procedure starts from generation of scenarios of the project task costs on the basis of combination of the scenario-based approach (with account for allowable and marginal growth ratios for certain tasks (or budget items) of the project, which are determined by experts) and the simulation-based stochastic model (geometrical Brownian movement model) [3]:

$$S_t - S_{t-1} = S_{t-1}(\mu \cdot \Delta t + \sigma\varepsilon\sqrt{\Delta t}), \quad (1)$$

where: Δt means duration of the interval between neighboring object observations; e means random noise represented in a form of a normalized normally distributed random variable; μ, y mean the stochastic process parameters (mathematical expectation and volatility) assessed by using the regression analysis method.

Simultaneous application of the scenario-based method and simulation modeling allows to construct the data sample using subjective judgments help to assess the unique nature of an implemented innovative idea and objective data of the stochastic modeling. This allows to get a sufficient data set for applying statistical methods of risk assessment.

On the basis of obtained scenarios, the losses caused by the risk factors are assessed using VAR analysis. It should be noted that such characteristics as the value and time of project implementation are correlating factors. Fluctuations of the project duration increase the cost, primarily due to growth in daily labor costs. That's why, project duration scenarios should be taken into account for assessment of possible absolute losses.

So, for the projects tasks (budget items), which are independent from fluctuation of the project completion time, the basic formula of VAR analysis should be used:

$$VAR = P_{t-1}e^{\mu - k_{1-\alpha}\sigma} \approx P_{t-1}(\mu - k_{1-\alpha}\sigma), \quad (2)$$

where P_{t-1} is the cost of the task in accordance with the project budget; μ, y means mathematical expectation and mean square deviation of the task cost fluctuation, respectively, calculated by the indicators growth rate; k_{1-6} means the quantile of probability distribution, and for the project tasks (budget items) subject to fluctuation of the project completion time, the basic formula of VAR analysis should be used with account for fluctuation in the project duration [4]:

$$VAR = P_{t-1}(\mu\frac{T}{t} - k_{1-\alpha}\sigma\sqrt{\frac{T}{t}}), \quad (3)$$

where T means duration of the task (project) in accordance with pessimistic forecast; t means duration of the task (project) in accordance with the calendar plan.

These models also allow to overcome non-applicability of the probability theory based on the classic standard distribution for consideration of subjective (axiological) probabilities, intrinsic to innovation processes, because the above models mean using the quantile of probability distribution k_{1-6} with real axiological probability values, not for normal distribution.

Further risk management process and determining of risk mitigation measures are based on projected losses for project tasks (budget items).

3.3 Elaboration of Enforcement Actions Matching with Risk Identification and Assessment Results

If you touch on the issue of ways to influence the risk as a result of the obtained estimates, then in the scientific literature you can find quite a variety of methods. All methods widely used in the practice of innovative management can be divided into three groups:

- methods of risk localization;
- methods of risk dissipation;
- methods of risk compensation.

Methods of risk localization. The method of the containment of risk involves identifying the most risky tasks of the project, subject to the influence of the greatest risk factors. Having identified economically the most dangerous tasks of the innovation project, you can make them controllable and thus reduce the level of final risk. At the same time, the definition of the most risky tasks in the management of risks of a different nature and homogeneity requires the application of different selection criteria. Based on the localization of the project's tasks, it is possible to proceed to the specific risks of the innovation project by protecting these sites through the use of dissipation mechanisms and risk financing.

Methods of risk dissipation – an effective measure exposure to the risks of the project in the case of the development of horizontal and vertical structure of the organization [5].

This method can be implemented as:

- involving of specialists from related units experienced an implementation of an innovative project in their departments;
- creating of specialized groups of employees implementing the project in all structural departments;
- exchange of experience among specialists engaged in the development and implementation of the project.

Methods of risk funding imply losses arising from the risk situation.

For choosing the form of risk funding, it is important to take into account all the basic risk characteristics:

- Number of risks;
- Homogeneity of risks;
- Probability of damage;
- Risk losses.

The first two parameters are evaluated at the stage of risk identification, the last two – in the context of direct quantitative assessment (see Fig. 2).

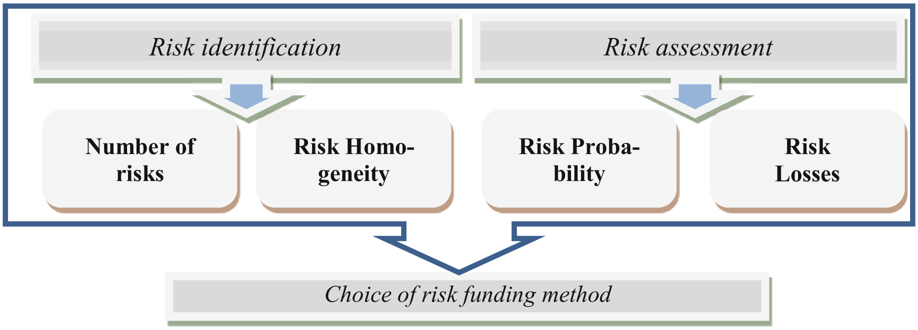


Fig. 2. Basic risk parameters using for choice of risk funding method (made by authors).

The scale and cost of risk treatment measures at the third stage of risk management – depend on the quality of creating risk map and their assessment.
The most commonly used methods of risk funding or covering the loss are:

– covering the loss from current income;
– covering losses from reserves;
– coverage of loss on the basis of self-insurance;
– coverage of loss on the basis of insurance;
– covering the loss by transferring this financing on the basis of a contract.

The first two methods represent a «risk reduction» procedure, and the last three methods are a «risk transfer» procedure.

Coverage of loss from current income assumes covering the damage by current cash flows. It's used in cases where the amount of damage projecting from VAR-method is small and the organization independently manages them. The increase in the number of risks limits the possibility of using this method, since the aggregate damage can become large enough to distort the cash flows of the enterprise. Factors such as homogeneity or heterogeneity of risks do not significantly affect the decision-making on the use of the method of financing risk from current income. However, it is necessary to take into account the financial situation of the enterprise at the time the risk is realized.

Formation of reserve funds – organization creates special reserve funds specifically formed to cover losses from the implementation of project risks. The main issue is to determine and to justify the size of reserves. Too small size does not allow to adequately protect against risk, and too high will cause distraction unreasonably large sums of turnover of the organization.

The characteristics of the risks for which this method can be applied are very close to those that were considered for the method of covering the risk from current income. The main difference is a larger amount of losses, which, in fact, requires the creation of these reserve funds. Therefore, the conditions for its application are primarily related to wider threshold values of losses getting from risk assessment [11].

Coverage of projected losses based on self-insurance means the creation of specialized insurance funds designed to cover losses.

One of the common used ways of self-insurance is the creation of captive insurance organizations (captive insurance), especially common in the financial environment. A captive company is a separate legal entity that is part of a non-insurance organization that insures the risks of the entire group.

The difference between this method and the previous methods is as follows:

– self-insurance working with a large number of similar risks when creating separate business entity is justified;
– reserves are created here within the framework of one business unit.

A significant disadvantage of this method is that in case of loss by the captive company itself, it is distributed among all the participants of the group.

Coverage of projected losses on the basis of insurance – responsibility for compensation of possible damage is transferred to a specialized organization – the insurance company. Moreover, compensation for damage can be transferred both partially and completely.

In practice, this method is used in the following cases:

– Regardless of homogeneity or heterogeneity of the risks and the number of risks if the risk probability is low, and the damage is relatively high;
– If the risks are not homogeneous;
– If there is a great set of risks, the probability of which is large, and the amount of the alleged in impairing small. Of course, in view of the small size of the possible damage, the firm can keep them, but the mass nature of such risks can lead to considerable damage. When the risks are mass and homogeneous, firm can use self-insurance;
– When the probability and (or) the amount of possible losses exceed the specified threshold values. At the same time, the number of such risks and the degree of their homogeneity can be any;
– If there are catastrophic risks.

Coverage of damage through the transfer of liability on the basis of the contract. In this case, the financing of the damage is transferred to another entity on the basis of the concluded contract. An example of such a method is hedging – the procedure for transferring price risk, aimed at minimizing it [10].

Simultaneous use of all forms of financing risks is inexpedient. The specific way to cover the loss depends on the scope of application of risks, their homogeneity, the data obtained during the quantitative assessment, and the adequacy of project financing at the time of the project [9].

In case of risk implementation, the organization will cover losses out of the reserved funds or insurance coverage. If preventive measures in respect of risk factors are not taken, the organization will cover the occurred losses out of its current income, or it may use the mechanism for raising additional funding, if losses are significant. The mechanism of selection of the financing method for losses is implemented in algorithm of project risk response (see Fig. 3):

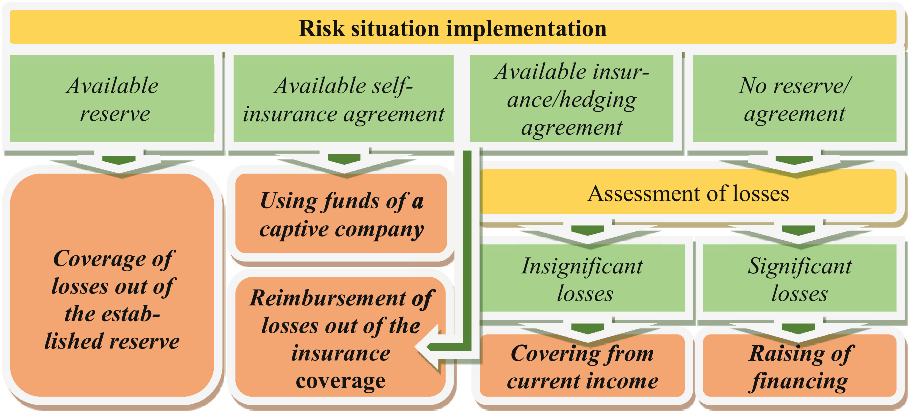


Fig. 3. Order of project risk response (made by authors).

These scheme allows to avoid raising of financing to cover occurring losses.

4 Conclusion

Developed system of innovative project risk management implies consecutive systematic implementation of management phases and matching of assessment results with risk treatment measures, taking into account the innovative nature of considering projects.

Risk identification is the first phase of the elaborated risk management system, which means mapping of key and specific project risks taking into account heterogeneous nature of risks.

Assessment procedures starts with creating several project scenarios on the basis of combination of objective statistical methods and subjective expert judgments which helps further for getting projected losses results using VAR-method real axiological probability values.

Risk treatment measures are based on the data received in the course of assessment and depends on different combination of risk parameters.

The developed risk management system has visible advantages at all key stages of risk management:

Risk identification:

– it takes into account the heterogeneous nature of project risks;
– it operates a set of risks subject to quantitative measurement and assessment.

Risk assessment:

– it includes the minimum required set of correlating assessment methods with the purpose of reduction of efforts required for the estimates and receipt of reliable results;

– it overcomes the insufficiency of project data to effectively apply statistical methods of risk assessment;
– it is based on the study of subjective axiological probabilities rather than beta distribution, which allows to estimate the statistical non-homogeneity of risk probabilities and innovativeness of the studied projects.

Risk mitigation:

– it means matching of risk treatment measures with identification and assessment results.

Generally it operates the parameters, which are uniform in project planning; so it may be adapted to different areas of innovations.

The results of this study may be used for development of application software (in the form of an independent software product or in the form of an automated system module), which would implement a set of risk assessment procedures and determining the key measures to mitigate risks.

References

1. Allen, F., Gale, D.: Financial innovation and risk sharing. MIT, Cambridge (1994)
2. Bowers, J., Khorakian, A.: Integrating risk management in the innovative project. Eur. J. Innov. Manage. **17**(1), 25–40 (2014)
3. Demkin, I.V.: Ocenka riska innovacionno-investicionnyh proektov na osnove imitacionnogo stohasticheskogo modelirovanija. Risk assessment of innovative and investment projects on the basis of imitating stochastic modeling. VJeO. VEO, 249–273 (2005)
4. Demkin, I.V.: Ocenka integrirovannogo innovacionnogo riska na osnove metodologii Value At Risk. Integrated innovative risk assessment on the basis of methodology of Value At Risk. Problemy analiza riska (Risk Analysis Problems) 3(4), 362–378 (2005)
5. Fathutdinov, R.A.: Innovacionnyj menedzhment. Innovative management. Saint Petersburg: Piter (2003)
6. Gracheva, M.V., Ljapina, S.V.: Upravlenie riskami v innovacionnoj dejatel'nosti Risk management in innovative activity. Moscow: JuNITI-DANA (2010)
7. Marle, F., Gidel, T.: Assisting project risk management method selection. Int. J. Proj. Organ. Manage. **6**(3), 254–282 (2015)
8. Nedosekin, A.O., Abdulaeva, Z.I.: Upravlenie korporativnymi riskami i shansami. Corporate risks and chances management. Saint Petersburg: "Saint Petersburg" (2010)
9. Olsson, N.O.E., Spjelkavik, I.: Assumption surfacing and monitoring as a tool in project risk management. Int. J. Proj. Organ. Manage. **6**(1–2), 179–196 (2015)
10. Popov, V.L.: Upravlenie innovacionnymi proektami. Innovative project management. Moscow: INFRA-M (2009)
11. Vertakova, J.V., Simonenko, E.S.: Upravlenie innovacijami: teorija i praktika. Innovation management: theory and practice. Moscow: Exmo (2008)

Estimation and Evaluation of Risk in the Railway Infrastructure

Piotr Smoczyński(✉) and Adam Kadziński

Poznan University of Technology,
Pl. Marii Skłodowskiej-Curie 5, 60-965 Poznań, Poland
piotr.smoczynski@put.poznan.pl

Abstract. Management of railway infrastructure in the European Union (EU) must be performed in accordance with the regulations of the Railway Safety Directive, i.e. it must be covered by a safety management system. The legislation specifies criteria which must be met by such a system, the first being the implementation of risk-related tools. Unfortunately, most of the internationally recognized research is either too general or too specific for direct application in the areas where safety management is still under development and there are not enough resources for complex reliability analyses. Therefore, in the paper, we show a proposal of a relatively simple risk model designed for estimation and evaluation of risk, which could be applied by railway infrastructure maintenance staff to identify the most urgent needs for maintenance works. For this reason, we propose to virtually split railway lines into segments of several types, according to their accessibility by emergency services. As a result, the criteria used in the model for risk estimation can better reflect the reality.

Keywords: Risk analysis · Safety management systems
Railway infrastructure

1 Introduction

Safety management systems have become more and more popular due to changing the emphasis from detailed technical safety to issues of decision making at higher level of management, following the conclusions of investigation reports on major disasters such as Piper Alpha in 1988 [1]. In the railway context, the changes were accelerated by structural reforms performed from the 1990s, aiming at vertical separation, i.e. dividing the infrastructure management from other rail activities. As a result, the market was opened to new railway undertakings and all the railway companies were made to adopt safety management systems based on harmonized principles. It was intended to at least maintain the same level of safety in the railway system [2].

The harmonized principles for safety management systems were gathered in European regulations, issued under the Railway Safety Directive [3]. The legislation specifies criteria which must be met by such a system, the first being the implementation of risk-related tools. Much research was done, especially in Central and Eastern Europe, to better understand the requirements and to establish guidelines for companies [4–6]. All of the research emphasizes the role of the assessment of risk, which is even

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_17

regarded as the basis for proper safety management system design, necessary for defining its objectives [7].

The concept of risk in railway infrastructure maintenance is studied from both technical [8] and organizational point of view [9, 10]. Very often the matter of risk is strictly connected with the reliability of the infrastructure elements [11, 12]. However, most of the internationally recognized research is either too general or too specific for direct application in the areas where safety management is still under development and there are not enough resources for complex reliability analyses. One of counterexamples is the paper of Bureika et al. [13] for Lithuanian railways.

Through the present paper, we aim to present a method for estimation and evaluation of risk on railway infrastructure, which could be used by middle-level staff of an infrastructure manager to rank and justify maintenance activities. In Sect. 2, we shortly present the understanding of risk which has been adopted in this paper. In Sect. 3, we present the results of our study and conclude them in Sect. 4.

2 Generalized Risk Model

There are numerous methods applied to assess risk, e.g. Failure Mode and Effect Analysis [14] and models used in occupational safety [15]. All of them can be formally described in a unified way with the use of a generalized risk model [16]. Let:

$$Z = \{z_1, z_2, \ldots, z_n\}, \tag{1}$$

be a finite set of hazards which have been identified in the analysis domain. For the decision-making in the risk management we need to introduce a risk metric in the form of value of an R function, given as follows:

$$R : Z \to V \subset \mathbf{R}, \tag{2}$$

which assigns to each of the hazards from the set Z (Eq. (1)) a value from the subset V of the set of all the real numbers $\mathbf{R}$. Each of the hazards will be assessed according to m criteria $K_i (i = 1, 2, \ldots, m)$, whereas each criterion must refer to either the possibility or the consequences of hazard activation.

The result of the risk assessment of hazard $z_k (k = 1, 2\ldots, n)$ according to criterion $K_i (i = 1, 2, \ldots, m)$ is a risk component $r_i (i = 1, 2, \ldots, m)$ which reflects one or more of the hazard attributes from the set $X_i (i = 1,\ 2, \ldots, m)$:

$$\begin{gathered} r_i : X_i \to \Omega_i;\ \ r_i(z_k) = \omega_{i,j};\ \ \omega_{i,j} \in \Omega_i; \\ i = 1, 2, \ldots, m;\ \ j = 1, 2, \ldots, s_i;\ \ k = 1, 2, \ldots, n, \end{gathered} \tag{3}$$

where $X_i (i = 1, 2, \ldots, m)$ are sets of hazard attributes (e.g. the history of hazard activation in the analysis domain; the possibility of hazard activation; the number of fatalities caused by hazard activation) and $\Omega_i (i = 1, 2, \ldots, m)$ are sets of s_i values (denoted as $\omega_{i,j}$), which can be used in the assessment according to the criterion i.

Apart from the values of risk components, in the risk estimation we can include the weight of each of the criteria $K_i (i = 1, 2, \ldots, m)$. Let the set A be defined as:

$$A = \{a_1, a_2, \ldots, a_m\}, \tag{4}$$

where a_1 denotes the weight of the criteria K_1 etc. Thus, the risk metrics can be calculated as:

$$R(z_k) = f(a_i, r_i(z_k)); \quad i = 1, 2, \ldots, m; \quad k = 1, 2, \ldots, n. \tag{5}$$

The function f can be given in form of an equation, but it is not obligatory; e.g. risk matrix can be used instead. The obtained value of the risk metrics is then evaluated, i.e. the risk is assigned into one of several categories, such as categories of acceptable, tolerable and unacceptable risk.

3 Results

In the paper, we show a proposal of a risk model designed for estimation and evaluation of risk, which could be performed by railway infrastructure maintenance staff to identify the most urgent needs for maintenance works. For this reason, we propose to split railway lines into segments, according to their accessibility by emergency services:

- segments, where tracks are located at the same level as their surroundings,
- segments, where the tracks are in cuttings or on embankments,
- segments on bridges, viaducts etc.

Our proposal consists of five risk estimation criteria related to the possibility of hazard activation, as well as its anticipated consequences. The risk can be expressed as:

$$R(z_k) = a_1 \cdot r_1 \cdot \sum\nolimits_{i=2}^{5} a_i \cdot r_i, \tag{6}$$

where z_k denotes the k-th hazard z whose value of risk metrics R is being estimated; a_i denotes the weight and r_i denotes the value of the risk component according to the criterion $i = 1, 2, \ldots, 5$. One of the criteria depends on the type of segment. The overview of the risk estimation is shown in Fig. 1.

The criteria can be roughly divided into three groups, where risk is estimated in respect to: the operating conditions, the given segment of infrastructure and the hazard. Each of the criteria, as well as the way of evaluating the risk, will be presented in detail in the following subsections.

3.1 Risk Estimation in Respect to the Operating Conditions

The first criterion proposed in our model is the criterion C1 'risk magnifier', which should reflect the operating conditions on the analysed railway line. The idea of 'magnifying risk' comes from [17] and the reason why it is proposed is our belief that

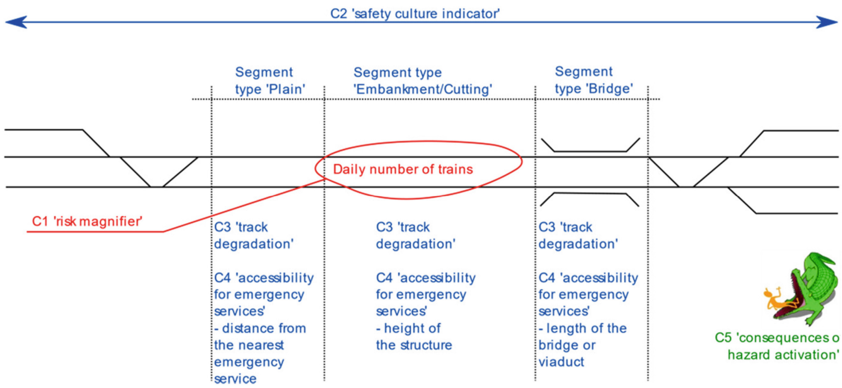


Fig. 1. Overview of the risk estimation criteria C1…C5 proposed in the model.

the risk of hazards increases with the number of trains operating on the line. This is because a moving train is an inevitable risk factor in formulation of virtually all hazards in the domain. The more trains use the line, the more probable is the coincidence of risk factors which can lead to negative consequences. Additionally, the more trains, the more losses e.g. due to delays.

On the other hand, the influence of the operating conditions cannot be assessed too high, as it could lead to results which are clearly not true, such as intolerable risk for all hazards near main railway stations. Therefore, we propose to use the following formula:

$$r_1 = 1 + 0.1 \cdot \frac{TPD}{TPD_{\max}}, \tag{7}$$

where TPD denotes the number of trains per day on the segment under analysis and $TPD_{\max}$ denotes the maximum number of trains per day in one segment in the network. Consequently, the criterion C1 ‘risk magnifier’ cannot change the overall risk metrics value by more than 10%.

3.2 Risk Estimation in Respect to the Segment

The second criterion, C2 ‘safety culture indicator’, is meant to describe the overall safety performance of the railway system in the respective Member State. We propose to use the tool which already exists in the EU legislation, i.e. the monitoring of Common Safety Targets (CST). The CST are divided into six groups:

- risk for passengers,
- risk for employees,
- risk for level crossing users,
- risk for others,
- risk for unauthorized persons on railway premises,
- risk for the whole society.

The most suitable for our needs is the last category, represented by CST No. 6. The observed safety performance (OSP) in respect to this CST is defined as follows [18]:

$$\mathrm{OSP} = \frac{\text{total number of FWSIs per year arising from significant accidents}}{\text{number of train - km per year}} \tag{8}$$

where FWSI means fatalities and weighted serious injuries. The observed safety performance in a Member State, in respect to each of the CST, is then compared with the national reference values. Detailed information on applicable definitions and methodology can be found in the Decision cited above.

The national reference value represents the highest tolerable value of the respective risk metric defined by the CSTs. It can decrease over time, as it should reflect the current level of safety in the Member State. Obviously, the values differ also between the Member States. Therefore, we propose to refer to the national reference value in the qualification scheme for the criterion C2, as shown in the Table 1.

Table 1. Qualification scheme for criterion C2 'safety culture indicator' – observed safety performance (OSP) in respect to the national reference value (NRV) of the maximal tolerable risk to the whole society.

j	Level $\omega_{2,j}$	Description
1	Green	$\mathrm{OSP} \leq 0.5 \cdot \mathrm{NRV}$
2	Yellow	$0.5 \cdot \mathrm{NRV} < \mathrm{OSP} \leq \mathrm{NRV}$
3	Red	$\mathrm{OSP} > \mathrm{NRV}$

For example, the current Polish national reference value for the CST No. 6 amounts to $1590 \cdot 10^{-9}$, whereas the last reported observed safety performance in 2015 amounted to $1050.421 \cdot 10^{-9}$ [19]. It means that, according to the scheme for criterium C2, the current level for Poland would be yellow.

The next two criteria are defined separately for each segment. First of them is devoted to the technical state of the infrastructure as was called by us 'track degradation' (C3). There are several parameters which can describe this state, such as track geometry values or wavelength of rail surface defects [20]. We decided to base the qualification scheme on the track degradation indicator used on Polish railways [21]:

$$G = \frac{G_r + G_s + G_b}{3} \tag{9}$$

where G_r, G_s and G_b denote degradation indicators of rails, sleepers and ballast accordingly, and are defined in the infrastructure manager instruction cited above. The instruction also prescribes the way of interpreting the value of G, i.e. regular maintenance works for $G < 0.6$, planning of track renovation for $G > 0.8$ and detailed analysis of further steps for the values in-between. This was used in the qualification scheme shown in Table 2.

Table 2. Qualification scheme for criterion C3 'track degradation'.

j	Level $\omega_{3,j}$	Description
1	Green	$G < 0.6$
2	Yellow	$0.6 \leq G \leq 0.8$
3	Red	$G > 0.8$

The criterion C4 'accessibility for emergency services' depends on the type of segment and should represent the most crucial characteristics in respect to the time which will be necessary to bring help in case of a hazard activation. We assumed that for the segments where track is built on the same level as its surroundings (type Plain), the key role is played by the distance from the nearest emergency services. As this data can be difficult to acquire, we propose to assess this criterion by choosing the type of area where the segment is located, as shown in Table 3.

Table 3. Qualification scheme for criterion C4 'accessibility for emergency services'.

Segment type	j	Level $\omega_{4,j}$	Parameter	Description
Plain	1	Green	Distance from the nearest emergency services	Bigger town, city
	2	Yellow		Small town or suburbs
	3	Red		Forest, field, etc.
Embankment cutting	1	Green	Height of the structure	$h < 6\,\text{m}$
	2	Yellow		$6 \leq h < 12\,\text{m}$
	3	Red		$h \geq 12\,\text{m}$
Bridge	1	Green	Length of the bridge or viaduct	$l < 400\,\text{m}$
	2	Yellow		$400 \leq l < 750\,\text{m}$
	3	Red		$l \geq 750\,\text{m}$

The distance from the nearest emergency services is obviously important for the other types of segments as well. However, in our opinion, their other characteristics prevail. Therefore, for Embankment/Cutting type of segment we decided to use the height of the structure as the decisive factor for the accessibility. The limit values are based on the Polish regulation [22]. For the Bridge type of segment, the decisive parameter will be the length of the bridge or viaduct. The limit values represent the maximum permitted length of train (750 m) and the maximum length of passenger trains (400 m) and should be adjusted to the situation in the analysis domain.

There are other types of segments which could possibly be distinguished. One of the examples is the type Level crossing; some ideas for their description can be found in [23]. It can also be justified to introduce special segments e.g. for places where wild animals often cross railway lines or in the area of possible mining losses.

3.3 Risk Estimation in Respect to a Hazard

The last proposed criterion C5 ‘consequences of hazard activation’ is the only one which depends on the hazard itself and should be used to assess the risk without considering any particular spot on the railway network. Of course, in reality, the consequences may be considerably influenced by factors related to the specific place. Therefore, to make the assessment more transparent and straightforward, we propose to use only three easily distinguishable values, as shown in Table 4.

Table 4. Qualification scheme for criterion C5 ‘consequences of hazard activation’.

j	Level $\omega_{5,j}$	Description
1	Green	Incident, i.e. any occurrence, other than accident or serious accident, associated with the operation of trains and affecting the safety of operations
2	Yellow	Accident, i.e. any unwanted or unintended sudden event or a specific chain of such events which have harmful consequences
3	Red	Serious accident, i.e. any train collision or derailment of trains, resulting in the death of at least one person or serious injuries to five or more persons or extensive damage to rolling stock, the infrastructure or the environment, and any other similar accident with an obvious impact on railway safety regulation or the management of safety

In the proposed qualification scheme (Table 4) we used the definitions provided by the Railway Safety Directive [3], as they are already well-established on the European railways.

3.4 Overall Risk Estimation

In the Sects. 3.1, 3.2 and 3.3, we have presented a set of five criteria used in the overall risk estimation of a particular hazard. The criteria can be classified into two groups, which are derived from the basic concept of risk:

1. Criteria describing the possibility of hazard activation: C1 ‘risk magnifier’, C2 ‘safety culture indicator’ and C3 ‘track degradation’
2. Criteria describing the consequences of hazard activation: C1 ‘risk magnifier’, C4 ‘accessibility for emergency services’ and C5 ‘consequences of hazard activation’.

The criterion C1 ‘risk magnifier’ can be classified to both groups due to its dual nature. From one side, it describes the possibility of hazard activation – the more often a train comes, the more probable is an event with the train. On the other hand, with greater number of trains the consequences increase, e.g. due to delays and diversions.

The difference between the criterion C1 and the following criteria C2…C5 is reflected also in the set of permissible values. For the criterion C1 ‘risk magnifier’, defined by Eq. (7), it can be stated that:

$$\Omega_1 = \left\{ 1 + 0.1 \cdot \frac{1}{TPD_{\max}}, 1 + 0.1 \cdot \frac{2}{TPD_{\max}}, \ldots, 1 + 0.1 \cdot \frac{TPD_{\max}}{TPD_{\max}} \right\}, \quad (10)$$

where $TPD_{\max}$ is the maximum number of trains per day in one segment of the network. For the other criteria, the sets Ω_i are equal:

$$\Omega_i = \{1, 3, 5\}, \ \ i = 2, .., 5. \quad (11)$$

It is generally approved that the criteria related to the consequences should contribute up to 2 times more to the overall value of risk. Therefore, the following set of weights A is proposed:

$$A = \{1, 1, 2, 3, 3\}. \quad (12)$$

With the assumptions given in Tables 1, 2, 3 and 4 and Eqs. (7)–(12), the overall risk can be estimated with the formula given in Eq. (6).

3.5 Risk Evaluation

Let us assume that the risk component according to the criterion C1 equals 1. Then the solution space of Eq. (6) ranges from 9 for all the criteria assessed as ‘green’ up to 45 in case of all criteria assessed as ‘red’. We propose to divide this solution space into three equal parts and use them as limit values for risk categories, as shown in Table 5.

Table 5. Risk evaluation scheme.

Value of risk metrics	Risk category
$R \leq 21$	Acceptable risk
$21 < R \leq 33$	Tolerable risk
$R > 33$	Unacceptable risk

The proposed risk evaluation scheme (Table 5) deliberately takes no consideration of the criterion C1 to achieve the effect of ‘magnifying’ the overall risk. Depending on the resulting risk category, further actions must be planned. Especially if the risk is evaluated as unacceptable, risk reduction measures (i.e. maintenance activities) must be implemented.

4 Conclusions

The risk model presented in the paper allows to perform the assessment and evaluation of risk related to the railway infrastructure. It is meant to be used by the middle-level staff which is responsible for prioritization of maintenance activities with respect to the available financial means. The criteria proposed by us may be adjusted to better suit the local conditions.

The main purpose of the risk model application is to rank the needs in the infrastructure maintenance and therefore to use the money in a more effective way, i.e. to minimize the risk to the greatest extent possible. However, also the purpose of justification of decisions should not be omitted. As Hokstad and Steiro noticed in their paper [24], in an open and democratic society it is valuable to have a foundation for decisions which can be tested and re-examined if necessary.

Acknowledgements. All presented work is realized within the research project “Contemporary problems of operation, maintenance and safety of rail transportation systems”, No. 05/52/DSMK/0266, co-financed by the Dean of the Faculty of Machines and Transport, Poznan University of Technology.

References

1. Hale, A.R., Heming, B.H.J., Carthey, J., Kirwan, B.: Modelling of safety management systems. Saf. Sci. **26**, 121–140 (1997)
2. Koursi, E.M.E., Mitra, S., Bearfield, G.: Harmonising safety management systems in the European railway sector. Saf. Sci. Monit. **11**, 1–14 (2007)
3. Directive (EU) 2016/798 of 11 May 2016 on railway safety (recast). European Parliament and Council (2016)
4. Čičmancová, S.: Safety and risk as part of railway system. Trans. VŠB - Tech. Univ. Ostrava, Saf. Eng. Ser. **8** (2013)
5. Leitner, B.: A general model for railway systems risk assessment with the use of railway accident scenarios analysis. Procedia Eng. **187**, 150–159 (2017)
6. Chruzik, K., Sitarz, M.: Investigation and development of safety measures in the European Union railway transport. Mechanika **20**, 431–437 (2014)
7. Demichela, M., Piccinini, N., Romano, A.: Risk analysis as a basis for safety management system. J. Loss Prev. Process Ind. **17**, 179–185 (2004)
8. Podofillini, L., Zio, E., Vatn, J.: Risk-informed optimisation of railway tracks inspection and maintenance procedures. Reliab. Eng. Syst. Saf. **91**, 20–35 (2006)
9. Vatn, J., Aven, T.: An approach to maintenance optimization where safety issues are important. Reliab. Eng. Syst. Saf. **95**, 58–63 (2010)
10. Smoczyński, P., Kadziński, A.: Introduction to the risk management in the maintenance of railway tracks. J. Mech. Transp. Eng. **68**, 65–80 (2016)
11. Andrews, J., Prescott, D., De Rozières, F.: A stochastic model for railway track asset management. Reliab. Eng. Syst. Saf. **130**, 76–84 (2014)
12. Macchi, M., Garetti, M., Centrone, D., Fumagalli, L., Piero Pavirani, G.: Maintenance management of railway infrastructures based on reliability analysis. Reliab. Eng. Syst. Saf. **104**, 71–83 (2012)

13. Bureika, G., Bekintis, G., Liudvinavicius, L., Vaiciunas, G.: Applying analytic hierarchy process to assess traffic safety risk of railway infrastructure. Eksploat. i Niezawodn. - Maint. Reliab. **15**, 376–383 (2013)
14. International Electrotechnical Commission: IEC 60812:2006 Analysis techniques for system reliability - Procedure for failure mode and effects analysis (FMEA) (2006)
15. Polski Komitet Normalizacyjny: PN-N-18002:2011 Systemy zarządzania bezpieczeństwem i higieną pracy. Ogólne wytyczne do oceny ryzyka zawodowego (2011). [in Polish]
16. Kadziński, A.: Studium wybranych aspektów niezawodności systemów oraz obiektów pojazdów szynowych [in Polish]. Wydawnictwo Politechniki Poznańskiej, Poznań (2013)
17. Marais, K.B., Robichaud, M.R.: Analysis of trends in aviation maintenance risk: an empirical approach. Reliab. Eng. Syst. Saf. **106**, 104–118 (2012)
18. Decision of 5 June 2009 on the adoption of a common safety method for assessment of achievement of safety targets, as referred to in Article 6 of Directive 2004/49/EC of the European Parliament and of the Council. Commission of the European Communities (2009)
19. Urząd Transportu Kolejowego: Raport w sprawie bezpieczeństwa transportu kolejowego w Polsce w 2015 roku., Warszawa (2016)
20. Lichtberger, B.: Track Compendium: Formation, Permanent Way, Maintenance, Economics. Eurailpress (2005)
21. PKP Polskie Linie Kolejowe S.A.: Warunki techniczne utrzymania nawierzchni na liniach kolejowych (2015)
22. Minister Transportu i Gospodarki Morskiej: Rozporządzenie Ministra Transportu i Gospodarki Morskiej z dnia 10 września 1998 r. w sprawie warunków technicznych, jakim powinny odpowiadać budowle kolejowe i ich usytuowanie (1998)
23. Kobaszyńska-Twardowska, A., Kadziński, A.: The model of railway crossings as areas of analyses of hazard risk management. Logist. Transp. 85–92 (2013)
24. Hokstad, P., Steiro, T.: Overall strategy for risk evaluation and priority setting of risk regulations. Reliab. Eng. Syst. Saf. **91**, 100–111 (2006)

An Empirical Analysis of Time Headways on Two-Lane Roads with Mixed Traffic

Rupali Roy and Pritam Saha(✉)

Indian Institute of Engineering Science and Technology, Shibpur,
Howrah 711103, India
rupaliroy.nit@gmail.com, saha.pritam@gmail.com

Abstract. This paper demonstrates a method of describing headways of two-lane roads under mixed traffic situation using statistical distributions. Characteristically such distributions may have two forms: single and mixture of two or more distributions. A single distribution, however, cannot describe headways in the event of significant proportion of shorter headways in traffic. Use of mixed models is appropriate in such situation since they describe headways by decomposing them into free and following component. Based on experiences with mixed traffic and field studies on two-lane highways of India, this paper has shown that Cowan's M3 can be reasonably applied for modeling headways up to a flow level that corresponds to 'moderate to heavy flow'. However, since shifted negative exponential distribution part of the Cowan's M3 distribution cannot model short headways, Cowan's M3 distribution cannot model headway data at congested state of flow when almost all the vehicles in the traffic stream start moving in following.

Keywords: Mixed traffic · Headway distributions · Mixed model

1 Introduction

Time headway is defined as the time interval, expressed in seconds, between successive vehicles as they pass a point on a lane or roadway. It is an important microscopic traffic flow parameter which plays a fundamental role in many traffic engineering applications, such as capacity and level of service analysis, safety studies and also in simulation issues [1]. Accordingly, there is a need to describe headways by a suitable statistical model. Conventionally, negative exponential distribution is used for modelling headway data. However, there have been a number of researchers who reported the use of several other models in order to explain the headway distribution pattern more explicitly. This is quite relevant when the prevailing traffic is heterogeneous in character and car-following interaction is frequent at increased flow level, as usually observed on two-lane roads.

Over the years, traffic analysts across the globe have reported the necessity of investigating time headways more explicitly and they suggested several methods of headway modelling at different operating conditions. Principally, such models may have two forms: single statistical distribution and mixed models of two or more distributions [2]. The first one could be reasonably appropriate in the event of low flow

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_18

when the traffic interaction is insignificant, whereas the second one characteristically exhibits its aptness if such interaction is considerable. This is attributable to the fact that although a single distribution is simple and easy to apply, there are several instances when performance of such distributions was not observed satisfactory due to their limited capabilities in approximating shorter headways.

Notably, most of the studies reported so far in regard to this are based on homogeneous traffic, thereby, making it intrinsic to investigate the effect of heterogeneity in the distribution models. Further, such effect aggravates in the event of heavy flow when interaction between vehicles is considerably high. This has been the motives of the present study wherein field study was conducted on two-lane highways in India. Observations indicate a large proportion of shorter headways in the data set and thus, manifest the need of modelling shorter and longer headways separately, thereby, developing a mixed model. The present study, consequently, made an attempt to interpret mixed distributions for modelling car-following headways on two-lane roads.

2 Research Motivations

There have been a number of studies that investigated time headways and suggested appropriate model forms. Experiences on urban expressways of Bangkok and Isfahan have shown that GEV distribution [3] and shifted lognormal and gamma distribution [4] are quite effective for the purpose of modeling time headways. Couple of studies on Indian urban roads reported that Hyperlang distribution can be used to describe headways under mixed traffic [5] but in the event of substantial proportion of smaller vehicles in such traffic, negative exponential distribution is a good choice [6]. On the basis of a study on two-lane roads of North Carolina, it was found that Schuhl model is suitable in describing headways [7]. However, a few studies have shown that different distributions may work better at different flow conditions. A study on urban roads of Riyadh reported the use of negative exponential, shifted exponential and Erlang distributions respectively for low, medium and high flow of traffic [8]. By the same token, another study on Finnish two-lane roads indicates that gamma distribution fits well to headway data under low-to-moderate traffic volumes, while lognormal distribution can be used to describe follower headways [9].

Application of single distribution is, however, somewhat impractical when the proportion of shorter headways is significant. Mixed distribution model is found appropriate in such situation since it considers following and free vehicles separately at the time of modeling. Over the past few decades, a number of mixed models have been developed and tested. They are respectively Cowan's M1–M4 model [10], double displaced negative exponential distribution model [2] and the generalized queuing model [11]. Among them, Cowan's M3 model has been widely applied for modeling headways because it is relatively simpler than other types of mixed models and gives more realistic results particularly while describing longer headways [12, 13].

Most of the studies that have applied mixed models are, however, based on more or less homogeneous traffic. In context of mixed traffic, quite often slower vehicles force faster ones to move in following, thereby; increase the amount of shorter headways. Since, single distributions cannot approximate them appropriately, there is a need to examine mixed distribution model for such traffic. This fact along with the pressing need of developing capacity standards under such traffic have been the motives for conducting the current study.

3 Study Design

3.1 Conceptual Blueprint

On two-lane roads, formation of platoons is quite frequent especially under mixed traffic conditions. At moderate and heavy flow of traffic, large numbers of faster vehicles are entrapped inside platoons and they start moving in following with shorter headways. Observed headway data, thus, exhibits an 'increased initial rate' and thereby, makes a 'highly skewed distribution'. Since, single distribution functions cannot describe the sharp peak and long tail of the headway distribution [9], the current study explicates the need of considering composite distribution that can define both the vehicles, 'following' and 'free', while describing headways (h) (Eqs. 1 and 2). The term 'free', however, needs an elucidation and in this context a fairly recent study [14] indicates that under mixed traffic situation car following interaction generally ceases beyond 6 s headway. The present study, thus, considered vehicles that move with a headway of more than 6 s as 'free'.

$$\mathrm{h} = \mathrm{u} + \mathrm{v}, \tag{1}$$

$$\mathrm{f(h)} = \theta.\mathrm{f}_1(\mathrm{h}) + (1-\theta).\mathrm{f}_2(\mathrm{h}), \tag{2}$$

where, u = headway between free vehicles, v = headway between following vehicles, θ = proportion of following vehicles, f(h) = probability distribution of composite model and, $\mathrm{f}_1(\mathrm{h})$, $\mathrm{f}_2(\mathrm{h})$ = probability distribution of following and free headways respectively.

Since, several studies have reported the use of Cowan's M1–M4 [10] and Double Displaced Negative Exponential Distribution (DDNED) [15] as composite distributions for modeling headways, the present study made an attempt to evaluate the compatibility of these distributions (Eqs. 3, 4, 5, 6 and 7) in context to mixed traffic. Cowan M1 and M2 models, represented by Eqs. 3 and 4, are basically negative exponential and displaced negative exponential distributions. They are commonly used in headway modeling under low flow and when traffic is more or less homogeneous in character. However, for the purpose of modeling headway distribution as a mixture of 'following' and 'free' vehicle headways, Cowan proposed M3 and M4 model as an alternative to simple exponential model. In M3 model (Eq. 5), shorter headways are represented by 'minimum headway' and 'free' vehicles are assumed to follow a shifted exponential distribution. As a matter of fact, M3 does not model shorter headways with reasonable amount of accuracy since they are represented by a single headway. Consequently, M4 was introduced as a more general model (Eq. 6) wherein shorter headways are represented by G(h) in lieu of a single headway. An appropriate distribution function for G (h) can be chosen based on the characteristics of field data. By the same token, DDNED

(Eq. 7) was found to exhibit good fit to observed headways than simple displaced negative exponential distribution (Eq. 4).

$$F(h) = \begin{cases} 0 & h < 0 \\ 1 - \exp(-\lambda.h) & h \geqslant 0 \end{cases}, \tag{3}$$

$$F(h) = \begin{cases} 0 & h < \Delta \\ 1 - \exp\{-\lambda.(h - \Delta)\} & h \geqslant 0 \end{cases}, \tag{4}$$

$$F(h) = \begin{cases} 0 & h < \Delta \\ 1 - (1 - \theta).\exp\{-\lambda.(h - \Delta)\} & h \geqslant 0 \end{cases}, \tag{5}$$

$$F(h) = \begin{cases} 0 & h < 0 \\ \theta.G(h) + (1 - \theta)\int_0^h G(h - u).\lambda.\exp(-\lambda.u)du & h \geqslant 0 \end{cases}, \tag{6}$$

$$F(h) = \begin{cases} 0 & h < d \\ \int_0^t \theta.\gamma_1.\exp\{-\gamma_1(h - d)\}dh \ + \ \int_t^\infty (1 - \theta).\gamma_2.\exp\{-\gamma_2(h - d)\}dh & h \geqslant 0 \end{cases}, \tag{7}$$

where, λ = scale parameter, Δ = minimum headway, θ = a parameter that represents proportion of following vehicles, $G(h)$ = distribution of following headway component, γ_1 & γ_2 = constants associated with the flow status and, d = displaced parameter.

Although, Cowan M4 and DDNED seem to give more realistic results, studies have reported that it is difficult to calibrate the parameters particularly using field observed headways [2]. Such difficulty, however, further increases in the event of heterogeneity in traffic mix. A wide range of vehicle types in terms of static and dynamic characteristics and also, drivers' behavior collectively have an impact on vehicle-arrival characteristics [16] and thus, make analysis of inter-arrival time or time headways somewhat complex. Cowan's M3 distribution, on the other hand, is simple in the sense that parameter estimation (Eq. 5) is not very complicated. The present study, therefore, considered M3 as an alternative to single distribution for modeling headways under mixed traffic.

3.2 Collection of Headway Data

Field studies were, accordingly, carried out on two-lane roads that exhibit heterogeneity in its traffic composition. Two road sections were selected for the study: they are respectively, a national (NH-44) and a state (SH-13) highway passing through the states of Tripura and West Bengal of India. The pavement condition of both the study sections was good and they were free from the effect of intersection and curvature. Pavement width of 7 m with 1 m earthen shoulder on its either side was observed for both the sections and their capacity was found to be 2200 & 2300 pc/h respectively [17].

Field data was collected using video photographic survey technique. At 'study site-1 (NH-44)' traffic data was captured covering a wide range of flow levels. Two video cameras were installed on either side of a reference line marked on the pavement and the time when a vehicle just crosses the reference line was recorded. Traffic data was collected next to a bottleneck and also, by stopping traffic movements in order to observe car following situation. Several such trials (about 2 min duration) were made to ensure the adequacy of sample size. Traffic police help was taken for conducting the study [14]. For the purpose of field validation of the proposed model, a pilot study was conducted on another road section (study site-2) which has almost similar traffic characteristics and composition. Traffic data was extracted from the video files and necessary readings i.e. vehicle type and crossing time were noted down. Figure 1 displays the composition of traffic wherein proportion of non motorized vehicles (NMV) is considerably high for both the study sites. Figure 2 exhibits the distribution of observed headways and indicates higher fraction of shorter headways for both the cases.

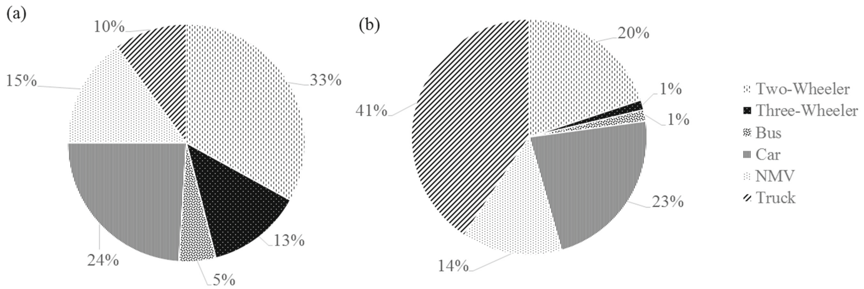


Fig. 1. Observed traffic compositions at: (a) study site-1 (NH-44) and (b) study site-2 (SH-13).

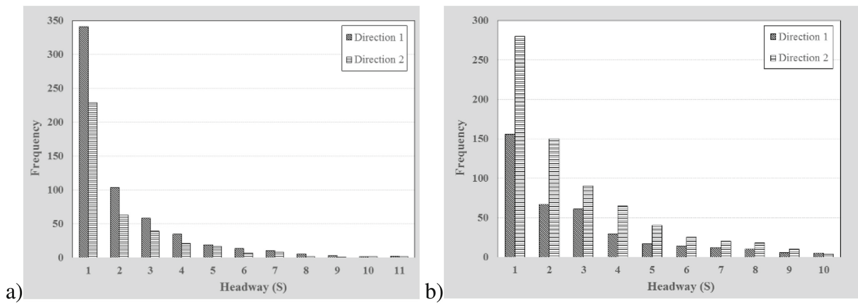


Fig. 2. Distribution of observed headways: (a) study site-1 (NH-44) and (b) study site-2 (SH-13).

Traffic data was characterized as moderate to heavy flow range based on platooning in the traffic stream. Since moderate flow corresponds to stable flow, the average number of headways inside and outside platoons under such flow is assumed to be approximately equal. Considering this fact, the present study found that the flow level corresponding to a volume to capacity ratio of about 0.5–0.6 represents moderate flow. Further, the average speed of non-following vehicles at such flow was observed to be in the range of 30–35 kmph, which, however, reduces to about 20–25 kmph near capacity. Accordingly, based on such speed the flow levels were expressed in terms of v/c ratio as follows:

- 0.5–0.6: Moderate traffic,
- 0.7: Moderate to heavy traffic without significant slowing,
- 0.8–0.9: Heavy traffic with minor slowing,
- 1.0: Congested traffic involving slowing and stopping.

4 Study Results

4.1 Development of Headway Models

Capacity formula described in the highway capacity manual is based on the assumption that headway follows negative exponential distribution [18, 19]. However, this distribution exhibits limitations in approximating shorter headways since it does not take the effects of platoons into consideration. As a result, it gives realistic results only in the event of low flow (<200 veh/h). This problem further aggravates under mixed traffic where formation of platoons are frequent resulting in significant proportion of shorter headways in traffic data. In these circumstances, headway data has two components: following and free and therefore, simple distribution cannot describe them properly. The current study has dealt with this problem and applied Cowan's M3 distribution as a composite model while modeling car-following headways.

Accordingly, an attempt was made to fit the distribution function to field data and obtain a calibrated expression. There are several methods to estimate the parameters: Maximum likelihood estimation (MLE) [20], method of moments [21], minimum chi-square method [22], and simultaneous numerical estimation [18]. The maximum likelihood method is, however, considered to be effective especially when the sample size is not large [23]. Since the current study is based on field data, maximum likelihood method was applied as an effective heuristic method. Thus, the parameters were estimated for the field data collected at 'study site-1' by maximizing the log-likelihood function (Eqs. 8 and 9) and presented in Table 1.

$$L(\hat{\varphi}) = \prod_{i=1}^{n} f(h_i|\hat{\varphi}), \tag{8}$$

$$\ln L(\hat{\varphi}) = \sum_{i=1}^{n} \ln f(h_i|\hat{\varphi}), \tag{9}$$

where, $\hat{\varphi}$ = parameter vector and $f(h_i|\hat{\varphi})$ = probability density function.

Table 1. Estimated parameters and goodness-of-fit results of Cowan's M3 model.

	v/c ratio	λ	$(1-\theta)$ (percent)	Δ (s)	K-S test statistic	P-value	Critical value	Significance level	Hypothesis test
West bound traffic	0.4	0.119	0.556	3.07	0.175	0.716	0.348	0.05	Accept
	0.5	0.136	0.495	2.98	0.200	0.699	0.391	0.05	Accept
	0.6	0.170	0.434	2.97	0.245	0.398	0.375	0.05	Accept
	0.7	0.172	0.356	2.97	0.310	0.192	0.391	0.05	Accept
	0.8	0.192	0.343	2.98	0.196	0.582	0.348	0.05	Accept
	0.9	0.238	0.306	2.97	**0.466**	**0.016**	0.409	0.05	**Reject**
	1.0	0.249	0.264	3.05	**0.456**	**0.019**	0.409	0.05	**Reject**
East bound traffic	0.4	0.097	0.530	3.30	0.139	0.894	0.358	0.05	Accept
	0.5	0.110	0.463	3.12	0.161	0.806	0.348	0.05	Accept
	0.6	0.136	0.413	2.92	0.294	0.143	0.384	0.05	Accept
	0.7	0.191	0.365	2.92	0.332	0.137	0.391	0.05	Accept
	0.8	0.289	0.382	2.98	0.329	0.146	0.362	0.05	Accept
	0.9	0.305	0.352	3.10	**0.366**	0.078	0.312	0.05	**Reject**
	1.0	0.393	0.280	2.98	**0.427**	**0.016**	0.337	0.05	**Reject**

Cumulative distribution functions (Eq. 5) were plotted for all the flow levels (Fig. 3). Subsequently, an attempt was also made to test the goodness-of-fit of the selected distributions to the data points. Kolmogorov–Smirnov (K-S) test was used since it can use data with a continuous distribution and there is no minimum frequency per test interval [1]. The K-S test statistic is calculated by determining the difference between the cumulative percentage of the measured frequency and the cumulative percentage of the expected frequency and the values obtained are displayed in Table 1. The null hypotheses for each test were as follows:

- The compatibility hypothesis of headway distribution with fitted model was rejected (P-value < α) or not rejected (P-value > α);
- The compatibility hypothesis of headway distribution with fitted model was rejected (critical value < test statistic) or not rejected (critical value > test statistic)'.

The '*p-value*' is defined as the probability of obtaining a result equal to or "more extreme" than what was actually observed, when the null hypothesis is true. A critical value is the point on the scale of the test statistic beyond which the null hypothesis is rejected, and is derived from the level of significance (α) of the test. A close look into Table 1 reveals the fact that at congested state of flow that corresponds to a v/c ratio of 0.9 and 1.0 respectively compatibility hypothesis of the fitted distributions are rejected. This could be attributed to the fact that at such flow levels almost all the vehicles start moving in following particularly under mixed traffic. Accordingly, the proportion of longer headways in the traffic stream reduces considerably. This fact along with the theoretical assumptions of Cowan M3, which demonstrates that shifted negative exponential distribution part of it avoids short headways at the time of modeling, make application of such distribution inappropriate under such condition.

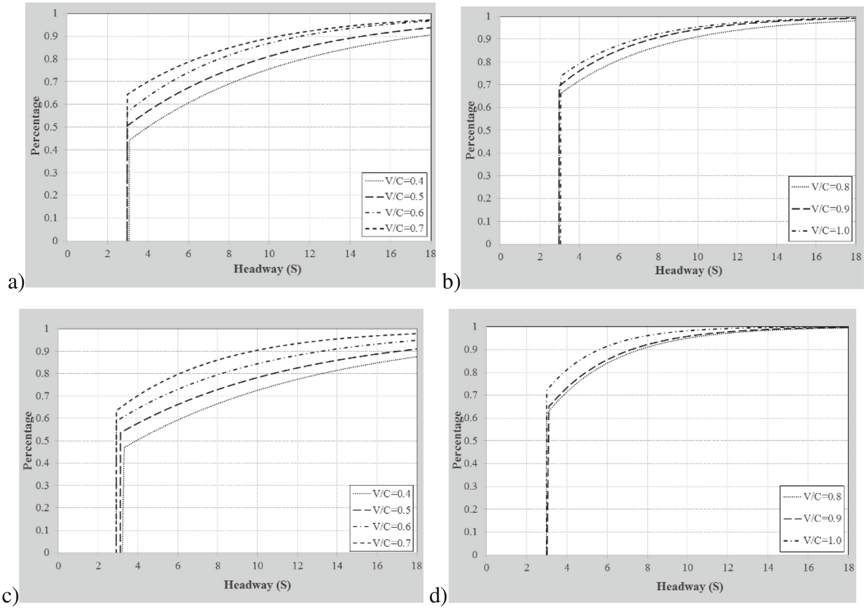


Fig. 3. Estimated distribution function (Cowan's M3 model) for 'study site-1': (a) Direction 1-Moderate to heavy flow; (b) Direction 1-Heavy flow; (c) Direction 2-Moderate to heavy flow and (d) Direction 2-Heavy flow.

The goodness-of-fit of the fitted distributions was visualized by Q-Q (quantile-quantile) plots (Fig. 4). The plot demonstrates a satisfactory agreement between empirical and estimated quantiles in view of the fact that most of the data points are very close to the straight line, except the points that are derived under congested state of flow. Those points are scattered widely above and below the 45^0 plots and indicate an incompatibility of Cowan's M3 distribution at such flow level under mixed traffic.

Further, a pilot study was conducted on a different highway section (study site-2) to observe the agreement between model outcomes and the field data. As the Cowan's M3 was found incompatible under congested state of flow corresponding to v/c ratio of 0.9 and 1.0, flow levels that correspond to v/c ratio of 0.7 and 0.8 were chosen while collecting traffic data at 'study site-2'. The amount of accuracy was computed and expressed in terms of standard error of estimate (SEE), a measure of the dispersion or variability. The calculated SEE values are: 0.022 (Fig. 5a), 0.056 (Fig. 5b), 0.039 (Fig. 5c) and 0.044 (Fig. 5d). The values are reasonably small and, thereby, it can be concluded that most of the observed values cluster fairly closely to the anticipated line of agreement.

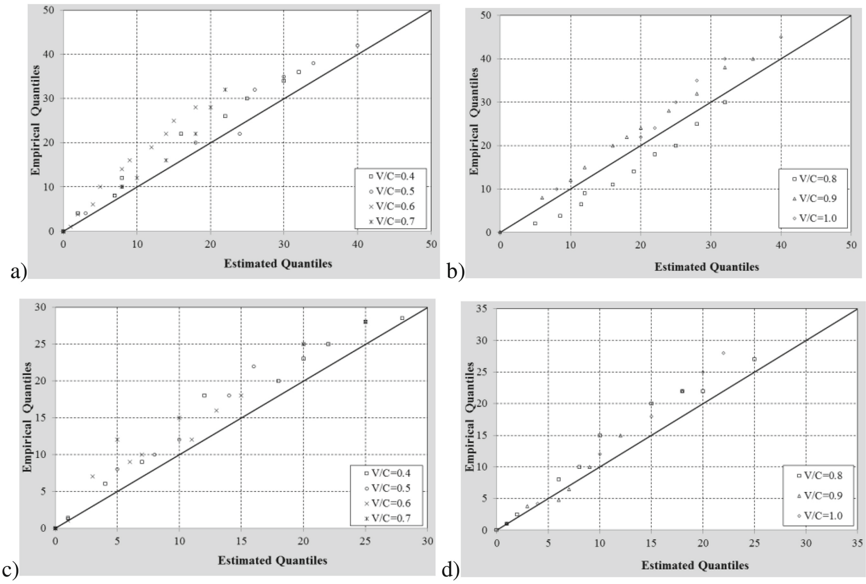


Fig. 4. A Q-Q plot comparing empirical and estimated (Cowan's M3 model) quantiles for 'study site-1': (a) Direction 1-Moderate to heavy flow; (b) Direction 1-Heavy flow; (c) Direction 2-Moderate to heavy flow and (d) Direction 2-Heavy flow.

4.2 Model Interpretation

The current study was conducted under mixed traffic condition where headways are observed to have two components, 'following' and 'free'. Thus, composite headway distribution was selected to model headway data. The current study considered Cowan's M3 because of its simplicity over Cowan's M4 and DDNE. Based on goodness-of-fit and also, hypothesis test it was found that such distribution cannot be considered for modeling headways that correspond to unstable flow.

A further attempt was made to compare the probabilities obtained based on the composite and single models. Probabilities based on single distribution model were obtained from a study conducted on similar traffic characteristics in India [14]. Table 2 demonstrates that probabilities for longer headways do not vary significantly. However, probability of shorter headways obtained based on single distribution was lower than those obtained from composite model. Accordingly, composite model exhibits its aptness in modeling headways under mixed traffic in view of the fact that considerable amount of slower vehicles in the traffic stream consequences higher proportion of short headways which a single model cannot estimate. However, at unstable flow since all the vehicles move at following mode such values did not differ considerably.

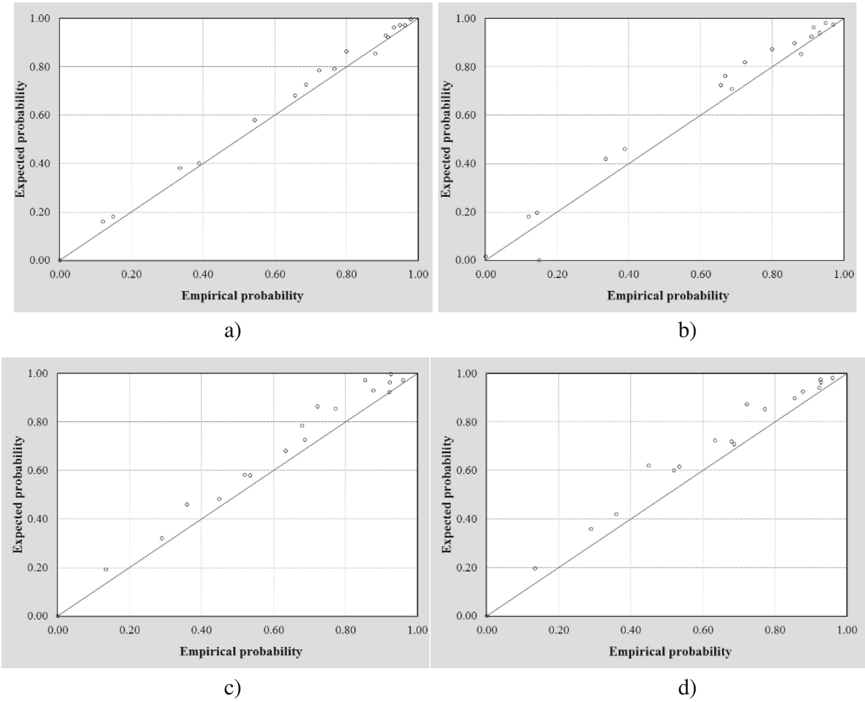


Fig. 5. Agreement between the probabilities estimated using Cowan's M3 model and field data collected from the 'study site-2': (a) direction 1-model and field data; (b) direction 1-model and direction 2-field data; (c) direction 2-model and Direction 1-field data and (d) direction 2-model and field data.

Table 2. Comparison of headway probabilities: composite model (current study) and single model (a case study).

	Headway (Sec)	V/C ratio						
		0.4	0.5	0.6	0.7	0.8	0.9	1.0
Composite distribution	<3.0	0.438	0.506	0.567	0.645	0.657	0.696	0.733
	<4.5	0.529	0.597	0.665	0.725	0.743	0.787	0.816
	<7.5	0.671	0.733	0.798	0.834	0.856	0.896	0.913
	<10.5	0.770	0.822	0.879	0.901	0.919	0.949	0.958
	<13.5	0.839	0.882	0.927	0.941	0.954	0.975	0.981
	<16.5	0.887	0.922	0.956	0.964	0.974	0.987	0.991
Single distribution [14]	<3.0	0.312	0.352	0.412	0.523	0.567	0.621	0.668
	<4.5	0.451	0.512	0.572	0.672	0.714	0.762	0.801
	<7.5	0.568	0.625	0.712	0.772	0.788	0.862	0.911
	<10.5	0.712	0.778	0.822	0.862	0.888	0.935	0.951
	<13.5	0.758	0.832	0.884	0.912	0.925	0.952	0.963
	<16.5	0.868	0.887	0.916	0.935	0.952	0.962	0.975

5 Conclusions

Frequent formation of platoons on two-lane roads under mixed traffic increases the proportion of follower vehicles in the traffic stream and shorter headways as a consequence. While characterizing headways statistically using a distribution function, mixed distribution was observed to exhibit its aptness. On the basis of critical review of these distributions and experiences with mixed traffic, the current study considered Cowan's M3 for the purpose of modeling. Empirical investigations revealed the fact that this distribution fits the data well up to a flow level that corresponds to moderate to heavy. Since, at congested state of flow almost all the vehicles start moving in following, Cowan's M3 cannot describe the data set appropriately. The current study thus creates a starting point of further initiatives aimed at developing a robust method for modeling headways under such traffic.

Acknowledgement. The part of the analysis presented in the paper has used the data collected in the CSIR-CRRI, New Delhi sponsored project "Development of Indian Highway Capacity Manual (INDO-HCM)". The authors sincerely acknowledge CSIR-CRRI. The authors are grateful to the anonymous referees for their suggestions to improve the paper.

References

1. Ye, F., Zhang, Y.: Vehicle type-specific headway analysis using freeway traffic data. Transp. Res. Record. **2124**, 222–230 (2009)
2. Zhang, G., Wang, Y., Wei, H., Chen, Y.: Examining headway distribution models with urban freeway loop event data. Transp. Res. Record. **1999**, 141–149 (2007)
3. Panichpapiboon, S.: Time-headway distributions on an expressway: case of Bangkok. J. Transp. Eng. **141**(1), 05014007 (2014)
4. Abtahi, S.M., Tamannaei, M., Haghshenash, H.: Analysis and modelling time headway distributions under heavy traffic flow conditions in the urban highways: case of Isfahan. Transport **26**(4), 375–382 (2011)
5. Chandra, S., Kumar, R.: Headway modelling under mixed traffic on urban roads. Road Transp. Res. **10**(1), 61–79 (2001)
6. Arasan, V.T., Koshy, R.Z.: Headway distribution of heterogeneous traffic on urban arterials. J. Inst. Eng. (India) **84**, 210–215 (2003)
7. Khasnabis, S., Heimbach, C.L.: Headway distribution models for two-lane rural highways. Transp. Res. Record. **772**, 44–51 (1980)
8. Al-Ghamdi, A.: Analysis of time headways on urban roads: case study from Riyadh. J. Transp. Eng. **127**(4), 289–294 (2001)
9. Luttinen, R.T.: Statistical Analysis of Vehicle Time Headways. Ph.D. thesis, Helsinki University of Technology, Finland (1996)
10. Cowan, R.J.: Useful headway models. Transport. Res. **9**(6), 371–375 (1975)
11. Branston, D.: Models of single lane time headway distributions. Transport. Sci. **10**(2), 125–148 (1976)
12. Zhang, G., Wang, Y.: A Gaussian kernel-based approach for modelling vehicle headway distributions. Transport. Sci. **48**(2), 206–216 (2013)
13. Tanyel, S., Yayla, N.: A discussion on the parameters of Cowan M3 distribution for Turkey. Transport. Res. A. **37**, 129–143 (2003)

14. Saha, P., Roy, R., Sarkar, A.K., Pal, M.: Preferred time headway of drivers on two-lane highways with heterogeneous traffic. Transp. Lett. (2017). https://doi.org/10.1080/19427867.2017.1312859
15. Griffiths, J.D., Hunt, J.G.: Vehicle headways in urban areas. Traffic Eng. Control **32**(10), 458–462 (1991)
16. Rengaraju, V.R., Rao, V.T.: Vehicle-arrival characteristics at urban uncontrolled intersections. J. Transp. Eng. **121**(4), 317–323 (1995)
17. Roy, N., Roy, R., Talukdar, H., Saha, P.: Effect of mixed traffic on capacity of two-lane roads: case study on Indian highways. Procedia Eng. **187**, 53–58 (2017)
18. Vasconcelos, L., Silva, A.B., Seco, Á., Silva, J.P.: Estimating the parameters of Cowan's M3 headway distribution for roundabout capacity analyses. Balt. J. Road Bridge E. **7**, 261–268 (2012)
19. National Academics: Highway Capacity Manual. TRB, Washington D.C. (2010)
20. Myung, I.J.: Tutorial on maximum likelihood estimation. J. Math. Psychol. **47**, 90–100 (2003)
21. Wang, X., Peng, Z.: Method of moments for estimating uncertainty distributions. J. Uncertainty Anal. Appl. **2**(1), 5 (2014)
22. Kominek, Z.: Minimum chi-squared estimation of stable distributions parameters: an application to the Warsaw Stock Exchange. J. Appl. Stat. **29**(5), 729–744 (2002)
23. Luttinen, R.T.: Properties of Cowan's M3 headway distribution. Transp. Res. Record. **1678**, 189–196 (1999)

Shared Risks at Interface Between Railway Undertakings and Infrastructure Managers

Piotr Smoczyński and Adrian Gill(✉)

Poznan University of Technology, Pl. Skłodowskiej-Curie 5, 60-965 Poznań, Poland
{piotr.smoczynski,adrian.gill}@put.poznan.pl

Abstract. Organisational changes on European railways significantly affected the way how the railway transportation works. The state railways were divided and new entities came into market. The cooperation between the actors is to be managed with the help of mandatory safety management systems, based on risk management principles. In this paper, we propose a novel notation for describing 'shared risks', i.e. the hazards which sources, possible risk reduction measures and/or consequences are not covered by one safety management system. The standardised description of such hazards should improve understanding of their nature and thus make the application of risk reduction measures more effective.

Keywords: Risk · Safety · Safety management · Safety management system Railway safety

1 Introduction

Organisational changes on European railways, which have been taking place since early 1990-ties, significantly affected the way how the railway transportation is functioning. Former state railways have been vertically separated and divided into railway infrastructure managers and railway undertakings – independent entities which should compete and cooperate at the same time. Competition is needed for lowering the prices and generally making this means of transport more accessible, as proposed by the European Commission in the relevant white paper [1]. Cooperation, however, results directly from the railway system properties, where vehicles of several undertakings must share the same tracks and other infrastructure elements.

The opposing goals of competition and cooperation are to be managed with the help of safety management systems, which were made mandatory for all the infrastructure managers and railway undertakings. The tendency of introducing such systems can be observed in many domains of human activity and is driven by 'audit society' phenomena [2]. The constant anxiety of being verified results in people securing themselves with the use of more and more sophisticated procedures [3]. The development of safety management systems in high-risk domains such as railways was investigated e.g. by Reiman et al. [4] and Grote [5] and is only possible if the risk is being managed properly [6]. The subject of risk management in the railway system was examined lately e.g. by Bureika et al. [7],

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_19

as well as Smoczyński and Kadziński [8]. It should be emphasised that all the aforementioned research was constrained to managing safety inside one business entity.

In this paper, we propose a novel notation for describing 'shared risks', i.e. the hazards which sources, possible risk reduction measures and/or consequences are not covered by one safety management system. In Sect. 2 we present our approach for understanding the basic terminology; in Sect. 3 we introduce the notation with subsequent discussion of the results. The paper ends with conclusions in Sect. 4.

2 Theoretical Background

Even though safety-related research has been successfully conducted for many years, there is still no unified terminology for describing these issues in an unambiguous way. There are many publications devoted to make the understanding more unified, such as [9] for definition of 'safety', [10, 11] for 'risk' and [12, 13] for 'safety barriers'. In the following, we will discuss terms applied by us throughout the paper.

2.1 Relation Between 'Risk' and 'Hazard'

Aven [10] compared definitions of risk used across some scientific disciplines. He proposed assigning them into three distinct categories:

- risk as a concept based on events, consequences and uncertainties;
- risk as a modelled, quantitative concept; and
- risk descriptions.

The approach used in Polish railway context can be assigned to the category which treats risk as a modelled, quantitative concept [14–16]. According to these definitions, risk of a hazard depends on the probability of the activation of the hazard under assessment, as well as on severity of its foreseen consequences. As a result, risk deals for assessing the seriousness of a hazard, and the terms 'risk' and 'hazard' have noticeably different meanings.

The 'hazard' itself is defined in the aforementioned Polish publications as 'a state of the analyses domain which can lead to loss or damage'. What is important, the state can be dependent on one or several variables. It means that each hazard can have one or several different causes and all the causes must occur at the same time to activate the hazard. This is different to e.g. bow-tie methodology, where each part of the diagram can independently account for the top event.

The causes of hazards, called hazard sources or risk factors, can be defined as 'physical, chemical, biological, psychophysical, organisational or human factors which presence, state or properties are the cause for formulating a hazard' [14]. We believe that, in addition to the causes such as 'car driver distraction' and 'train approaching a level-crossing', we should take into consideration a distinctive set of hazard sources, which is possible abnormal operation or inactivity of risk reduction measures, discussed in Sect. 2.2.

2.2 Risk Reduction Measures

To avoid the hazard activation or to mitigate its consequences, some safety actions are needed. According to the Commission Regulation (EU) 402/2013 [17], which constitutes legal basis for risk management on European railways, a set of these safety actions is called 'safety measures' or 'risk control measures'. It should be added, that the safety measures activity usually comes down to the elimination of hazard sources, breaking their impact path (i.e. isolating the source or isolating the receiver on which the source has an impact), and informing about hazard source activity.

Alternatively, the risk reduction measures can also mitigate the severity of the potential consequence of the hazard activation, although this kind of measures will not be considered here; further discussion on this topic can be found e.g. in [12, 18].

2.3 Consequences

The consequences of a hazard activation is the loss or damage of elements located in the analyses domain where the hazard was identified. The loss or damage is typically understood as death, physical injury or health impairment, loss of property or destruction of an object, environmental degradation, and economic loss [19].

Sometimes more specific type of consequences are also included in the definition, e.g. business interruption costs [20] or loss/compromise of data that the organization is required to protect, like confidential data and customers' personally identifiable information [21]. Additionally, several types of consequences can be differentiated, such as:

- instant and non-instant losses [22],
- direct and indirect consequences [23],
- residual and inherent, as used e.g. in bow-tie simulation software.

In the following, we will treat consequences assets of loss or damage that are significant in terms of the railway system as a whole, not just for one particular business entity. Depending on the hazard, all the aforementioned consequence types can be taken into consideration.

3 Results

The definitions discussed in Sect. 2 allow us to present our understanding of safety-related terminology in graphical form, see Fig. 1.

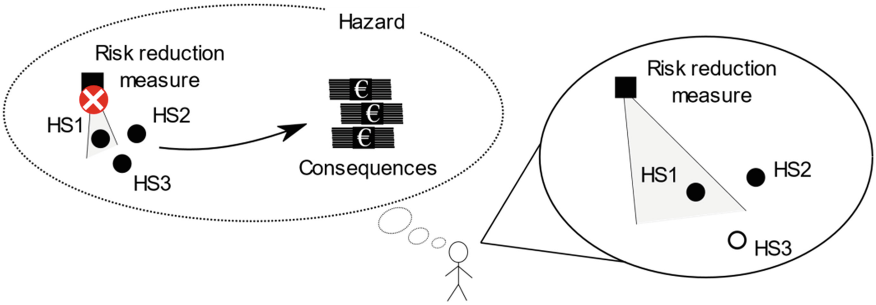


Fig. 1. Schematic representation of the hazard (on the left), formulated by a risk analyst, who considers the state of the domain (on the right).

In the Fig. 1 there is a person performing risk analysis. He or she looks on the domain (right side of Fig. 1), seeing two hazard sources, denoted as HS1 and HS2. There is also one risk reduction measure, which is applied on the hazard source HS1 to reduce the possibility of its presence in the domain. In the following, we will call the risk reduction measure 'active' when it fulfils its task and 'inactive' if it is not working or working incorrectly. Additionally, the person is aware of a hazard source HS3, which can, in some circumstances, be present in the analyses domain.

The observed state of the domain, with RRM1 active and HS3 not present, cannot lead to any negative consequences (expressed e.g. in terms of money). Similarly, following domain states:

- RRM1 active and HS3 present,
- RRM1 not active and HS3 not present,

would not lead to negative consequences. Still, the risk analyst knows that the losses can happen if the risk reduction measure is not active and the HS3 arises. Therefore, he or she identifies and formulates a hazard, presented graphically on the left side of Fig. 1. In the following figures, we will use this representation, omitting the symbol "X" for more clarity.

The relationship between hazard sources, hazards, risk reduction measures and consequences are beginning to be much more complex if they are in different subdomains, e.g. covered by different safety management systems. In case of railways we can distinguish three such subdomains; two of them are covered by the safety management system of an infrastructure manager and a railway undertaking, whereas the third subdomain are the surroundings: road traffic, pedestrians, inhabitants, etc. Please note that both railway entities are obliged by the European legislation to take into consideration also the hazards arising from the activities of other parties, external to the railway system. In Fig. 2 there is a graphical representation of two sample groups of hazard sources at the junction of the subdomains.

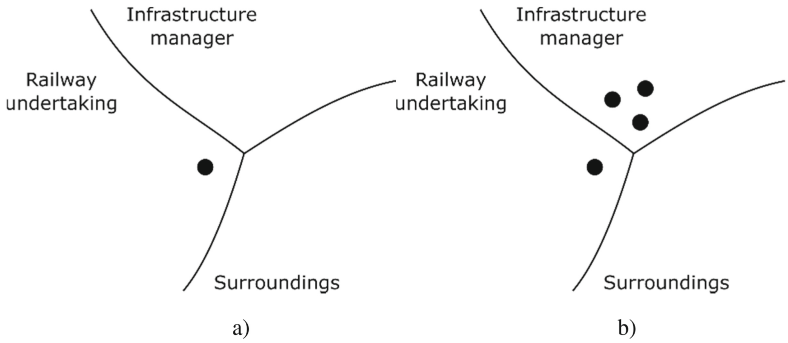


Fig. 2. Schematic representation of hazard sources (black dots) located in different subdomains: (a) hazard with one source in one subdomain; (b) hazard with sources in two subdomains.

The hazard sources can be in one or more of any of the subdomains under consideration. The total number of possible combinations of hazard source distributions CHS can be established with the following equation:

$$C_{\mathrm{HS}} = \sum\nolimits_{i=1}^{3} \binom{3}{i} = \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 3 + 3 + 1 = 7. \tag{1}$$

As in the case of the hazard sources, also possible inactive risk reduction measures and consequences can be located in one or more subdomains. Obviously:

$$C_{\mathrm{RRM}} = C_{\mathrm{CON}} = C_{\mathrm{HS}} = 7, \tag{2}$$

where C_{RRM} denotes the number of possible inactive risk reduction measure distributions and C_{CON} denotes the number of possible consequence distributions. The number of distinguishable hazard structures amounts to:

$$C_{\mathrm{H}} = 7 \cdot 7 \cdot 7 = 392, \tag{3}$$

where C_H denotes the number of possible hazard structures, i.e. distribution of hazard sources, possible risk reduction measures and consequences across the subdomains of infrastructure manager, railway undertaking and surroundings. In three cases, all the three elements of hazard structure are located in one subdomain. The remaining 389 structure types fall into the category of 'shared risks'.

Unambiguous notation of the hazard structure can be achieved with help of a hazard structure matrix, defined as follows:

$$H_0 = \begin{bmatrix} a_{HS,IM} & a_{HS,RU} & a_{HS,SU} \\ a_{RRM,IM} & a_{RRM,RU} & a_{RRM,SU} \\ a_{CON,IM} & a_{CON,RU} & a_{CON,SU} \end{bmatrix}, \tag{4}$$

where HS – hazard source, RRM – risk reduction measure, CON – consequences, IM – infrastructure manager, RU – railway undertaking, SU – surroundings.

The entry is set to 1 if a hazard source/risk reduction measure/consequences is/are located in the subdomain of an infrastructure manager, railway undertaking and surroundings, respectively; in the opposite case, the entry is set to 0.

We will now discuss two examples from practice. Firstly, let us assume that there is a railway undertaking which has identified a hazard ‘possibility of rolling stock damage due to wild animals on track’. This hazard activates regularly, causing considerable losses due to delays, cancelled trains and rolling stock damage. Schematic representation of this hazard is shown in Fig. 3.

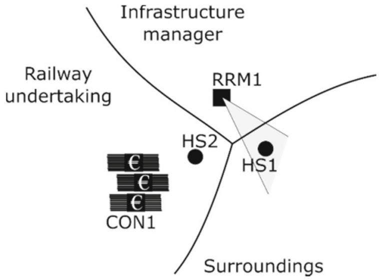


Fig. 3. Schematic representation of a hazard ‘possibility of rolling stock damage due to wild animals on track’ (explanation of abbreviations in text).

For the hazard shown in Fig. 3, there are two hazard sources which take part in formulating the hazard, i.e. a wild animal coming from forest (HS1) and a moving train (HS2). The railway undertaking can do nothing to prevent the animals from crossing the railway line, so the only option it has for lowering the risk of this hazard is to stop the service, which is not plausible. However, the infrastructure manager has a variety of options, e.g. installing a fence or an animal repellent device (RRM1). The manager is expected to take such measures, although all the primary consequences (CON1) are being incurred by the railway undertaking; it must pay compensation to the passengers, organise replace bus service and repair the vehicles. The structure of this hazard can be denoted by the following matrix:

$$H_{\text{animals}} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{5}$$

The situation is different for the hazard ‘possibility of health deterioration due to railway noise’, schematically shown in Fig. 4.

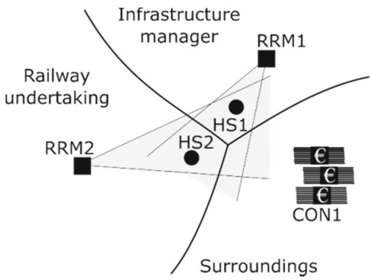


Fig. 4. Schematic representation of a hazard ‘possibility of health deterioration due to railway noise’ (explanation of abbreviations in text)

The consequences (CON1) are located in the surroundings in form of health problems, fall in property values etc. Neither the infrastructure manager nor the railway undertaking suffer directly from the activation of this hazard, although the hazard sources are located in their subdomains, mainly in the wheel-rail interface (HS1 and HS2). Similarly, to the first hazard, the most plausible risk reduction measure can be applied by the infrastructure manager in form of noise barriers or additional rail grinding (RRM1). Still, also the railway undertaking can take risk reduction measures, e.g. exchanging cast iron brake blocks by blocks made from composite materials (RRM2).

Alternatively, the hazard structure can be denoted by the following matrix:

$$H_{\text{noise}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

Using the matrixes instead of the graphical representation allows to propose in the future some methods for analytical determination of the characteristics of particular hazards. It may be useful e.g. to clearly divide the cost of introducing new risk reduction measures if one of the parties considers that the risk level is unacceptable.

4 Conclusions

The idea of ‘shared risks’ is specific to railways. Different business entities are legally obliged to manage risk of hazards, which sources, possible risk reduction measures and/or consequences are split between their safety management systems. There is a need for standardisation and formalisation of data wherever it is necessary to collect it, especially for subsequent data processing (e.g. in the form of queries in databases). Therefore, we propose a form of hazard structure description which allows to present a complex hazard ‘picture’.

Linguistic representation of hazard structure, which has been applied so far e.g. in rail transport systems in Poland, does not facilitate the execution of the indicated processes. Additionally, the proposed hazard structure description provides new opportunities to characterize the hazard in numerical terms. In this perspective, it can become a vital supplement for risk measurement models.

Knowing the hazard structure is important for harmonising hazard records of different actors and allows for better understanding of the situation. The proposed form of hazard picture is meant to motivate the risk analyst to look on the hazards in a more abstract level and consider its consequences across the borders of railway safety management systems. We believe that the standardisation and formalisation of the presented hazard structure will contribute to its wide application across railway entities. The idea can be also further developed to be the basis for splitting costs of introducing and maintaining risk reduction measures.

Acknowledgements. The research work was financed with the means of statutory activities of Faculty of Machines and Transport, Poznan University of Technology, 05/52/DSPB/0259.

References

1. European Commission: Roadmap to a single European transport area – towards a competitive and resource efficient transport system (2011)
2. Power, M.: The Audit Society. Rituals of Verification. Oxford University Press, Oxford (1997)
3. Jeffcott, S., Pidgeon, N., Weyman, A., Walls, J.: Risk, trust, and safety culture in U.K. train operating companies. Risk Anal. **26**, 1105–1121 (2006)
4. Reiman, T., Rollenhagen, C., Pietikäinen, E., Heikkilä, J.: Principles of adaptive management in complex safety-critical organizations. Saf. Sci. **71**, 80–92 (2015)
5. Grote, G.: Safety management in different high-risk domains - all the same? Saf. Sci. **50**, 1983–1992 (2012)
6. Demichela, M., Piccinini, N., Romano, A.: Risk analysis as a basis for safety management system. J. Loss Prev. Process Ind. **17**, 179–185 (2004)
7. Bureika, G., Gaidamauskas, E., Kupinas, J., Bogdevičius, M., Steišūnas, S.: Modelling the assessment of traffic risk at level crossings of Lithuanian railways. Transport **4142**, 1–9 (2016)
8. Smoczyński, P., Kadziński, A.: Introduction to the risk management in the maintenance of railway tracks. J. Mech. Transp. Eng. **68**, 65–80 (2016)
9. Aven, T.: What is safety science? Saf. Sci. **67**, 15–20 (2014)
10. Aven, T.: A holistic framework for conceptualising and describing risk. In: SSARS 2010, pp. 7–14 (2010)
11. Aven, T., Kristensen, V.: Perspectives on risk: review and discussion of the basis for establishing a unified and holistic approach. Reliab. Eng. Syst. Saf. **90**, 1–14 (2005)
12. Sklet, S.: Safety barriers: definition, classification, and performance. J. Loss Prev. Process Ind. **19**, 494–506 (2006)
13. Gill, A.: The method of analysis of safety systems dedicated to the systems in rail vehicle. Logist. Transp. **18**, 67–76 (2013)
14. Kadziński, A.: Studium wybranych aspektów niezawodności systemów oraz obiektów pojazdów szynowych. Wydawnictwo Politechniki Poznańskiej, Poznań (2013). [in Polish]
15. Kadziński, A.: Integracja pojęć. W: Krystek, R. (red.) Zintegrowany system bezpieczeństwa transportu. T. 2, Uwarunkowania rozwoju integracji systemów bezpieczeństwa transportu, Warszawa, pp. 285–289 (2009)

16. Chruzik, K.: Zarządzanie bezpieczeństwem w transporcie kolejowym, Radom (2014)
17. Commission Implementing Regulation (EU) No 402/2013 of 30 April 2013 on the common safety method for risk evaluation and assessment and repealing Regulation (EC) No 352/2009. European Commission (2013)
18. Gill, A.: Koncepcja zastosowania reguł decyzyjnych w doborze środków redukcji ryzyka zagrożeń. Pr. Nauk. Politech. Warsz. Transp. 181–190 (2013)
19. Jamroz, K.: Metoda zarządzania ryzykiem w inżynierii drogowej. Wydawnictwo Politechniki Gdańskiej, Gdańsk (2011)
20. Center for Chemical Process Safety: Guidelines for developing quantitative safety risk criteria (2009)
21. Ucedavélez, T., Morana, M.M.: Risk Centric Threat Modeling (2015)
22. Zadakbar, O., Khan, F., Imtiaz, S.: Development of economic consequence methodology for process risk analysis. Risk Anal. **35**, 713–731 (2015)
23. Santos, J.R., May, L., Haimar, A.El: Risk-based input-output analysis of influenza epidemic consequences on interdependent workforce sectors. Risk Anal. **33**, 1620–1635 (2013)

Risks Associated with the Use of High-Strength Titanium Alloys in Transportation Systems

Mykola Chausov[1], Pavlo Maruschak[2(✉)], Olegas Prentkovskis[3], and Myroslav Karpets[4]

[1] National University of Life and Environmental Sciences of Ukraine, Kyiv 03041, Ukraine
[2] Ternopil National Ivan Pul'uj Technical University, Ruska 56 str., Ternopil 46001, Ukraine
Maruschak@tu.edu.te.ua
[3] Vilnius Gediminas Technical University, Plytinės g. 27, 10105 Vilnius, Lithuania
olegas.prentkovskis@vgtu.lt
[4] Frantsevich Institute for Materials Science Problems, National Academy of Sciences of Ukraine, Krzhizhanovs'kogo 3 str., Kyiv 03680, Ukraine

Abstract. On an example of testing sheet high-strength (α + β) titanium alloys with different percentages of α and β phases, the danger of using titanium alloys with a large α phase content in transportation systems subjected to impact-oscillatory loading is shown. Under impact-oscillatory loading, dynamic nonequilibrium processes (DNP) with self-organization of the structure can be realized in titanium alloys. As a result, depending on the initial percentage of α and β phases in alloys, the impact-oscillatory loading can significantly affect fluctuations in the initial plastic deformation of the alloys upward or downward without appreciably reducing the strength of the alloy.

Keywords: Impact-oscillatory loading · Dynamic nonequilibrium processes (α + β) titanium alloys · Mechanical properties

1 Introduction

Almost all modern critical structures use high-strength titanium alloys as structural materials [1, 2]. Titanium alloys are used, inter alia, for airborne and aerospace systems of various purpose [3]. High strength, low density, high specific strength, wide temperature range of titanium alloys can significantly reduce the weight of structures of transportation systems and increase the reliability of their operation. At present, structural and special titanium alloys with varying strength and, most importantly, high ductility have been developed. These alloys include two-phase high-strength titanium alloys VT23 ($\sigma_{us} \geq 1150$ MPa, $\delta = 15\%$) and VT23 M ($\sigma_{us} \geq 1100$ MPa, $\delta = 20\%$) [4, 5]. To achieve such high parameters of the mechanical properties of these two-phase titanium alloys, the developers of semifinished products implement complex multi-stage thermo-mechanical processing (TMP) modes with different percentages of α and β phases and with various additives of alloying elements [6, 7]. One should pay attention to the fact that most stages of complex TMP modes occur in conditions far from thermodynamic

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_20

equilibrium. Therefore, as early as at the stages of manufacturing semifinished products of titanium alloys, dynamic non-equilibrium processes (DNP) are realized in melts, which also contribute to the appearance of unique mechanical properties of titanium alloys of the class in question [7, 8]. On the other hand, the real loading modes of transportation systems can also realize dynamic nonequilibrium processes with self-organization of the structure in titanium alloys. The authors were the first to establish experimentally that such real modes definitely include impact-oscillatory loading with a loading frequency of 1 … 2 kHz [9]. Such loading modes can lead to significant changes in the initial structure of alloys carefully prepared by technological methods and, consequently, to changes in the initial mechanical properties of industrial titanium alloys used in transportation systems. Previous studies on titanium alloys of different classes showed that, for example, for a two-phase titanium alloy VT22 after impact-oscillatory loading, the plastic deformation increases significantly in comparison with the initial state, while the strength practically does not decrease [10]. Similar tests on a submicrocrystalline titanium alloy VT1-0, on the contrary, have found a significant decrease both in plastic deformation and strength after the realization of DNP due to impact-oscillatory loading [11]. The limited experimental data in this area of research does not allow to unequivocally answering the question of what parameters of the initial structure of two-phase titanium alloys are most sensitive to dynamic nonequilibrium processes realized during the operation of transportation systems due to impact-oscillatory loading.

The purpose of this work is to assess the impact of the initial percentage ratio of α and β phases in the sheet titanium alloys VT23 and VT23 M on changes in the mechanical properties of alloys after impact-oscillatory loading.

2 Materials and Methods

The method of mechanical testing is implemented on the basis of a modified hydraulic setup ZD-100Pu for static testing and is described in detail in [12, 13]. The main idea of the proposed technique is the high-speed stretching of the material with a high frequency oscillation (several kilohertz) applied to it, which corresponds to the own frequency of the testing machine. Structurally this is achieved due to the fact that, in addition to the outer contour (loaded frame of the testing machine), an inner contour is introduced into the testing machine. The inner contour is a simple statically indeterminate structure in the form of three parallel elements that are loaded simultaneously – a central specimen and two satellite specimens (brittle samples) of different cross sections made from hardened steels 65G or U8 … U12. When this structure is stretched, the satellite specimens are destroyed, and impulse energy is introduced into the material of the specimen under study. The satellite specimens can be activated at any stage of the preliminary static stretching, so that the effect of pulse introduction of energy on the degradation of mechanical properties due to material damage at static stretching can be investigated. By varying the initial diameter of the satellite specimens, it is possible to control the intensity of the pulse introduction of the power energy into the material.

Mechanical tests were conducted on specimens (Fig. 1) from sheet industrial titanium alloys VT22 and VT23 M with a thickness of 3 mm. The strain measurement base was 16 mm.

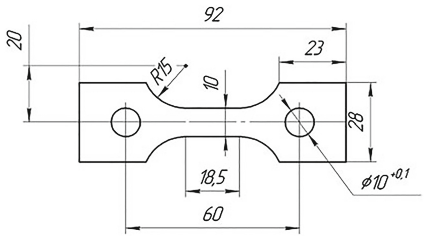


Fig. 1. Test specimen.

The mechanical properties of the alloys in the initial state are given in Table 1.

Table 1. Mechanical properties of titanium alloys VT23 and VT23 M.

Titanium alloys	Mechanical properties		
	σ_{ys}, MPa	σ_{us}, MPa	δ, %
VT23	980–1180	1080–1280	15
VT23M	1000–1150	1080–1180	20

To evaluate the percentage ratio of α and β phases in the titanium alloys VT23 and VT23 M, an X-ray study was used in the initial state. The X-ray study was performed in monochromatic Cu Kα-radiation on a DRON-UM1 diffractometer. A single crystal of graphite was installed on a diffracted beam as a monochromator. The diffractograms were taken by step scanning in the range of angles 2Θ 10–90°. The scan step was 0.05°, the exposure time at a point was 3 s. The data of the diffractometric experiment were processed using a program for full spectrum analysis of X-ray spectra from a mixture of polycrystalline phase components of Powder Cell 2.4. In full-profile analysis of diffraction patterns, texture patterns were recorded using the March-Dollase model [14, 15]. The correction for intensity of any maximum hkl in this model was performed by introducing an effective multiplier of repetition Mhkl:

$$M_{hkl} = \sum_{i=1}^{m} \left(\tau^2 \cdot \cos^2 \varphi_i + \tau^{-1} \cdot \sin^2 \varphi_i\right)^{-3/2}, \tag{1}$$

where τ is the correction factor of the texture, m is the crystallographic multiplier of repetition of this family of equivalent planes $\{hkl\}$, φ is the smallest value of the angle between the normal to a given plane $(hkl)_i$ and a given vector of the texture. According to Eq. (1), in the absence of the texture, $\tau = 1$, at $0 < \tau < 1$, the value $M_{hkl} > 1$, and the intensity of maximum *hkl* increases compared to the non-textured specimen, and at $\tau > 1$, the value $M_{hkl} < 1$.

3 Results and Discussion

The results of the diffractometric study have shown (Fig. 2) that the β phase occupies 43% of weight and the α phase – 57% of weight in the titanium alloy VT23. The β phase occupies 22% of weight and the α phase – 78% of weight in the titanium alloy VT23 M. It is notable that both phase components in specimens have a texture along the crystallographic direction (002). The latter can result from the rolling of specimens or other mechanical impact.

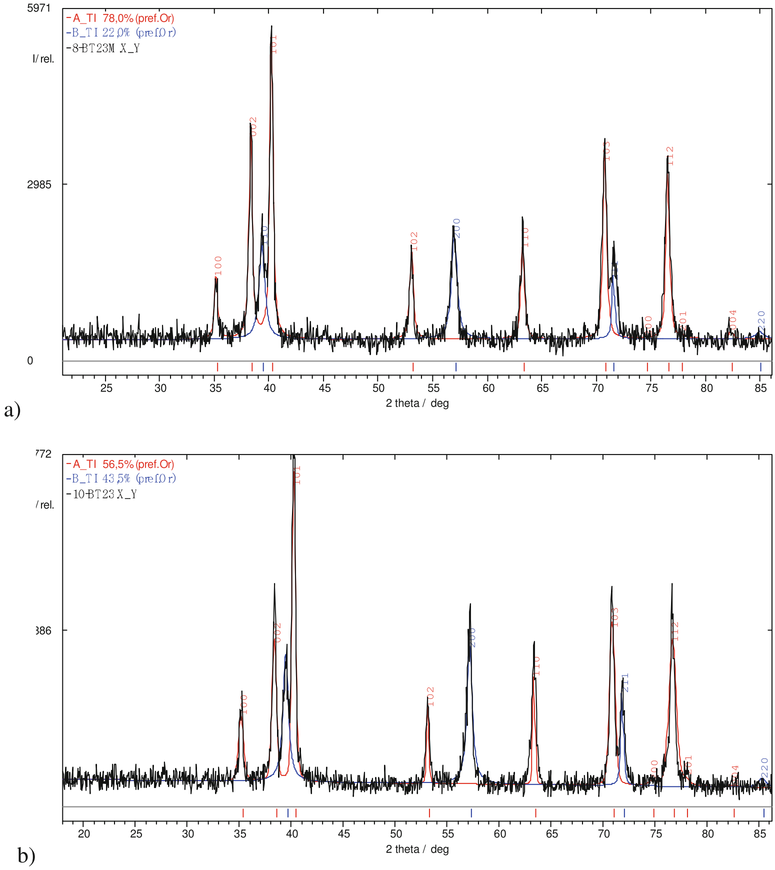


Fig. 2. The results of the diffractometric study on the percentage of α and β phases in the investigated titanium alloys: (a) alloy VT23 M; (b) alloy VT23.

Attention should be paid to the periods of the lattice of phases in the investigated specimens of alloys. Thus, in the transition from the specimens of the alloy VT23 M to the specimens of the VT23 alloy, the volume of the elemental cell of the hexagonal lattice alpha-Ti decreases from 34.84 A^3 to 34.58 A^3, while for the elementary cell of the cubic lattice beta-Ti, the volume decrease from 33.49 A^3 to 33.08 A^3. Such changes can be due to the substitution of the titanium atoms by atoms with a smaller atomic radius, or the formation of a certain number of vacancies in crystalline lattices.

Preliminary experiments on specimens from alloys VT23 and VT23 M have shown that the loading modes, at which the total stress in specimens under the action of preliminary static stretching and additional impulse loading does not exceed the yield strength of the alloy, lead to an insignificant decrease in plastic deformation of the alloys upon repeated static stretching (up to 10%). The strength of alloys is practically unchanged in this case.

Therefore, the method of further testing was as follows. The batches of specimens from alloys VT23 and VT23 M, 17 pieces each, were subsequently loaded with static stretching to the initial loading of 8.0 kN, and further in 2.0 kN increments. Under these fixed static loads, specimens were subjected to a given additional force impulse load $F_{imp.} = 45.0 \ldots 50.0$ kN. Under the pulse introduction of energy into the alloys, slips of plastic deformation $\varepsilon_{imp.}$ were recorded, and after the realization of DNP specimens were completely unloaded. Next, all the tested specimens were stretched statically to fracture, and the effect of pulse introduction of energy on the increase in plastic deformation of alloys as compared to the initial state was determined using the parameter ε_{imp}.

Figure 3 shows the obtained experimental curve for the dependence of $\varepsilon_{plast.}$ on $\varepsilon_{imp.}$ for alloy VT23, and Fig. 4 shows stress-strain diagrams for some specific tested specimens (here, specimen 19 corresponds to the stress-strain diagram under static stretching of the VT23 alloy in the initial state). The analysis of the experimental results obtained showed that for the VT23 alloy the optimum value is $\varepsilon_{imp.} = 3.5 \ldots 4.0\%$.

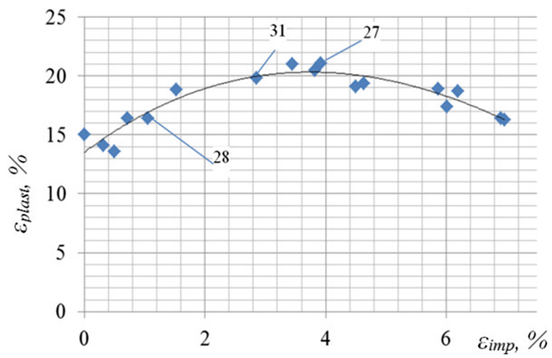


Fig. 3. The curve showing the dependence of plastic deformation of specimens from titanium alloy VT23 upon repeated static stretching on deformation slips in the process of pulse introduction of energy into the alloy.

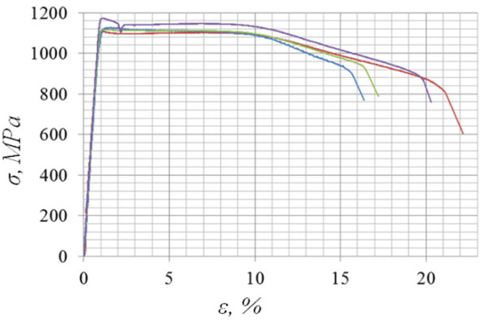


Fig. 4. Stress-strain diagrams of tested specimens from alloy VT23 (explanations are given in the text).

In general, there is a positive effect of pulsed introduction of energy at the stage of plastic deformation of VT23 alloy on the increase in plastic deformation upon repeated static stretching.

Interesting results were obtained for alloy VT23 M. No loading modes, under which there is a noticeable increase in plastic deformation after the DNP realization with a symmetric introduction of the pulse energy, have been recorded for this alloy. Only with $\varepsilon_{imp.} = 0.2 \ldots 1.0\%$ it is possible to reach the upper limit of scatter of the mechanical properties of the alloy under standard static tension, Fig. 5. It should be noted that curves 2y, 5y and 7y practically coincide with curve 1, which corresponds to the upper limit of scatter of the mechanical properties of the alloy under standard static tension. Figure 5 also shows the stress-strain diagram of alloy VT23 M under impulse loading.

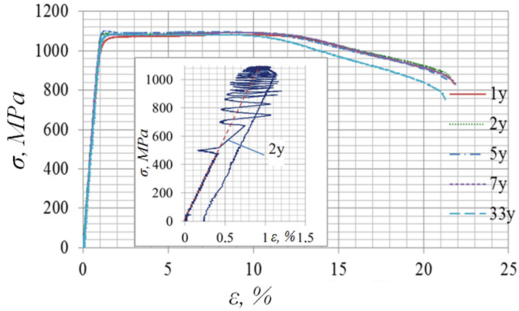


Fig. 5. Stress-strain curves of tested specimens from alloy VT23: curves 1y and 33y correspond to static stretching of specimens in the initial state, curves 2y, 5y and 7y correspond to specimens after impulse loading in the range $\varepsilon_{imp.} = 0.2 \ldots 1.0\%$.

Since in addition to longitudinal vibrations, bending vibrations also arise in the case of impact-oscillatory loading, we pay attention to the relative symmetry of the specimen bend in one direction or another under the action of additional impulse loading on the specimen as a result of practically simultaneous fracture of the satellite specimens.

Specific experiments on the effect of non-simultaneous fracture of satellite specimens, i.e. in the case of asymmetric bending of the specimen in one direction or another during the pulse introduction of energy, revealed an unexpected result, which directly depends on the ratio of α and β phases in the titanium alloys in the initial state. Asymmetric bending of specimens under the pulse introduction of energy practically did not affect plastic deformation of alloy VT23 with a higher content of a more ductile β phase and significantly affected plastic deformation of alloy VT23 M with a significantly lower content of β phase. After the DNP was realized upon repeated static stretching, the plastic deformation of alloy VT23 M decreased to 8.0 … 10.0%, Fig. 6.

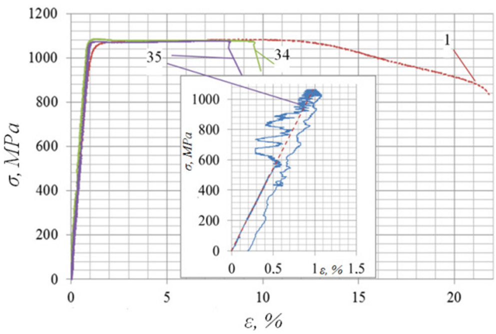


Fig. 6. Stress-strain curves of tested specimens from alloy VT23 M: curve 1 corresponds to static stretching of the alloy in the initial state. Stress-strain diagrams of specimens 34 and 35 correspond to the experimental data on the repeated static stretching after the realization of DNP.

Thus, when high-strength titanium alloys with α phase content of about 80% are used in transportation systems, the risk of a significant reduction in the plastic deformation of alloys under the action of a short-term asymmetric impact-oscillatory loading can be encountered.

4 Conclusions

It is shown that two-phase high-strength titanium alloys VT23 and VT23 M are very sensitive to impact-oscillatory loading with a loading frequency of 1 … 2 kHz, in which dynamic nonequilibrium processes are realized in alloys with self-organization of the structure.

It is found that after the action of asymmetric impact-oscillatory loading, two-phase titanium alloys with a high percentage of α phase of about 80% show a tendency for a dramatic decrease in plastic deformation upon repeated static stretching. Thus, in particular, plastic deformation of alloy VT23 M after the realization of dynamic nonequilibrium processes due to impact-oscillatory loading decreased 2.6 times.

Considering a high probability of realization of impact-oscillatory loading with high frequencies in the process of operation of transportation systems, there are risks associated with using two-phase high-strength titanium alloys with a high percentage of α phase in transportation systems.

References

1. Peters, M., Kumpfert, J., Ward, C.H., Leyens, C.: Titanium alloys for aerospace applications. Adv. Eng. Mater. **5**, 419–427 (2003)
2. Khanna, N., Davim, J.P.: Design-of-experiments application in machining titanium alloys for aerospace structural components. Measur. J. Int. Measur. Confederation **61**, 280–290 (2015)
3. Brewer, W.D., Bird, R.K., Wallace, T.A.: Titanium alloys and processing for high speed aircraft. Mater. Sci. Eng. A **243**(1–2), 299–304 (1998)
4. Khorev, A.I.: High-strength titanium alloy VT23 and its applications in advanced welded and brazed structures. Weld. Int. **24**(4), 276–281 (2010)
5. Khorev, A.I.: Alloying titanium alloys with rare-earth metals. Russ. Eng. Res. **31**(11), 1087–1094 (2011)
6. Illarionov, A.G., Popov, A.A., Illarionova, S.M.: Effect of microalloying, with REM inclusively, on the structure. Phase composition and properties of ($\alpha + \beta$)-titanium alloy. Met. Sci. Heat Treat. **57**(11–12), 719–725 (2016)
7. Gordienko, A.I., Dymovskii, A.S., Kozina, IYu.: Correlation of grain-structure parameters with the properties of thermally strengthened titanium alloys VT6 and VT23. Met. Sci. Heat Treat. **33**(2), 137–141 (1991)
8. Babareko, A.A., Egiz, I.V., Khorev, M.A.: Effect of supplementary alloying of titanium alloy VT23 on its phase composition and tendency to recrystallization. Met. Sci. Heat Treat. **35**(2), 116–120 (1993)
9. Chausov, M.G., Berezin, V.B., Pylypenko, A.P., Hutsaylyuk, V.B.: Strain field evolution on the surface of aluminum sheet alloys exposed to specific impact with oscillation loading. J. Strain. Anal. **50**, 61–62 (2014)
10. Chausov, M., Maruschak, P., Pylypenko, A., Markashova, L.: Enhancing plasticity of high-strength titanium alloys VT 22 under impact-oscillatory loading. Phil. Mag. **97**, 389–399 (2017)
11. Chausov, M.G., Maruschak, P.O., Pylypenko, A.P., Berezin, V.B., Prentkovskis, O.: Structural self-organization of titanium alloys under impulse force action. Mater. Test. **59**(6), 567–569 (2017)
12. Chausov, N.G., Voityuk, D.G., Pilipenko, A.P., Kuz'menko, A.M.: Setup for testing materials with plotting complete stress–strain diagrams. Strength Mater. **36**(5), 532–537 (2004)
13. Chausov, N.G., Pilipenko, A.P.: Influence of dynamic overloading on fracture kinetics of metals at the final stages of deformation. Mechanika **48**, 13–18 (2004)
14. March, A.: Mathematische theorie der regelung nach der korngestalt bei affiner deformation. Z. Kristallogr. **81**, 285–297 (1932)
15. Dollase, W.A.: Correction of intensities for preferred orientation of the March model. J. Appl. Cryst. **19**, 267–272 (1986)

Statistics and Modelling in Transport Applications

Markov-Modulated Linear Regression: Tasks and Challenges in Transport Modelling

Nadezda Spiridovska(✉)

Transport and Telecommunication Institute, 1 Lomonosova Str., Riga 1019, Latvia
Spiridovska.N@tsi.lv

Abstract. Different models and modelling techniques are used in all four stages of the classical transport model. Regression models are widely used in two of them, i.e. in trip generation modelling and transport choice modelling (modal split). Still probabilistic-statistical models generally accept that parameters (regression coefficients in our case) of the model remain unchanged throughout the period of the process of viewing the model. However in practice these parameters usually changing randomly.

Markov-Modulated linear regression brings the idea that the regression model parameters do not remain constant throughout the period of model viewing, but vary randomly with the external environment, the impact of which is described by a Markov chain with continuous time and final state set. This assumption seems quite natural, because the “external environment” is a random in every day’s sense of this word.

This study attempts to identify the advantages and disadvantages of using Markov-modulated linear regression models exactly in transport modelling, comparing with classical regression models and stochastic Markov-chain based models as well. The research gives a vision of Markov-modulated linear regression model’s place in the transportation field, describing new tasks and challenges when facing to the different circumstances such as missing data or big data.

Keywords: Transport modelling · Regression models
Markov-Chain based models · Markov-Modulated linear regression
External environment

1 Introduction

Transports modelling leading specialists, who have great works in this industry, are Juan de Dios Ortúzar and Willumsen [8], Hensher and Button [4], Jaume Barceló Bugueda, Tomer Toledo and others. Juan de Dios Ortuzar [8] describing transport planning process and its importance particularly emphasizes the use of analytical models in decision-making process. There is a huge amount of different models used in transport planning process and the choice of modelling methodology depends on the conceptual model structure, advances in statistical modelling, the data availability, and the computing capability.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_21

Since the object of the investigation is the Markov-modulated linear regression model (which is a combination of regression model and Markov chain based model) and its capabilities, it is natural firstly to describe the model itself (Sect. 2), then consider the scope of regression models and Markov chain based models application within the 4-step classical transport model (Sects. 3 and 4).

2 Markov-Modulated Linear Regression: The Main Idea

The term 'Markov-Modulated' means that a random environment in which model operates is presented as a continuous-time homogeneous irreducible Markov chain $J(t)$, $t \geq 0$, with_finite state set $N = \{1, 2, \ldots, k\}$ (Pacheco et al. [15]). The work in the direction of placing the linear regression into Markov's environment was performed by the professor Alexander Andronov [1].

Using probabilistic-statistical models generally is accepted that parameters of the model remain unchanged throughout the period of the process of viewing the model. In this case it refers to the regression model parameters, i.e., the regression coefficients. However in practice these parameters usually changing randomly. Saying about "random environment" in which investigated object is changing. Thus, this fact it is necessary to take into account developing models to ensure the adequacy of the model to more realistic conditions.

In Markov-Modulated linear regression the idea is offered, which is based on the assumption that the external environment is described by a Markov chain with continuous time. This assumption seems quite natural, because the "external environment" is a random in every day's sense of this word. Let's briefly describe the main idea of the model.

2.1 Model Description

The full description of the Markov-modulated linear regression model can be found in [1]. Let's see the model in matrix notation.

$$Y(T) = \left(Y_1(t_1), \ldots, Y_n(t_n)\right)^T = \begin{pmatrix} \vec{t}_1 \otimes x_1 \\ \vec{t}_2 \otimes x_2 \\ \cdots \\ \vec{t}_n \otimes x_n \end{pmatrix} vec\,\beta + diag(\sqrt{t_1}, \sqrt{t_2}, \ldots, \sqrt{t_n})Z. \tag{1}$$

where $Y_i\,(t)$ are scale responses which are time-additive ($Y_i\,(0) = 0$), n is the number of observations, the $1 \times m$ vector $\vec{t}_i = \left(t_{i,1}, \ldots, t_{i,m}\right)$, which component $t_{i,j}$ means a sojourn time for response Y_i in the state j (it is supposed that model operates in the so-called *external environment*, which has final state space $S = \{s_j, j = 1, \ldots, m\}$, for the fixed state $s_j \in S$, $j = 1, \ldots, m$, (note that $t_i = t_{i,1} + \ldots + t_{i,m}$), the $n \times m$ matrix $T = \left(\vec{t}_i^{\,T}, \ldots, \vec{t}_i^{\,T}\right)^T$, $\bigotimes$ is Kronecker product, $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,k})$ is $1 \times k$ vector, the

$k \times m$ matrix $\beta = (\beta_1, \ldots, \beta_m) = (\beta_{v,j})$ of unknown parameters, vectorization operator *vec(A)* of matrix *A*, the *n*-dimensional diagonal matrix $diag(v)$ with the vector *v* on the main diagonal, $Z = (Z_i)$ is the $n \times 1$ vector, where $Z_i(t)$ is Brown motion scale disturbance (Z_i are independently, identically normally distributed with mean zero and constant variance σ^2).

It is seen that the linear regression model has place here. The expectation and the covariance matrix of *Y(T)* are the following:

$$E(Y(T)) = \begin{pmatrix} \vec{t}_1 \otimes x_1 \\ \vec{t}_2 \otimes x_{21} \\ \ldots \\ \vec{t}_n \otimes x_n \end{pmatrix} vec\,\beta, \qquad (2)$$
$$Cov(Y(T)) = \sigma^2 diag(t_1, t_2, \ldots, t_n).$$

After that it is possible to use the generalized least square method to estimate parameter matrix β, supposing that the matrix of unknown coefficients by $vec\,\beta$ has full rank *mk*.

Further it is supposed that the *external environment* is a random one and is described by a continuous-time Markov chain $J(t)$, $t \geq 0$, with the finite state set $S = \{1, 2, \ldots, m\}$, where $\lambda_{i,j}$ is the known transition rate from state s_i to state s_j, and $\Lambda_i = \sum_{j \neq i} \lambda_{i,j}$.

Further all necessary formulas for a calculation of the conditional average sojourn time that allows to get the needed estimates are provided in previous researches [1, 2].

The obtained model can be used in any applied area; however we narrow our research to the field of transportation. Let's see how regression models are used in transport modelling.

3 Regression Models in Transport Modelling

Regression analysis is a technique that can help to describe the relationship between the variables using the analytical instruments. After the opinion of Washington [12] and another authors, the linear regression is one of the most widely studied and used statistical and econometric methods. And there are a number of reasons. Firstly, the linear regression is suitable for modelling with different relationships between variables. In addition, linear regression model assumptions are often properly executed in many practical applications. The results of the regression model are relatively easy to interpret, regression models estimates are relatively easy, and software for model estimation is available in many non-specialized packages. For sure there are also many restrictions on using regression, but we aren't discussing them in this paper.

3.1 The Structure of the Classical Transport Model

Years of experimentation and development have resulted in the general structure, called the classical transport model. Model structure is the practical result of 1960's, and

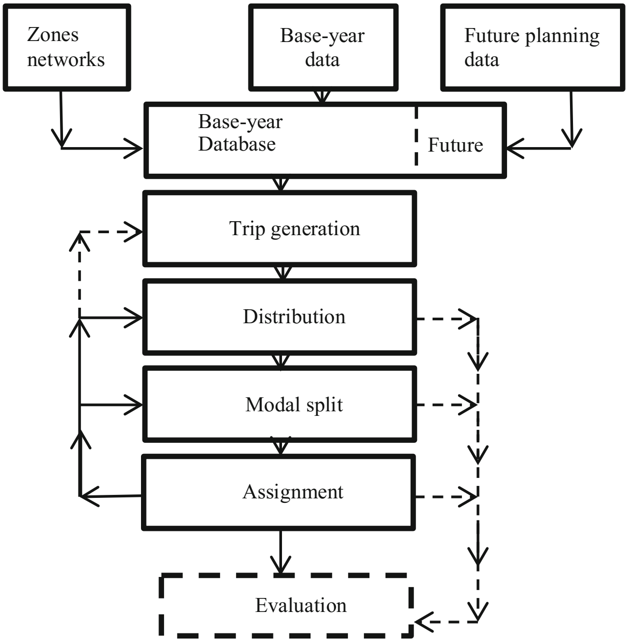


Fig. 1. The structure of the classical transport model.

despite significant improvements of modelling techniques, is still more or less effective from those times. Classical transport model structure is reflected in Fig. 1 [8].

Approach starts with the zone and network system review, and the collection and coding of planning, calibration and validation. These data are then used to estimate a model of the total number of trips generated and attracted by each zone of the study area (trip generation). In other words, determination of transport flow, their destination and origin, chose reason for the trip. The next step is the allocation of these trips to particular destinations, in other words their distribution over space, thus producing a trip matrix. The following stage normally involves modelling the choice of mode and this results in modal split, i.e. the allocation of trips in the matrix to different modes. Finally, the last stage in the classic model requires the assignment of the trips by each mode to their corresponding networks: typically private and public transport.

Different models are used in all four stages of the classical transport model.

Regression models are actively applied in two stages of the classical transport model: *trip generation* and *modal split*.

Regression Models in Trip Generation Modelling. The objective of classical transport model trip generation stage is to predict the total number trips, generated by the origin of each zone of the studied region (Origin O-i) and connected destination (Destination *Dj*). This can be achieved in different ways, not only with regression models.

There is *Zonal-based Multiple Regression* and *Household-based Regression.* In the first case, an attempt is made to find the ratio between the number of household trips, created by or connected to the zone, and average socioeconomic characteristics (for example, such as the level of income, the number of cars, household size and structure) in each zone. It is important to note that the zonal-based regression is conditioned with area environment and size (i.e. the spatial aggregation problem). This is clearly shown by the fact that the inter-zonal variation decreases with the zone size. In the second case, when the regression is based on households, each house is taken as the entry vector data with the purpose of including into the model all the observed variation range of households and their driving behaviour. The calibration process, as well as in the case of zonal-based models can be performed by matching, testing each independent variable until the best model is acquired (for a given confidence level in terms of common indicators).

Regression Models in Modal Split Modelling. Modal split is probably the most important stage in the classical model of transport planning. It is, because of the public transports' main role in policy design. Almost without exception, travel by public transport uses the road space more efficiently and causes less accidents and emissions than private car use.

Modal split models can be *aggregated*, if they are based on zonal (and inter-zonal) information. And also may be *disaggregated* models, which are based on households and/or individual data.

Work describes disaggregated models that are very widely and successfully used in the process of modal split. This is the model class called discrete choice models.

At the end of 1970's, Daniel McFadden, now a professor at the University of California (Berkeley), created and developed a statistical estimation methods, which promoted wide spread of the discrete choice models, and for this achievement Dr. McFadden in association with James Heckman received the Nobel Prize in economy in year 2000. Around the same time, Moshe Ben-Akiva at Massachusetts Institute of Technology submitted a doctoral thesis presented related to this topic. Some great names and their impressive work, which developed this theory in transport sector, are: Daniel McFadden [7], Moshe Ben-Akiva and Lerman [3], Chandra Bhat, Kenneth Train, David Hensher, Michael Berlarie, Charles Manski.

In general, discrete choice models postulate that: the probability of individuals choosing a given option is a function of their socioeconomic characteristics and the relative attractiveness of the option.

Nowadays the discrete choice models range is very wide: starting with simplest ones and finally with a very complex models. Let's look at some model classes.

- *Logit* models (logistic regression).

Logit models are the most popular discrete choice models. There are different logistic regression models such as:

- Multinomial *logit* models,
- Conditional *logit* models,
- Nested *logit* models,
- Cross-Nested *logit* models,

– Cross-Correlated *logit* models,
– Paired Combination *logit* models.

These models belong to the class of Generalized Extreme Value models formulated by Daniel McFadden in 1978. This class of models is characterized by the fact that residuals ε are independently and identically distributed as Gumbel distribution (or Weibull, or extreme value type I).

- *Probit* models (*probit* regression):

– Conditionally-heteroskedastic probit models.

- Robust choise models:

– Mixed *logit* models,
– Semiparametric single index discrete choice models,
– Nonparametric discrete choice models.

- *Tobit* models (censored regression).

Obviously, the number of models that can be used in transport mode choice modeling is very large. It is necessary to take into account the appropriate conditions during the model selection process.

4 Stochastic Markov Chain Based Models in Transport Modelling

Markov chains have been used over the years for statistical analyses of engineering, medical and environmental data. Markov chains have been used in transportation related fields including air traffic control, bridge deterioration rates, intersection queuing theory, pavement crack deterioration and pavement performance [11].

As well Markov chains are employed to model congestion and emissions in a manner analogous to how Google employs these tools to model congestion in the Internet [5]. According to the authors Crisostomi, Kirkland and Shorten in [13] "Markov chains offer considerable advantages over conventional road network simulators. They can be built from real data easily; they are fast and effective simulation tools. Also, Markov chains can be used to inform the design of control strategies that are suitable for regulating load in transportation networks; namely the design of load balancing strategies using infrastructure to shape the probabilities. Furthermore, Markov chain models allow users to glean structural information that is usually difficult to obtain using other modelling techniques. These include: identification of sensitive links in the network; identification how connected the network actually is (graph connectivity, sub-communities); the design of networks that are in some sense maximally mixing; and the ability to predict the effects of failure of a link (i.e. due to road works or an unexpected event). Such information cannot be extracted easily from conventional simulators (most sensitive road junctions, speed of mixing, identification of subgraphs, and degree of graph connectivity). As such they are excellent traffic engineering tools and provide a mechanism to respond to congestion conditions in near real time in a pre-emptive manner".

In this study we are interested in Markov chain application exactly in transport modelling, more precisely the place of Markov chains in classical transport model. Due to recent studies [10] Markov chains can be successfully implemented to the stage named distribution of classical transport model to form a trip matrix or OD-matrix (Origin-Destination). The method is based on the transportation network, which is associated with the graph of the corresponding Markov chain and on the canonical form of the graph. Previously, Li [6] considered a method to estimate OD matrices of the public transport using Markov approach.

Also there is a possibility to use Markov chains at Modal Split Modelling (3rd stage of classical transport model), as well as regression models are used. There are some investigations, for example [14], where a Markov chain based choice model is considered and showed that it provides a simultaneous approximation for all random utility based discrete choice models including the multinomial logit, the probit, the nested logit and mixtures of multinomial logit models. [14] showed that the choice probabilities computed by the Markov chain based model are a good approximation to the true choice probabilities for any random utility based choice model under mild conditions. Unfortunately, this example does not consider the application of the Markov chain to the transport area.

5 Markov-Modulated Linear Regression: Tasks and Challenges

In general we can conclude that regression models and Markov chain based models are not competitors, they are used at different stages within the frame of transport modelling. They can rather complement each other as we can see in the proposed Markov-modulated regression model. Let's briefly summarize advantages of both model types.

Some of the general advantages of regression analysis techniques are:

- Simplicity of computation,
- Software for model estimation is available in many non-specialized packages,
- The regression model results are relatively easy to interpret,
- Future trends can be predicted based on historical data,
- It provides a statistical platform for more advanced modelling,
- It has the ability to deal with multivariate components.

The advantages of Markov chain modelling can be considered asef both types n randomacan be mentioned as disadvantage of model appliance:

- Allows to develop a prediction model with just two years of data;
- Allows to make calculations even if data for some years is missing (special approach is needed for regression analysis);
- Allows users to glean structural information;
- Has the ability to treat data with stochastic tools including Bayesian processes.

Markov-Modulated linear regression model is still advanced regression model, so it can be applied in any subject area as well as the usual regression models where they are reasonable. And as advanced model Markov-Modulated linear regression model can bring new vision of system or can help to obtain new information. For example, the total number of trips in a zone may depend not only on the average socioeconomic characteristics but also on the weather conditions that can be included in the model not as an additional independent variable (what complicates the model), but to consider them as external environmental conditions (sunny or rainy weather, for example), thus obtaining various regression parameters at different states of the external environment.

Previous researches were executed on artificial data [1]. Simulation modelling included two stages: a small data sample and a large data set. In the first case, the Markov-modulated linear regression gave unacceptable estimates of the true parameter β, but dependent variable values were predicted well. This was explained by the fact that the available observation number was insufficient with the chosen number of parameters to be evaluated. A large sample yielded much better results and it was concluded that the assessments tend to be true parameter values, but rather slow. This can be improved if the evaluation procedure uses the true variance values for sojourn times. Of course it can be argued that in actual practice we rarely are in asymptotic situation. But nowadays, when specialists are confronted directly with big data problems, when data in huge quantities come from different mobile applications, or are collected by GPS assistance, the Markov-modulated linear regression can provide a significant contribution to the different transport problems' solving. This gives reason to believe that it can be applied to various challenges in the transport sector and other practical life areas.

6 Conclusions

Markov-Modulated linear regression brings the idea that the regression model parameters do not remain constant throughout the period of model viewing, but vary randomly with the external environment, the impact of which is described by a Markov chain with continuous time and final state set.

In the framework of the classical transport model regression models are actively applied in two stages: *trip generation* and *modal split*, in their turn Markov chains are successfully implemented to the distribution stage to form a trip matrix or OD-matrix and *modal split* stage as well.

In general regression models and Markov chain based models are not competitors, they are used at different stages within the frame of transport modelling. They can rather complement each other as we can see in the proposed Markov-modulated regression model.

Markov-Modulated linear regression model is still advanced regression model, so it can be applied in any subject area as well as the usual regression models where they are reasonable. And as advanced model Markov-Modulated linear regression model can bring new vision of system or can help to obtain new information. The lack of model approbation on real data can be mentioned as disadvantage of the application of Markov-Modulated linear regression model. Also a shortage of software is also a temporary disadvantage of using this type of model. But both shortcomings are temporary and quite solvable problems.

References

1. Andronov, A., Spiridovska, N.: Markov-modulated linear regression. In: Proceedings of the International Conference on Statistical Models and Methods for Reliability and Survival Analysis and Their Validation (S2MRSA), Bordeaux, France, pp. 24–28 (2012)
2. Andronov, A.M., Spiridovska, N., Santalova, D.: Markov-modulated linear regression for incomplete observations. In: Abstracts of Communications on 11th Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius, p. 72 (2014)
3. Ben-Akiva, M.E., Lerman, S.R.: Discrete Choice Analysis. Theory and Application to Travel Demand. MIT Press, Cambridge (1985)
4. Hensher, D.A., Button, K.J.: Handbook of Transport Modelling (Handbooks in Transport), 2nd edn. Emerald Group Publishing Limited, Bingley (2008)
5. Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond – The Science of Search Engine Rankings. Princeton University Press, Princeton (2006)
6. Li, B.: Markov models for Bayesian analysis about transit route origin-destination matrices. Transp. Res. Part B Methodological **43**(3), 301–310 (2009)
7. McFadden, D.: Econometric Models of Probabilistic Choice. The MIT Press, Cambridge (1981)
8. Ortuzar, J., Willumsen, L.: Modelling Transport, 4th edn. Wiley (2011)
9. Sivakumar, A.: Modelling Transport: A Synthesis of Transport Modelling Methodologies. Imperial College, London (2007)
10. Tesselkin, A., Khabarov, V.: Estimation of origin-destination matrices based on Markov chains. Procedia Eng. **178**, 107–116 (2017)
11. Uchwat, C., MacLeod, D.: Case studies of regression and Markov chain models. In: Proceedings of the Conference of the Transportation Association of Canada, Fredericton, New Brunswick (2012)
12. Washington, S.P., Karlaftis, M.G., Mannering, F.L.: Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC, Boca Raton (2003)
13. Crisostomi, E., Kirkland, S., Shorten, R.: A Google-like Model of Road Network Dynamics and its Application to Regulation and Control (2010). Hamilton Institute's Homepage. http://www.hamilton.ie/shravan/SpectralRoad_Report.pdf
14. Columbia University's Homepage. http://www.columbia.edu/~vg2277/MC_paper.pdf. https://doi.org/10.1287/opre.2016.1505
15. Pacheco, A., Tang, L.C., Prabhu, N.U.: Markov-Modulated Processes and Semiregenerative Phenomena. World Scientific Publishing Company, Hackensack (2009)

Spatiotemporal Big Data Challenges for Traffic Flow Analysis

Dmitry Pavlyuk(✉)

Transport and Telecommunication Institute, Riga, Latvia
Dmitry.Pavlyuk@tsi.lv

Abstract. This paper contains a survey of spatiotemporal big data challenges in the area of urban traffic flow analysis. Existing sources and types of traffic flow data were reviewed and evidences that traffic flow data can be considered as spatiotemporal big data were provided. Current trends in spatiotemporal big data analytics and in urban traffic flow modelling and forecasting were consolidated and a list of joint emerging challenges was composed. The stated challenges cover different spatiotemporal aspects of big data and are linked to optimal time and space data resolution, spatial and temporal relationships in traffic data, computational complexity of spatiotemporal algorithms, fusion of traffic data from heterogeneous data sources into a single predictive scheme, and development of responsive streaming algorithms. The raised challenges are supported by an extensive literature review, and suggestions for future work are offered.

Keywords: Spatiotemporal big data · Urban traffic flows · Modelling Forecasting

1 Introduction

Nowadays a wide network of distributed sensors and tracking devices provide overwhelming amount of geospatial traffic data with high velocity. Modern urban traffic data include high-resolution values from hundreds of traffic sensor loops, installed at different space points; thousands GPS car routes from navigation software and probe cars; millions of traffic-related short messages in social networks. As a result, traffic data volume and dimensions grow rapidly and datasets become heterogeneous, which make application of conventional methods difficult and inefficient [1]. Recently developed methods of big data processing also weakly appropriate for spatiotemporal data as most of them don't support even basic spatial properties and relationships.

Classical big data is associated with features of volume, velocity, variety, variability, veracity, visualization, and value. In addition to these 7V's, spatiotemporal big data includes a myriad of potential relationships between observations in spatial and temporal dimensions. Inputs of spatiotemporal big data inputs are more complicated than classical ones, because they include discrete representations of continuous space and time.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_22

Spatiotemporal relationships and continuous-to-discrete conversion raise new challenges for researchers in visualisation, modelling, and forecasting of spatiotemporal big data.

Due to a highly interrelated structure, spatiotemporal data processing poses statistical and computational challenges [2]. Mining information from spatiotemporal datasets and estimating of statistical model parameters is more difficult than solution of corresponding problems for traditional data due to the complexity of spatiotemporal data types and relationships [3]. The problem of efficient spatiotemporal big data processing and utilisation for modelling become even more important in combination with real-time requirements, peculiar to traffic analysis [4]. Automated traffic lights control, variable speed limits and other methods of active traffic management require results of data processing in real-time, and weakly developed background of spatiotemporal big data processing methods creates obvious obstacles for their effective application.

In this paper challenges of spatiotemporal big data processing are reviewed, focusing on their application to real-world traffic flow datasets. Most important methodological shortages of spatiotemporal big data analytical methodology were identified and can be considered as directions for intensive interests of the scientific community. Recently Vlahogianni et al. [5] published a list of most important challenges in traffic flow modelling and short-term forecasting, and Jiang and Shekhar [6] presented their vision for spatiotemporal big data challenges. In this paper these lists of challenges were proved to be highly related and thus many modern traffic flow modelling tasks have the spatiotemporal big data nature.

2 Spatiotemporal Big Data of Traffic Flows

Urban traffic flow data is naturally spatiotemporal big data. Nowadays enormous volume of traffic data is collected in real-time from various data source, and almost all observations has spatial and temporal labels. Spatiotemporal data can be classified into three types: raster, vector and graph data [6]. Raster spatiotemporal data includes temporal series of raster images (video, geo-images, etc.); vector data includes trajectory polylines in time (time series) or both in time and space (routes); graph data includes information about dynamics of network structures.

The problem of new traffic data sources is emerging: Yatskiv et al. [7] presented a comprehensive review of modern traffic flow data sources; recently a special workshop on new traffic data sources was organised [8]. An extended list of traffic data sources is summarised in Table 1 and classified by spatiotemporal data types.

Recently several researchers utilised data source of different types for model estimating and forecasting [9, 10], and fusing data of various types into a single predictive scheme becomes an important direction for academic studies.

Note that spatiotemporal nature of traffic flow data should lead to a significant concordance of spatiotemporal methodological developments and traffic flow modelling and forecasting challenges. In the following section a list of emerging problems of traffic flow forecasting is composed and their links to spatiotemporal big data challenges are provided.

Table 1. Spatiotemporal traffic flow data sources.

Data type	Spatial dimension	Temporal dimension	Traffic flow data source
Raster	Territory or road segment	Time series of raster images	• Wide area motion imagery (WAMI): geo-images and video records of traffic flows, i.e. from unmanned aerial vehicle (UAV) cameras • Traffic cameras • Car digital video recorders • Autonomous cars' light/laser imaging detection and ranging (LIDAR) sensors
Vector	Spatial point	Trajectories in time	• Sensor loops (video, infrared, ultrasonic, microwave, acoustic, piezoelectric, photoelectric, inductive, magnetic, RFID) • Human (counters)
	Spatial trajectory	Trajectories in space and time	• GPS-enabled road mapping devices and mobile phones • Cellular-based mobile phones • Probe cars • Bluetooth car trackers
Graph	Road network	Dynamics of a road network structure	• Road network structure data (OpenStreetMaps, etc.) • Volunteered geographic information from drivers' software: accidents, traffic jams, danger conditions, etc. • Social media data (Twitter twits, Facebook comments, etc.)

3 Spatiotemporal Big Data Challenges of Traffic Modelling

Traditional big data challenges are typical to traffic flow analysis: many researchers paid their attention to core V's of traffic data [11–13]. This paper focused on additional challenges, related to the spatiotemporal nature of traffic data. Each stated challenge can be considered as a wide direction for future research.

3.1 Challenge 1: Determining Time and Spatial Data Resolution

A spatiotemporal process of traffic flow is continuous in space and time, but the majority of big data analytics collection and modelling tools deal with them as discrete. This continuous-to-discrete conversion can be implemented both at the layer of data collection (i.e. information from sensor loops could be available in a form of average values) and at the modelling layer (data could be aggregated to improve model's forecasting accuracy). Nowadays raw traffic flow data is available at very high temporal and spatial resolutions (i.e. 5 s time frames for every crossroad in a road network segment), and selection of an appropriate time and space resolution of data becomes an important researcher's decision. Both spatial and temporal aggregation of traffic flow data are not straightforward and related to multiple issues.

A list of recent urban traffic studies, where attention to spatial aggregation is paid is quite long [14, 15]. Spatial aggregation is frequently referred in literature as the modifiable areal unit problem (MAUP) [16]. MAUP present the fact that different selection of spatial regions could have a significant impact on the modelling results. Traffic flow data is very flexible for spatial aggregation: data can be grouped by road lane (for multi-lane roads), by road segments of different lengths, by crossroads and different directions, by residential quarters, etc., and there is no straightforward methodology for selection of an appropriate aggregation levels.

Although temporal aggregation of traffic data is addressed in studies more often [17–19], a lack of a general approach or methodology also should be noted. General conclusions of existing researches support the hypothesis about a critical role of temporal aggregation for properties of traffic flow time series like stationarity, homogeneity, and volatility, and as a result for data predictability. Also most of authors mentioned that further research is needed to determine the optimum aggregation level with respect to different traffic modelling techniques.

Finally, to the best of our knowledge, there are no empirical researches, where spatial and temporal aggregation are analysed simultaneously. A strict methodology, which is recommended for selection of an appropriate temporal and spatial aggregation level for a modelling technique, is called.

3.2 Challenge 2: Discovering Spatial and Temporal Relationships

Spatial and temporal links are natural for traffic flows and should be addressed in modelling applications. Mostly researchers consider physical properties of the flow for modelling, i.e. estimate relationships (lagged in time) between consecutive space points on freeways. But the problem is more complex: different segment of road network can be considered as complementary one to each other and traffic flow could be relocated under specific conditions (congestion, traffic jams, etc.). Thus modelling of spatial relationships between space points within a small road segment (i.e. one freeway) could be inaccurate and big data of the whole road network should be utilized.

There are two general approaches to identification of spatial links: road-network-based and correlation-based. Within the network-based approaches a physical structure of the road network is utilized for identification of potential relationships. Usually this structure is introduced in a form of spatial contiguity matrix, which is a necessary exogenous component of many spatial and space-time models. Popular spatial autoregressive models (SAR) and space-time autoregressive integrated moving average (STARIMA), based on fixed spatial contiguity matrixes, have been recently applied for traffic flow forecasting [20–22]. The formal specification of STARIMA model is [3]:

$$Y_{i,t} = \sum_{k=1}^{p} \sum_{s=1}^{\lambda_k} \alpha_{ks} W_s Y_{i,t-k} + \sum_{k=1}^{q} \sum_{s=1}^{m_k} \gamma_{ks} W_s \varepsilon_{i,t-k} + \varepsilon_{i,t}, \quad (1)$$

where p is the autoregressive order, q is the moving average order, λ_k is the spatial order of the kth autoregressive term, m_k is the spatial order of the kth moving average term,

α_{ks} and γ_{ks} are parameters to be estimated and W_s is the matrix for spatial order s and $\varepsilon_{i,t}$ is the random disturbance vector at time t.

The assumption about a fixed spatial contiguity matrix W_s is rarely satisfied in practice, so several recent researches applied correlation-based analysis to dynamic identification of spatial relationships [23–26]. Correlation analysis in conjunction with regime-switching techniques allows discovering of time-varying spatial relationships between traffic flows at remote road segments, observed under specific road conditions. Although correlation-based approaches are more flexible, they have their own drawbacks. In particular, big data analysis allows identification of correlation links between remote space points without a reasonable background for causal spatial relationships. This high correlation across spatial time series at a long distance are referred as teleconnections [6], but also could be just spurious. Modern least absolute shrinkage and selection operator (LASSO) techniques, rarely applied for traffic modelling [27–29], allow limit spatial links with the most significant ones. Unexplained teleconnections in spatiotemporal data could be useful for predictive techniques, but identification of "useful" spatial links is still challenging.

Recent trends in big data analysis in general and specifically in traffic flow forecasting lay in shifting from correlational relationships to causal ones. Consistent methods of causal spatiotemporal relationships identifications are an emerging area of traffic flow analysis [12, 30].

3.3 Challenge 3: Dealing with Computational Complexity

Growing volume, velocity, and variety of spatial datasets exceed the capacity of commonly used spatial modelling techniques and spatial database technologies to handle the data with reasonable efforts. Estimation of classical spatial models' parameters is usually related to intensive matrix algebra (for example, estimation of all spatial autoregressive models requires multiple calculations of the spatial log-determinant $log|I - \alpha_{ks} W_s|$), which becomes challenging for high-dimensional spatial data. Presence of multiple spatial and temporal relationships, described in Challenge 2, also highly increases computational complexity of traffic modelling techniques.

There are two main directions of development in this area: algorithmic and computational. Algorithmic advances are related to computationally effective methods of matrix algebra (sparse spatial matrix handling, effective matrix decomposition and tridiagonalization, etc.) and new parallel and approximate parameter estimation techniques [31]. To the best of our knowledge, such techniques are not adapted for spatiotemporal model parameter estimation.

Computational advances are related to extending of big data mining techniques to take advantages of parallel processing [32–34]. Currently such intensive calculations are executed using distributed computing techniques like Hadoop framework. The new techniques of data warehousing (i.e. Apache Hive), querying (i.e. Pig and Pigeon) and mining (i.e. Apache Mahout) of distributed file systems are rapidly developing last years. Mahout on Spatial-Hadoop [35] and several other similar solutions extend the parallel processing to spatial data. Mahout offers several important data mining functionalities, but its features are still very limited for real-world urban traffic spatiotemporal big data analysis.

3.4 Challenge 4: Developing Responsive Streaming Algorithms

Nowadays the challenging task of traffic flow big data analysis is not only to produce an optimal solution, but to do this in a timely manner. A traditional statistical approach, which is based on historical data collecting, model parameters' estimating and further model-based forecasting, doesn't match requirements for timely and accurate information in a highly dynamic transportation environment. Nowadays there are only a few cities in the world where streaming traffic data is provided to central servers that can process the data in real-time, but the rapid growth of hardware and telecommunication technologies will lead this to an everyday reality soon. New responsive algorithms, which use spatiotemporal data in a streaming mode and update a predictive scheme, are highly required. Development of these new algorithms is challenging due to a relatively weak methodological background of data streams [36] and complexity of modern traffic flow predictive schemes.

Streaming algorithms of model parameters estimation is still in the emerging state and there are no such algorithms for multivariate autoregressive models (like STARIMA), widely utilised for traffic flow forecasting. A list of studies, where spatiotemporal traffic data is handled in a streaming manner, is very limited [37].

From another side, modern traffic flow predictive schemes are rarely based on a single statistical model, but includes a set of candidates models and an algorithm that switch between them or combine different forecasts. The positive effects of combining forecasts have been discussed in several papers [38, 39], and a methodology for a responsive model selection on the base of streaming spatiotemporal data may enhance the decision making process in case of dynamic traffic flow conditions.

3.5 Challenge 5: Fusing Data from Various Data Sources

In Sect. 2 different types of spatiotemporal traffic data, used for modelling purposes, were presented. Fusing data from different data sources into a single predictive scheme is often found to be useful for forecasting accuracy and consistency. Multiple data sources may provide complementary data, and data fusion can produce a better understanding of the observed traffic flow process [40]. Nowadays this is not enough to estimate model parameters on the base of a predefined traffic data type, this is also necessary to raise the question – how the model should be updated, using date from other sources. For example, STARIMA models, mentioned above, are traditionally estimated on the base of data from several sensor loops (temporal trajectories at fixed spatial points). Updating of the model parameters on the base of probe cars' tracks (trajectories in space and time) or raster data (the current traffic state, obtained from UAV photo) seems very promising and is perceived as an answer to growing heterogeneity of traffic flow data.

4 Conclusions

This paper contributes to emerging usage of spatiotemporal big data for urban traffic flow analysis. Nowadays traffic data is obtained from various data sources, has heterogeneous structure (raster, vector, and graph in spatial and temporal dimensions) and naturally spatiotemporal and big. Current trends in development of the spatiotemporal big data analytics and in urban traffic flow analysis were consolidated and a list of joint emerging challenges is composed:

- Determining time and spatial data resolution,
- Discovering spatial and temporal relationships,
- Dealing with computational complexity,
- Developing responsive streaming algorithms,
- Fusing data from various data sources.

The stated challenges and current approaches to their resolving are supported by an expensive literature review and directions for future researches are offered.

Acknowledgements. This work was financially supported by the post-doctoral research aid programme of the Republic of Latvia (project no. 1.1.1.2/VIAA/1/16/112, "Spatiotemporal urban traffic modelling using big data"), funded by the European Regional Development Fund.

References

1. Vlahogianni, E.I., Park, B.B., van Lint, J.W.C.: Big data in transportation and traffic engineering. Transp. Res. Part C Emerg. Technol. **58**, 161 (2015)
2. Shekhar, S., Jiang, Z., Ali, R., Eftelioglu, E., Tang, X., Gunturi, V., Zhou, X.: Spatiotemporal data mining: a computational perspective. ISPRS Int. J. Geoinf. **4**, 2306–2338 (2015)
3. Cressie, N.A.C., Wikle, C.K.: Statistics for Spatio-Temporal Data. Wiley, Hoboken (2011)
4. Xu, J., Deng, D., Demiryurek, U., Shahabi, C., van der Schaar, M.: Mining the situation: spatiotemporal traffic prediction with big data. IEEE J. Sel. Topics Signal Process. **9**, 702–715 (2015)
5. Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C.: Short-term traffic forecasting: where we are and where we're going. Transp. Res. Part C Emerg. Technol. **43**, 3–19 (2014)
6. Jiang, Z., Shekhar, S.: Spatial Big Data Science: Classification Techniques for Earth Observation Imagery. Springer, Cham (2017)
7. Yatskiv, I., Grakovski, A., Yurshevich, E.: An overview of different methods available to observe traffic flows using new technologies. In: The NTTS 2013 Proceedings, p. 10, Brussel, Belgium (2013)
8. Bureau of Infrastructure, Transport and Regional Economics: GHD report: New traffic data sources. In: presented at the New Data Sources for Transport Workshop, Sydney, Australia (2014)
9. Wang, S., He, L., Stenneth, L., Yu, P.S., Li, Z., Huang, Z.: Estimating urban traffic congestions with multi-sourced data. In: presented at the 17th IEEE International Conference on Mobile Data Management (MDM) (2016)

10. Bhaskar, A., Chung, E., Dumont, A.-G.: Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks: fusing loop detector and probe vehicle data. Comput. Aided Civil Infrastruct. Eng. **26**, 433–450 (2011)
11. Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.-Y.: Traffic flow prediction with big data: a deep learning approach. IEEE Trans. Intell. Transp. Syst. **16**, 1–9 (2014)
12. Vlahogianni, E.I.: Computational intelligence and optimization for transportation big data: challenges and opportunities. In: Lagaros, N.D., Papadrakakis, M. (eds.) Engineering and Applied Sciences Optimization, pp. 107–128. Springer, Cham (2015)
13. He, Z., Lv, Y., Lu, L., Guan, W.: Constructing spatiotemporal speed contour diagrams: using rectangular or non-rectangular parallelogram cells? Transportmetrica B Transp. Dyn. 1–17 (2017). http://www.tandfonline.com/doi/full/10.1080/21680566.2017.1320774
14. Kuhn, K., Nicholson, A.: Traffic flow forecasting and spatial data aggregation. Transp. Res. Rec. J. Transp. Res. Board **2260**, 16–23 (2011)
15. Loidl, M., Wallentin, G., Cyganski, R., Graser, A., Scholz, J., Haslauer, E.: GIS and transport modeling-strengthening the spatial perspective. ISPRS Int. J. GeoInf. **5**, 84 (2016)
16. Wong, D.: The modifiable areal unit problem (MAUP). In: Fotheringham, A.S., Rogerson, P. (eds.) The SAGE Handbook of Spatial Analysis, pp. 105–124. SAGE Publications, Los Angeles (2009)
17. Qiao, F., Yu, L., Wang, X.: Double-sided determination of aggregation level for intelligent transportation system data. Transp. Res. Rec. J. Transp. Res. Board **1879**, 80–88 (2004)
18. Oh, C., Ritchie, S., Oh, J.-S.: Exploring the relationship between data aggregation and predictability to provide better predictive traffic information. Transp. Res. Rec. J. Transp. Res. Board **1935**, 28–36 (2005)
19. Vlahogianni, E., Karlaftis, M.: Temporal aggregation in traffic data: implications for statistical characteristics and model choice. Transp. Lett. **3**, 37–49 (2011)
20. Kamarianakis, Y., Prastacos, P.: Space-time modeling of traffic flow. Comput. Geosci. **31**, 119–133 (2005)
21. Min, X., Hu, J., Chen, Q., Zhang, T., Zhang, Y.: Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. In: 2009 12th International IEEE Conference on Intelligent Transportation Systems, pp. 1–6 (2009)
22. Salamanis, A., Kehagias, D.D., Filelis-Papadopoulos, C.K., Tzovaras, D., Gravvanis, G.A.: Managing spatial graph dependencies in large volumes of traffic data for travel-time prediction. IEEE Trans. Intell. Transp. Syst. **17**, 1678–1687 (2016)
23. Stathopoulos, A., Karlaftis, M.G.: A multivariate state space approach for urban traffic flow modeling and prediction. Transp. Res. Part C Emerg. Technol. **11**, 121–135 (2003)
24. Cheng, T., Haworth, J., Wang, J.: Spatio-temporal autocorrelation of road network data. J. Geogr. Syst. **14**, 389–413 (2012)
25. Min, W., Wynter, L.: Real-time road traffic prediction with spatio-temporal correlations. Transp. Res. Part C Emerg. Technol. **19**, 606–616 (2011)
26. Schimbinschi, F., Moreira-Matias, L., Nguyen, V.X., Bailey, J.: Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. Expert Syst. Appl. **82**, 301–316 (2017)
27. Kamarianakis, Y., Shen, W., Wynter, L.: Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. Appl. Stoch. Models Bus. Ind. **28**, 297–315 (2012)
28. Haworth, J., Cheng, T.: Graphical LASSO for local spatio-temporal neighbourhood selection. In: Proceedings the GIS Research UK 22nd Annual Conference, presented at the GISRUK, pp. 425–433 (2014)

29. Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., Li, Y.: Robust causal dependence mining in big data network and its application to traffic flow predictions. Transp. Res. Part C Emerg. Technol. **58**, 292–307 (2015)
30. Ermagun, A., Levinson, D.M.: Spatiotemporal traffic forecasting: review and proposed directions. In: presented at the 96th Annual Transportation Research Board Meeting, USA (2016)
31. Kazar, B.M., Celik, M.: Spatial AutoRegression (SAR) Model. Springer, New York (2012)
32. Lee, J.-G., Kang, M.: Geospatial big data: challenges and opportunities. Big Data Res. **2**, 74–81 (2015)
33. Eldawy, A., Mokbel, M.F.: The era of big spatial data: a survey. Inf. Media Technol. **10**, 305–316 (2015)
34. Alkathiri, M., Abdul, J., Potdar, M.B.: Geo-spatial big data mining techniques. Int. J. Comput. Appl. **135**, 28–36 (2016)
35. Eldawy, A., Mokbel, M.F.: SpatialHadoop: a MapReduce framework for spatial data. In: presented at the 2015 IEEE 31st International Conference on Data Engineering (ICDE) (2015)
36. Aggarwal, C.C. (ed.): Data Streams. Springer, Boston (2007)
37. Herring, R.J.: Real-time traffic modeling and estimation with streaming probe data using machine learning (2010)
38. Tan, M.-C., Wong, S.C., Xu, J.-M., Guan, Z.-R., Zhang, P.: An aggregation approach to short-term traffic flow prediction. IEEE Trans. Intell. Transp. Syst. **10**, 60–69 (2009)
39. Stathopoulos, A., Karlaftis, M., Dimitriou, L.: Fuzzy rule-based system approach to combining traffic count forecasts. Transp. Res. Rec. J. Transp. Res. Board **2183**, 120–128 (2010)
40. Faouzi, N.-E.E., Leung, H., Kurian, A.: Data fusion in intelligent transportation systems: progress and challenges – a survey. Inf. Fusion **12**, 4–10 (2011)

Wide-Scale Transport Network Microscopic Simulation Using Dynamic Assignment Approach

Mihails Savrasovs(✉), Irina Pticina, and Valery Zemlyanikin

Transport and Telecommunication Institute, Lomonosov Street 1, Riga, Latvia
{savrasovs.m,Pticina.i}@tsi.lv, Valery.zemlynikin@gmail.com

Abstract. Traffic microscopic simulation is a powerful decision supporting tool, which could be applied for wide range of tasks. In a past microscopic traffic simulation was used to test local changes in transport infrastructure, but the growth of computers performance allows now to simulate wide-scale fragments of the traffic network, but this require usage of more advanced simulation techniques, like dynamic traffic assignment (DTA). There are several benefits of DTA application, but practical application of DTA make the simulation task more complex and require additional parametrisation, calibration and validation of the model. The goal of the current paper is to present the case-study from city of Riga (capital of Latvia), which demonstrates the application of dynamic assignment approach in the complex network simulation with use of DTA in PTV VISSIM simulation software. Additional finding of this paper is the set of the recommendations of the VISSIM dynamic assignment module configuration for solving above mentioned problem.

Keywords: Simulation · Dynamic assignment · Wide-scale networks · Traffic

1 Introduction

Traffic simulation is a powerful tool to forecast the impact on the traffic flow circulation in the selected area in case of introduction changes in the transport network. In a past microscopic traffic simulation was used to test local changes in transport infrastructure, but the growth of computers performance allows now to simulate wide-scale fragments of the traffic network and to estimate the impact of new objects into traffic flows and the surrounding infrastructure. But this leads to the problem of how to simulate such complex networks. One of the feasible solutions is to apply dynamic traffic assignment (DTA) approach in the simulation. The DTA is the approach, which is used to define the paths for the traffic flows throw the network. But use of DTA puts additional tasks in phase of model development and application: (1) DTA require the use of an origin-destination matrix (OD matrix), it means, that OD matrix should be estimated before; (2) the calibration process of the DTA is a complex, time-consuming and intuitive process, but to get a valid model for further use, this process should be completed.

The first of the above-mentioned tasks for the case-study regarding OD matrix evaluation was presented and described in detail in the earlier publication [1]. The problem

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_23

of OD matrix evaluation was solved based on using video recording data with manual transcription of the license plates numbers and referencing them to the origin and destination zone in a study area. The origin destination zones defined for this case study are presented in Fig by the circles. The referencing was done based on following steps: (1) The recorded video is decoded by means of vehicle type and registration number at license plate; (2) Next the data were processed by searching exit point for each entering vehicle (by license plate number); (3) The initial versions of origin-destination matrix were analysed and validated in manual mode; (4) Next, the calibration of the matrices was performed by application of TFlowFuzzy method; (5) After performing calibration of the matrix, the calibration results were validated based on NAIVE approach.

The goal of the current paper is to presents the case-study from the city of Riga (capital of Latvia), which demonstrates the application of dynamic routing approach in the complex network simulation and to provide recommendation based on case-study implementation with use of DTA.

2 Simulation Object of the Case-Study

The object is a fragment of the Riga transport network located in the eastern part of the city on the right side of the Daugava River. The study object includes a number of signalised and not signalised crossroads and three-two level flyovers. The territory is a part of Riga transport system, which connects several residential districts of Riga city: Purvciems, Kengarags, Centre and the left riverside. Additionally, the study area includes parts of the following main streets: Krasta, Maskavas Lubanas, Katlakalna, Krustpils, Darzciema, Piedrujas, Ilukstes and A. Saharova street. Maskavas and Krasta streets are highly used by the traffic from the city centre to neighbouring cities. The study object is represented in Fig. 1.

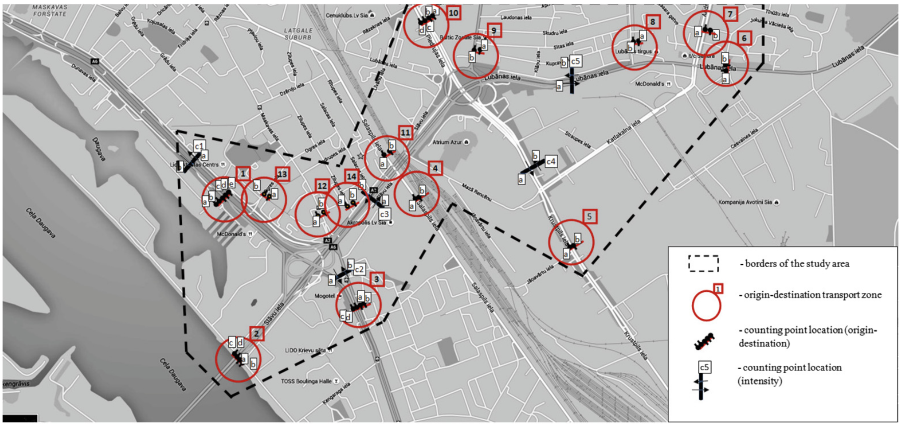


Fig. 1. Study object (background from Google Maps).

2.1 Current State of the Network

The study area is congested in morning (7:30–9:00) and evening peak hours (17:00–18:30). The Google Maps service provides the following information about the traffic in the area (see Figs. 2 and 3). Moreover, this area of the transport network is complex and treated as one of the “black spots” in Riga (high number of traffic accidents). From the Figs. 2 and 3, it could be seen that congestion level in the evening is much higher compared to morning peak hours, so it was decided to simulate only the evening peak hours in this research.

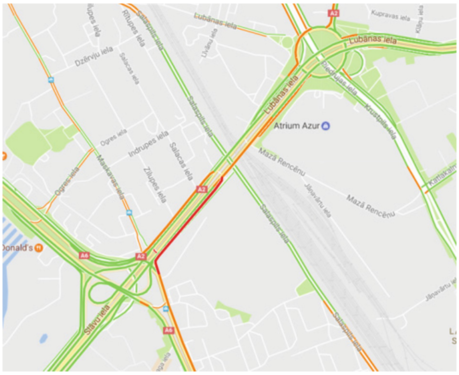


Fig. 2. Congestion during morning peak hour (data from Google Maps).

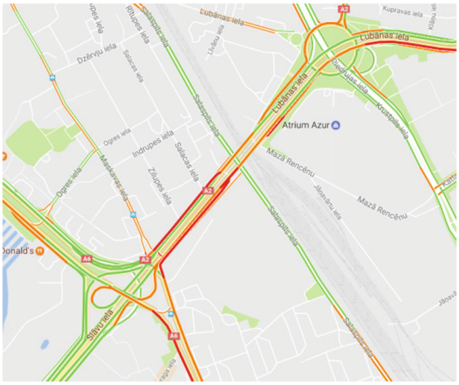


Fig. 3. Congestion during evening peak hour (data from Google Maps).

2.2 Input Data for Simulation

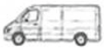
To complete the simulation of the study object, it is necessary to have the preliminary data for the model construction. The following data are required for the microscopic simulation: Geometrical data; Controlling devices and signs data; Demand data; Validation data; Future demand data; Future geometrical and controlling signs data. As current paper is targeted only on initial model development, the last two datasets (future demand data and future geometrical and controlling signs data) are not presented here.

- **Geometrical data.** To implement the transport network, the images from Google Maps application were used as background. The quality of the images was recognised as satisfactory for the study goal.
- **Controlling devices and signs data.** The data for defining the traffic light operation were obtained from Riga City Traffic Department. In total, there are 19 signalling control points in the transport network and up to 38 none-signalized intersections.
- **Demand data.** The demand was obtained in form of the OD matrix, which includes the travel patterns for 14 zones indicated in the study area. The OD matrix was obtained based on a license plate survey [1]. The OD matrix represents travel patterns for 1 h 15 min and includes the information about following types of vehicles (see Table 1).

The study area has several routes of public transport. There are 16 public transport lines with schedules. As the source of information, the data, of the "Rigas Satiksme" Ltd. were used.

- **Validation data.** The data for validation were obtained during the traffic counting, described in [1] and consist of traffic volume data for 5 counting points (see Fig. 1, black marks).

Table 1. Vehicles types.

Graphical representation	Code	Description
	V	Passenger vehicles
	C1	Light cargo vehicles
	C2	Midcargo vehicles
	C3	Cargo vehicles
	C4	Cargo vehicles with trailer
	S	Buses (without public transport of Riga city)

3 Initial Model Development

To develop a simulation model of the study object, the PTV VISSIM software was selected as the main simulation tool. The model development methodology in PTV VISION VISSIM simulation software could be presented as a sequence of the following steps: Defining background; Description of the vehicles types; Traffic flows composition; Network modelling; Traffic flow input data modelling; Routing decisions and routes definition; Public transport definition; None-signalized intersection definition; Signalized intersection definition; Output data definition; 2D/3D model development.

3.1 Definition of Model Supply Data

Based on the background images the transport network was implemented in VISSIM simulation software using standard objects: links and connectors. In total, more than 670 objects were used to model the transport network in the software. Figure 4 represents the coded network. Additionally, 33 public transport stops and 16 public transport lines were described in the model. In current model 19 signal controllers were described and 129 signal heads were inserted in the model to implement traffic lights infrastructure. To describe the traffic flows behaviour on un-signalized intersections, the Conflict area zones object was used. In total 38 objects are applied to define the priorities for traffic movement on crossroads (signalised and none-signalized intersection).

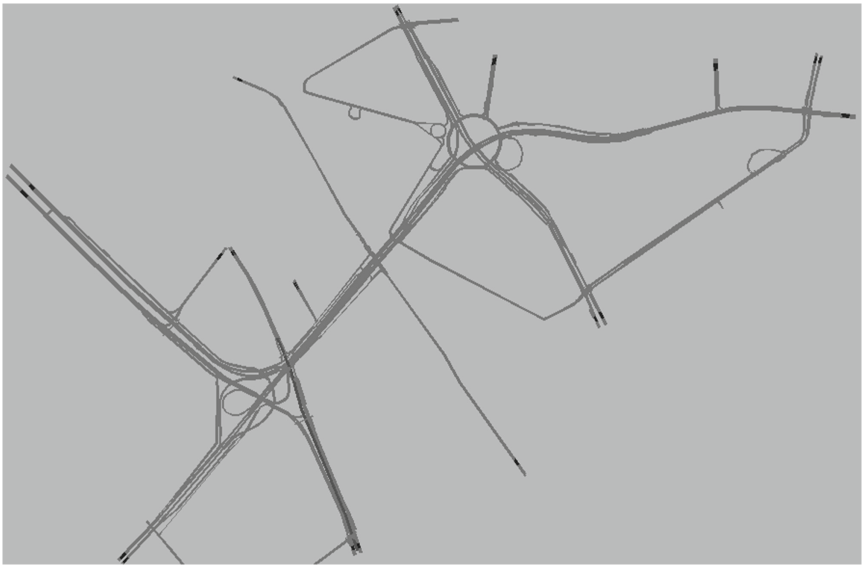


Fig. 4. Coded network.

3.2 Definition of Model Demand Data

The demand of the model was described in a form of the OD matrix for each type of the vehicles (presented in Sect. 2.2). Additionally, 28 parking lots were used in the model to represents transport zones. In total 14 zones were included in the model. Each zone is described by 2 parking lots (entering and leaving parking lot). The described OD matrices were defined for 1 h 15 min, to have 15 min as a warm-up period during simulation. A warm-up period was estimated as the doubled time for vehicles to travel through the network.

4 Dynamic Assignment Application and Results

The dynamic routing was used in this case-study as the network is wide-scale and includes more than 60 crossroads. Additionally, the final goal of the research is to forecast the re-distribution of the traffic flows after introducing significant changes (new elements, crossroads reconstruction etc.) in the transport network. The general principles of the dynamic assignment in VISSIM could be found in [2].

To implement DTA, VISSIM simulation software allows to provide the demand data in form of OD matrix (FMA files are used) or trip chain files (FZP files are used). In contrast to OD matrices a trip chain file allows to supply the simulation with more detailed travel plans for individual vehicles; however, the coding efforts are much higher. It is possible to mix traffic demand by OD matrices and trip chain files in the same simulation. In the initial (base) model as the demand data, only OD matrices were used.

There are a number of parameters in VISSIM dynamic assignment module, which could be used to control the behaviour of the model and do the calibration. In the same time, not so much information how to set up parameters for dynamic assignment, to get a valid model, could be found in literature, most significant sources of information are the reports and guidelines [3, 4], and software manuals [5]. Further recommendations of DTA use in VISSIM applied for this case-study are listed and commented.

4.1 DTA Configuration

Default Parameters. It is recommended to use VISSIM defaults parameters for dynamic assignment initially. The default values should not be changed, if you are not sure about the influence of the parameters on dynamic assignment procedure. In most cases, the options described below could give satisfactory results.

Multirun Simulation Options. The random seed increment option should be set to 0, in the case of applying dynamic assignment. This should speed up the convergence reaching, but it should not be forgot to provide the seed increment while doing simulation for data collection.

Dynamic Assignment Volume Increment. If the simulated network is congested much, it is recommended to apply dynamic assignment with volume increment. It is recommended to start with 30% of traffic volume and add 5% by each iteration (but this is highly dependent on how the network is congested). In the current case-study, the recommendation above led to successful results.

Convergence Criteria. VISSIM provides the following convergence conditions of the dynamic assignment (see Fig. 5):

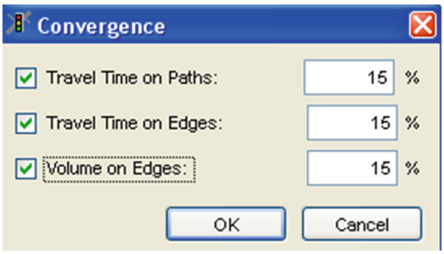


Fig. 5. Convergence options in VISSIM.

- *Travel Time on Paths* computes the change of the travel time on every path compared to its travel time in the previous run. If this change for all paths is lower than the user-defined factor, convergence for this criterion is detected;
- *Travel Time on Edges* the old and new travel times on edges are compared in the same way as described above;
- *Volume on Edges* the absolute difference of old and new volume on every edge is determined and compared with the user-defined number of vehicles.

VISSIM allows to use all three convergence criteria at the same time, but it is strongly unrecommended, as it could be a very high possibility, that the model will never reach the convergence. Travel Time on Paths option is recommended in most cases. It should be noted, that Volume on Edges is an option which could lead to slow convergence process, because the absolute values of volumes are compared and usually, if the network is loaded it is hard to reach. The tolerance values of 15% are recommended (default value). In current case study as a convergence criteria Travel Time on Paths option was selected with tolerance values of 15%.

Local Calibration. In case the model reached convergence, but the traffic behaviour is still unrealistic, for example, the traffic is using one route in a more intensive way, while others are used in reality, it is necessary to apply the local calibration. This option could be treated as an alternative way to add factors, which are not considered by VISSIM's decision model.

The cost of VISSIM links and connectors can be increased or decreased using surcharges. Surcharges are added to the total cost once per visit of a link/connector section (i.e. not per km). There are two types of surcharges (see Fig. 6):

- Surcharge 1 is sensitive to the weight for financial cost in the vehicle type cost coefficients;
- Surcharge 2 is simply added to the general cost and is independent of the cost coefficients of the vehicle type.

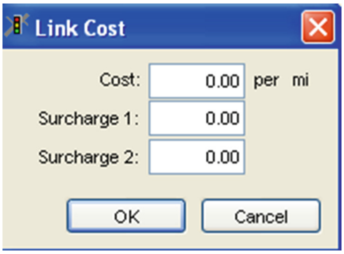


Fig. 6. Link costs.

It is recommended to apply Surcharge 2 as it provides more visibility to the calibration procedure. Surcharge 1 could be recommended, if there are problems with specific types of the vehicles, which are selecting not realistic paths (at example cargo vehicles). The cost per km could be another option, but usually, it is used to represent the financial part of the decision model. The general recommendations, based on case-study, of using Local calibration are following:

- Re-check the current state of the network for the errors in network coding, the usual case why traffic selects strange paths are the errors in network coding;
- Try to understand, why traffic is routed by different paths by comparing the cost values for each path (using VISSIM internal tools).

Only in case, if there are no errors in network coding and it is possible logically explain why VISSIM's decision model provides the obtained results apply some Surcharge 2 values. The open issue here is a value of Surcharge 2. There are no any well-defined guideline on the relationship between the surcharge value and the traffic divergence. In [6] is recommended to start with a surcharge value of 30. In this case-study the staring value was selected as 25.

The more rigid methods, available in VISSIM, as Edge closure and Route closure, are not recommended for practical use, especially in the cases when the model will be used to forecast future distributions of the traffic flows.

Restricting the Number of Routes. The number of routes are not limited by default in VISSIM. It means, that all routes are used for traffic distribution; it could lead to the problem, that very expensive routes still be used by vehicles. To avoid this situation the number of routes restriction possibility could be used in VISSIM. There are 2 options for restricting the number of routes (see Fig. 7):

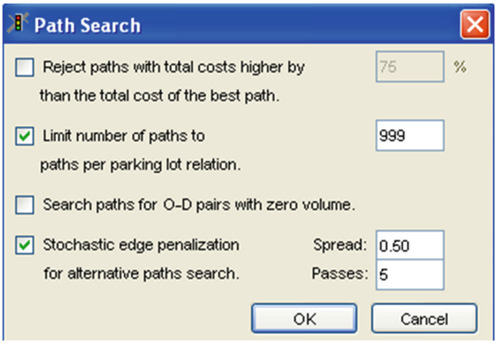


Fig. 7. Path search option.

- Defining an upper limit for the number of routes;
- Defining a maximum of the cost difference between the best and the worst route.

The use of the Path search options highly depends on the network, which has been simulated. If there are only a few realistic routes from zone to zone, the definition of the upper limit of routes could give a positive result. If for some OD pairs exist a significant number of routes and they are used and at the same time for some OD pairs only a few routes exist, it is recommended to define a maximum of the cost difference between the best and the worst route.

Costs Coefficients. Cost coefficients are an effective way to influence the behaviour of vehicles of different types. If the analysis of the traffic flows circulation shows the difference in behaviour, it is possible to control with providing coefficients for following factors: Travel time, Distance, Link Cost. Usually, the default values give satisfactory results during simulation, if no difference in behaviour could be foreseen (Fig. 8).

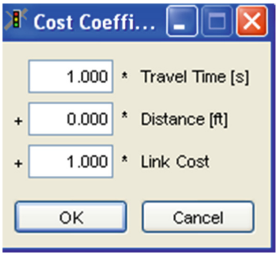


Fig. 8. Cost coefficient.

4.2 Results Analysis

Even in the case, if convergence conditions are true it is always necessary to check the validity of the model. Based on described above technical issues regarding dynamic traffic assignment procedure parametrization and turning the final model was obtained. The validation of the model was completed and validation results summary are presented in Table 2.

Table 2. Validation results summary.

Observation point	Direction	Simulated values	Counted values	Difference, %	GEH
C1	A	2658	2804	5.2	2.71
	B	4649	5016	7.3	5.27
C2	A	3003	3222	6.8	3.9
	B	1846	1967	6.2	2.77
C3	A	4626	4909	5.8	4.09
	B	4960	5267	5.8	4.29
C4	A	1297	1379	5.9	2.24
	B	1136	1223	7.1	2.53
C5	A	1562	1658	5.8	2.39
	B	2157	2302	6.3	3.07

Table 2 demonstrates the summary of validation results. Simulated values column refers to the model data (volume of traffic flow), after applying DTA, counted values column is the data collected during traffic survey. The GEH (Geoffrey E. Havers) values could be used to decide about the validity of the transport model. The GEH statistics is calculated as follows:

$$GEH = \sqrt{\frac{2(M - C)^2}{M + C}}, \tag{1}$$

where: M – is the hourly traffic volume from the model; C – is the real-world hourly traffic counts.

According to the GEH application rules, 85% of the volumes in a model should have GEH less than 5. As could be seen from the data only one value is higher, based on obtained data of GEH.

Additionally, to GEH calculation, the NAIVE approach was used to compare simulated and counted values by constructing the linear regression model between real data and the outputs from the model as defined in [7]. To evaluate validation results standard regression model quality indicators were used: R2, RMSE (root mean square error). The Fig. 9 represents the comparison of the counted values (real data) and simulated values (the data from the model) with approximation with regression line. The high value of R2 (close to 1) confirms that validation procedure was completed successfully.

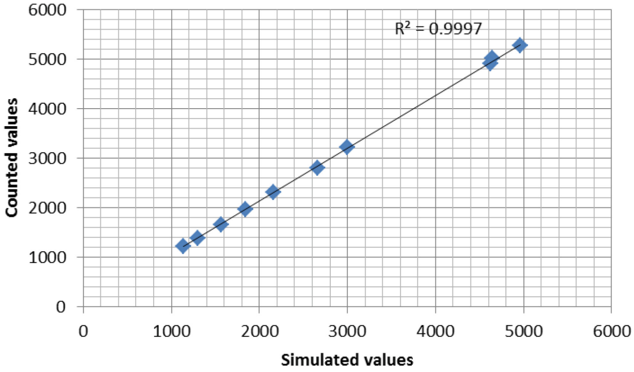


Fig. 9. Validation results by NAIVE approach.

Summing up validation results based on GEH test and NAIVE approach application it could be concluded that developed model with DTA is valid and could be used for further research.

Not looking to the fact, that application of DTA in this case-study was successful, there are some limitations of DTA [8], which should be considered by applying DTA in traffic flow simulation and analysis:

- Higher requirements to the initial data for simulation (OD matrix, data required for validation etc.);
- DTA models take more time and resources to be developed and calibrated (compare to static traffic assignment models);
- For long-term planning, the available level of input data required for DTA may not be available;
- In some cases, the traffic flow models within a DTA model may produce counterintuitive results that are difficult to explain, because they are the consequence of interactions taking place over the entire network and across multiple time periods.

5 Conclusions

The paper demonstrates the case-study of applying DTA procedure for simulation of the wide-scale transport network and draws recommendations about DTA use in PTV VISSIM simulation software. The provided recommendations are the summary of the analysis of the simulation guidelines, software manuals, scientific papers and the outcome from the lessons learned, implementing the presented case-study.

The DTA approach allows to build and run microscopic wide-scale traffic simulation models, but in the same time adds complexity in the stage of model development, calibration and validation. The DTA is beneficial in a decision making, as it provides the ability to forecast the behaviour of the traffic flow in case of the changes in transport infrastructure (use of new routes), based on predefined initial model. This gives a possibility to receive more substantiate results of the simulation for further decision making. Meanwhile, the use of DTA adds additional complexity to the model parametrisation, calibration and validation.

The presented case-study demonstrates the development of base model of the study area for different development scenarios evaluation in the future. The validity of the models was proved by visual inspection, calculation of GEH values and applying simplified NAIVE validation approach.

Acknowledgements. This work was supported by Latvian state research program 2014–2017 project "The next generation of information and communication technologies (NexIT)".

References

1. Savrasovs, M., Pticina, I.: Methodology of OD matrix estimation based on video recordings and traffic counts. Procedia Eng. **178**, 289–297 (2017)
2. Fellendorf, M., Vortisch, P.: Microscopic traffic flow simulator VISSIM. In: Fundamentals of Traffic Simulation, pp. 63−93 (2010)
3. FHWA: Traffic Analysis Toolbox Volume XIV: Guidebook on the Utilization of Dynamic Traffic Assignment in Modeling, Publication NO. FHWA-HOP-13-015 (2012)
4. Transport for London: Traffic Modelling Guidelines version 3 (2010). Edited by Dr. James Smith and Robert Blewitt
5. PTV Planung Transport Verkehr AG: PTV VISSIM MANUAL, PTV Planning Transport Verkehr AG (2012)
6. Alam, M., Ahsanul Habib, M., Quigley, K.: Vulnerability in transport network during critical infrastructure renewal: lessons learned from a dynamic traffic microsimulation model. Procedia Comput. Sci. **109**, 616–623 (2017)
7. Kleinen, J.P.C.: Validation of models: statistical techniques and data availability. In: Winter Simulation Conference, p. 7 (1999)
8. Chiu, Y., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., Hicks, J.: Dynamic Traffic Assignment: A Primer. Transportation Research Circular E-C153 (2011)

The Riga Public Transport Service Reliability Investigation Based on Traffic Flow Modelling

Irina Yatskiv (Jackiva)(✉), Irina Pticina, and Kateryna Romanovska

Transport and Telecommunication Institute, Lomonosova 1, 1019 Riga, Latvia
{Jackiva.I,Pticina.I}@tsi.lv,
kateryna.romanovska@inbox.lv

Abstract. The target of research is the investigation of an aspect of public transport service quality – namely, reliability from passenger's viewpoint. The paper presents the reliability assessment methodology with respect to the Riga Public Transport System (PTS) services by using microscopic traffic flow modelling and constructing and developing the integral indicator of the model outputs. The system of indicators analysing the reliability at different levels of hierarchy – from route and transport mode to the integral public transport system (or a fragment thereof) – is developed. The proposed approach makes it possible to model and incorporate some factors affecting the reliability – like weather, congestions, and the number of passengers – for evaluating the integral reliability indicator on the model.

To test the approach, the authors have used the simulation model of a fragment of the Riga public transport system, developed based on PTV VISION VISSIM software and data from the Riga traffic survey, which had been carried out in 2016. The implementation of this type of assessment furnishes the Riga transport authority with useful information on non-robust PTS fragments in the context of reliability of PTS service.

Keywords: Public transport · Reliability · Punctuality · Integral indicator
Traffic model

1 Introduction

1.1 Motivation

Public transport is considered to be a sustainable, viable alternative to private car use [1]. Public transport reliability is an important characteristic of transportation service quality both from service recipients' (passengers) and service providers' (public transport operators) points of view. Public transport system service quality is characterised by the components as follows: availability, accessibility, information, time, customer care, comfort, security and environment [2] and each of them is presented by a set of indicators. Peek and Van Hagen [3] have introduced the 'pyramid of Maslow for public transport' (Fig. 1) for prioritizing the quality factors in public transport. The lower part of the pyramid shows the elements that must be sufficient without doubt. The elements safety and reliability form the base of the pyramid. It may be stated that unreliability of public transport drives away both the existing and the prospective

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_24

passengers. Passengers will be dissatisfied and they are likely to avoid using public transport; the upper part shows the satisfiers which are additional quality aspects. These aspects satisfy passengers. At the same time, König and Axhausen concluded in [4] that the research done over the last decade shows that the reliability of the transportation system is the governing factor in the choice behaviour of people.

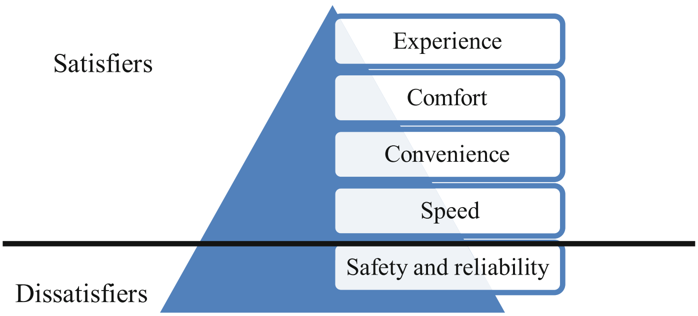


Fig. 1. Quality factors for public transport presented in pyramid of Maslow [3].

The goal of this paper is to develop and apply a methodology for assessing the reliability of the Riga Public Transport System (PTS) services by using integral indicator developed based on the microscopic traffic flow models' outputs. The proposed methodology introduces the use of Travel Time, Arrival Time, Probability of non-delay etc., which are estimated on simulation models for the lower hierarchy level for each route of public transport. Then, with the calculated weights for routes, the indicators are grouped into reliability indicator characterizing the specific mode of transport and, subsequently – into the integral indicator characterizing the whole PTS. The route weights are estimated based on attributes like passenger volumes, route length, frequency, and the number of stops; the weights of transport modes – on passenger volumes and number of routes.

The paper is structured as follows: the next section reviews the term reliability and different indicators in the context of Urban Public Transport System (UPTS). Section 2 presents the methodology used to evaluate UPTS reliability through the integral indicator scheme. Section 3 details the specification of the model for a fragment of Riga Transport System that was used for testing the approach. The results are presented in Sect. 4 and are subsequently discussed along with suggestions for further research in Sect. 5.

1.2 Urban Public Transport Reliability and Indicators

To target reliability as an objective, it is important to understand various definitions because it differently depends on the field to which it applies. For a UPTS service, the reliability is defined as the ability of the service to provide a consistent service over a period of time [5]. A definition of PT reliability associates with such concepts as punctuality, cancellations, variability and waiting time variability [6] and it is based on the analysis of the different types of time: travel time reliability, waiting time reliability, and arrival time reliability. In order to achieve reliability, the appropriate indicators should be identified for measurement, which makes it possible to assess the current reliability level and to target reliability by setting the corresponding benchmarks. In Table 1 different variants of PT reliability indicators are presented.

Table 1. Public transport system reliability indicators.

Characteristic of reliability	Indicators	Source
Cancellations	Missed trips	[7]
	Distance travelled between mechanical breakdowns	[7]
Punctuality	% passengers earlier/later than advertised	[2, 8]
	% of passengers arriving on time	[2]
	% of passengers arriving at the destination point in the scheduled time	[9]
	% of connection met	[7, 8]
	Adherence to schedules	[7]
	Percentage of travellers who are sent with delay/advance	[9]
	On-time performance	[7, 10]
	Run/time ratio	[7]
	Punctuality index	[7]
	Waiting time	[14]
Variability	Excess journey time	[2, 11, 12]
	Excess waiting time	[2, 13, 14]
	Excess access, egress and interchange time	[2]
	Excess ticket purchase time	[2]
	Travel time variability	[7]
	Additional time	[9, 11, 15, 16]
	Average departure deviation	[16]
	Percentage regularity deviation mean	[16]
	Standard deviation of actual transit times	[17]
	The coefficient of variation of route travel time	[17]
Additional	Transit-auto travel time	[7]
	Passenger perceptions of time	[7]
	Time accessibility indicators	[18]

All of these variants of reliability estimation approaches in Table 1 are connected with the analysis of real travel time, waiting time or arrival time compliance with the planned (or scheduled) time.

Methods that can be used to measure reliability typically include analytic, simulation, and heuristic approaches. Simulation approach, of course, is very time-consuming and it usually requires a great amount of data. However, the simulation modelling used to assess the UPTS reliability indicator supports the modelling of various sources of service uncertainty and different volumes of passenger flows, and their impacts on reliability can be easily assessed. The results showed in [14, 19–21] demonstrate that the simulation modelling framework will facilitate a more realistic reproduction of service reliability rather than generating it by embedding independent statistical processes of individual system components, and will make it possible to consider different measures of reliability.

2 Methodology

The offered approach to assess the reliability of urban public transport system is a hierarchical approach where initial reliability indicators for each route are estimated; subsequently, the reliability indexes are calculated for each transport mode and, finally, the integral public transport system reliability index for the analysed fragment of transport network (or for the whole city) is developed (see Fig. 2).

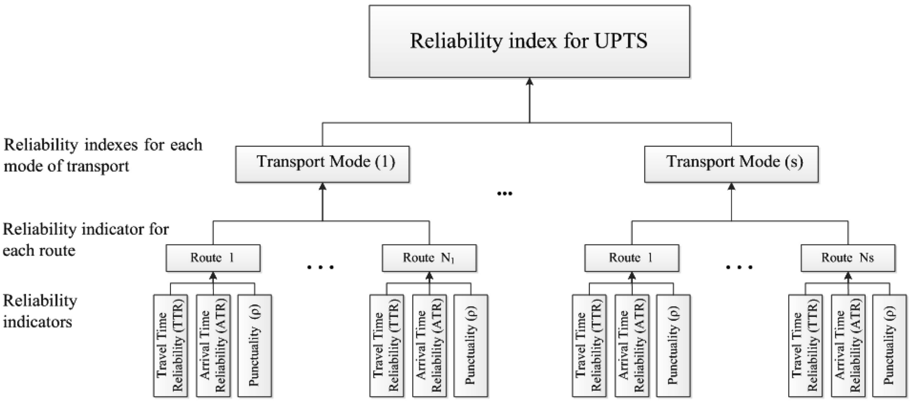


Fig. 2. The hierarchy of reliability index calculation.

The integral indicator (index) of the upper hierarchy level is the Reliability index for UPTS (or the UPTS fragment). The index was calculated on the basis of the additive function of Reliability indexes for transport mode *i* and weight of each mode in the specific UPTS:

$$RI = \sum\nolimits_{i=1}^{s} w_i K^i. \tag{1}$$

To calculate weight w_i, the attributes that characterized the role of this transport mode in UPTS were used according to the next formula:

$$w_i = \frac{1}{3}\left(w_i^{TP} + w_i^{Ni} + w_q^i\right) = \frac{1}{3}\left(\frac{TP_i}{\sum_{i=1}^{s} TP_i} + \frac{N_i}{\sum_{i=1}^{s} N_i} + \frac{q_i}{\sum_{i=1}^{s} q_i}\right), \tag{2}$$

where: TP_i – total passenger flow for transport mode i (i = 1, s), pass/h; N_i – number of routes for transport mode i; q_i – the probability of boarding of a vehicle for transport mode i.

Reliability index for transport mode i evaluates the reliability of this mode for all the routes (j = 1, N_i) where this transport mode operates, and is calculated as follows:

$$K^i = \sum\nolimits_{j=1}^{Ni} \gamma_j^i I_j^i, \tag{3}$$

where γ_j^i – weight of route j for transport mode i and I_j^i – reliability indicator of route j for transport mode i.

Weight of j route is estimated on the basis of M route attributes such as route length (L_i), km; service frequency on route (h_i), vech/h; number of stops on route (St_i), passenger flow (P_i), and passes/h. Let us consider that the weight of each attribute for route j and transport mode i is equal and in this case the weight of route j for transport mode i is calculated according to the formula as follows:

$$\gamma_j^i = \frac{1}{M} \sum\nolimits_{m=1}^{M} R_{i,j}^m, \tag{4}$$

and weight of route j for attribute m is calculated as follows:

$$R_{i,j}^m = \frac{W_{i,j}^m}{\sum_{i=1}^n W_{i,j}^m}, \tag{5}$$

where $W_{i,j}^m$ – the value of attribute m for route j and transport mode i.

Initial reliability indicators $\left(I_j^i\right)$ of route j for transport mode were determined through simulation and included the following [19]:

1. Travel Time Reliability (TTR) is the calculated difference between the modelled (TT_{mod}) and the scheduled (TT_{sched}) travel times:

$$TTR_{dif} = \Delta TT = |TT_{mod} - TT_{shed}|, \tag{6}$$

2. Arrival Time Reliability (ATR), which is measured as the delay time on the last stop:

$$ART = DT = |AT_{mod} - AT_{sched}|. \tag{7}$$

The next descriptive characteristics for TT_{mod} and AT_{mod} were used: mean, median, maximum, and 95% percentile. Also, the standard deviations of the modelled variant of travel time, as well as arrival time and coefficient of variation CoV, are considered.

3. The non-delay probability as the probability of arrival to the last stop with a delay not more than v minutes, and estimating it as the ratio between the journey number of vehicles which arrived to the last stop with the delay of no more than v minutes (n_v), and the total number of vehicle journeys (N):

$$\rho = \frac{n_v}{N}. \tag{8}$$

The procedure for these indicators estimation based on simulation was offered and described in [19].

3 Case Study. Riga Public Transport Service Reliability

3.1 Model Description

In this study, the microscopic model of a fragment of the Riga transport system was used, which was developed in TTI, in the Laboratory of Applied Software. The model was developed by using software application PT VISION VISSIM (version 7.0).

The total area of the simulated fragment of the transport network is 10 km^2. The general characteristics of the modelled object are as follows: the transport area is located approximately 5 km from the city center and covers 16 routes of public transport (bus, trolleybus, tram, and microbus) and 15 stops. Diagram of the simulated area of the Riga transport system and traffic volumes are presented in Fig. 3.

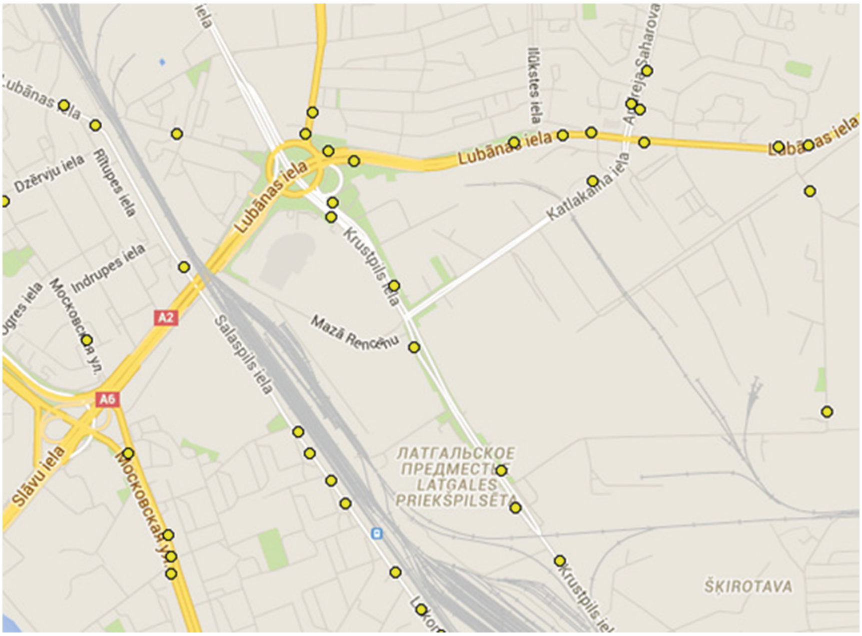


Fig. 3. The scheme of the modelled area of the Riga transport system.

The data sources for the model are as follows: the statistics on the existing traffic flow; the public transport routes and timetable; and traffic light signal cycles.

To simulate the morning rush hour, the time (8:15 am to 9:15 am) was selected because the morning rush hour is a common problem faced by the city of Riga, and because it affects the reliability of the whole public transport. The amount of simulation runs for each experiment is equal to 100; the dwell time for each stop is defined as the random value with normal distribution Normal (20, 2) s.

3.2 Data for Reliability Indexes Calculation

The investigated UPTS fragment includes 15 common stops and the next routes:

- 7 bus routes (No. 12, 13, 15, 18, 31, 49, 50);
- 3 trolleybus routes (№ 15, 16, 22);
- 3 tram routes (№ 3, 7, 9);
- 3 routes by minibuses (№ 204, 209, 216).

The values of route attributes and route weights for each transport mode are calculated based on the formulas (4, 5) in Table 2, and the weight of each transport mode in the Riga UPTS are calculated based on formula (2) and is presented in Table 3.

Table 2. Attributes of routes and route weights for each public transport mode.

	St_i	$R^1_{i,j}$	L_i	$R^2_{i,j}$	h_i	$R^3_{i,j}$	P_i	$R^4_{i,j}$	Weight γ^i_j
Buses									
No. 12	13	0.052	9.8	0.071	3	0.15	210	0.105	**0.095**
No. 13	46	0.184	24.06	0.175	2	0.1	240	0.120	**0.145**
No. 15	43	0.172	22.5	0.164	3	0.15	390	0.195	**0.170**
No. 18	34	0.136	17.9	0.131	3	0.15	190	0.095	**0.128**
No. 31	42	0.168	22.03	0.161	2	0.1	225	0.112	**0.135**
No. 49	38	0.152	23.04	0.168	4	0.2	460	0.229	**0.187**
No. 50	34	0.136	17.8	0.130	3	0.15	290	0.145	**0.140**
Sum	250	1	137.13	1	20	1	2005	1	1
Trolleybus									
No. 15	20	0.345	10.46	0.334	19	0.594	2280	0.646	0.480
No. 16	23	0.397	12.65	0.403	4	0.125	200	0.057	**0.245**
No. 22	15	0.259	8.25	0.263	9	0.281	1050	0.297	**0.275**
Sum	58	1	31.36	1	32	1	3530	1	1
Tram									
No. 3	39	0.402	21.4	0.412	1	0.091	230	0.086	**0.248**
No. 7	22	0.227	11.6	0.224	9	0.818	2250	0.836	**0.526**
No. 9	36	0.371	18.9	0.364	1	0.091	210	0.078	**0.226**
Sum	97	1	51.9	1	11	1	2690	1	1
Microbuses									
No. 204	38	0.404	20.4	0.393	3	0.1875	60	0.194	**0.295**
No. 209	31	0.330	18.4	0.355	3	0.1875	60	0.194	**0.266**
No. 216	25	0.266	13.08	0.252	10	0.625	190	0.613	**0.439**
Sum	94	1	51.88	1	16	1	310	1	1

Table 3. Weight for each public transport mode in Riga PTS.

Transport mode	M_i	w_M^i	TP_i	w_{TP}^i	q_i	w_q^i	Weight for transport mode, w_i
Bus	7	0.438	2005	0.275	1	0.311	**0.341**
Trolleybus	3	0.188	2280	0.313	0.962	0.299	**0.267**
Tram	3	0.188	2690	0.369	0.926	0.288	**0.282**
Microbus	3	0.188	310	0.043	0.325	0.101	**0.110**
Total	16	1	7285	1	3.213	1	1

4 Results

Results in Table 4 show that the most reliable mode of transport within this PTS fragment under study is the tram; with probability 1, the vehicle approaches the stop with a delay of no more than 3 min. With regard to microbus and trolleybus, the high reliability indexes are equal to 0.995 and 0.952, respectively. However, bus reliability is given below and it is equal to 0.734.

Table 4. Calculation of reliability indicators on routes.

	γ_j^i	ρ_i	$\gamma_j^i\rho_i$
Bus			
No. 12	0.095	1	0.095
No. 13	0.145	1	0.145
No. 15	0.170	0.48	0.082
No. 18	0.128	1	0.128
No. 31	0.135	0.49	0.066
No. 49	0.187	0.42	0.079
No. 50	0.140	1	0.140
Reliability Index (bus)			**0.734**
Trolleybus			
No. 15	0.480	0.99	0.475
No. 16	0.245	1	0.245
No. 22	0.275	1	0.275
Reliability Index (trolleybus)			**0.995**
Tram			
No. 3	0.248	1	0.248
No. 7	0.526	1	0.526
No. 9	0.226	1	0.226
Reliability Index (tram)			**1**
Minibuses			
No. 204	0.295	1	0.295
No. 209	0.266	1	0.266
No. 216	0.439	0.89	0.391
Reliability Index (minibus)			**0.952**

Based on reliability indicators and weights of each transport mode, and by using Eq. (1), the integral index of reliability of the public transport system fragment is calculated and presented in Table 5.

Table 5. Public transport system reliability calculation on analysed fragment of PTS.

Transport mode	w_i	K^i	$w_i K^i$
Bus	0.341	0.734	0.2503
Trolleybus	0.267	0.995	0.2657
Tram	0.282	1.000	0.282
Minibuses	0.110	0.952	0.1047
Integral index of public transport system reliability			**0.9027**

Thus, the reliability of the system of public transport in rush hours within the time interval 8:15 am to 9:15 am under normal conditions is high enough, and amounts to 0.9027 with the probability that a vehicle arrives at the stop with a delay of no more than 3 min.

5 Conclusion

UPTS reliability is considered as the most important attribute of quality from a passenger's viewpoint. The UPTS reliability monitoring should be day-to-day practice of transport authorities to ensure public transport development as a sustainable, viable alternative to private car use.

The PTS is a complex system with various sources of service uncertainty, and it is not possible to calculate the impacts of possible changes on PTS reliability level by applying the analytical approach. The simulation approach for reliability indicator estimation supports the modelling of various sources of service uncertainty, different volume of passenger flows etc.

This research offers an estimation of UPTS based on integral index construction, The approach consists of determining the primary indicators for each route on base simulation (continuing an example of evaluating a single route reliability in [19]), with the subsequent construction of the integral index taking in account the weights of routes based on its attributes and the weights of transport mode in a specific UPTS.

This approach was tested on the Riga TS fragment simulation model. In the further research, a set of experiments will be conducted to assess the impact of road traffic and passenger flow volume in PT on travel time and delay time.

References

1. Holmgren, J.: Meta-analysis of public transport demand. Transp. Res. Part A **41**(10), 1021–1035 (2007)
2. European Union, EN 13816 Transportation – Logistics and services – Public passenger transport – Service quality definition, targeting and measurement, European Standard (2002)
3. Peek, G.J., Van Hagen, M.: Creating synergy in and around stations: three strategies in and around stations. Transp. Res. Rec. J. Transp. Res. Board **1793**, 1–6 (2002)
4. König, A., Axhausen, K.W.: The reliability of the transportation system and its influence on the choice behaviour. In: Proceedings of the 2nd Swiss Transportation Research Conference, Monte Verità (2002)
5. Polus, A.: Modeling and measurements of bus service reliability. Transp. Res. **12**(4), 253–256 (1978)
6. Vincent, M.: Measurement valuation of public transport reliability. s.l.: Wellington, N.Z.: Land Transport New Zealand (2008)
7. Kittelson and Associates: A Guidebook for Developing a Transit Performance Measurement System. Transit Cooperative Research Program (TCPRT), report 88, Washington, D.C. (2003)
8. Cascetta, E., Cartenì, A.: A quality-based approach to public transportation planning: theory and a case study. Int. J. Sustain. **8**(1), 84–106 (2014)
9. Dragu, V., Roman, E.A., Roman, V.C.: Quality assessment in urban public transport. Theor. Empir. Res. Urban Manage. **8**(3), 32–43 (2013)
10. Eboli, L., Mazzulla, G.: A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. Transp. Policy **18**(1), 172–181 (2011)
11. Frumin, M., Uniman, D., Wilson, N.H.M., Mishalani, R., Attanucci, J.: Service quality measurement in urban rail networks with data from automated fare collection systems. In: Proceedings of CASPT Conference, Hong Kong (2009)
12. Uniman, D.: Service quality measurement using AFC smart card data–a framework for the London underground. Master thesis. MIT, Boston (2009)
13. Eboli, L., Mazzulla, G.: Performance indicators for an objective measure of public transport service quality. Eur. Transp. **51**, 1–21 (2012)
14. Cats, O., Gkioulou, Z.: Modelling the impacts of public transport reliability and travel information on passengers' waiting time uncertainty. EURO J. Transp. Logist. **6**, 247–270 (2014)
15. van Oort, N.: Incorporating enhanced service reliability of public transport in cost-benefit analyses. Publ. Transp. **8**(1), 143–160 (2016)
16. van Oort, N.: Service reliability and urban public transport design, TRAIL. Ph.D. thesis Series T2011/2, Delft, The Netherlands (2011)
17. Guido, G., Vitale, A., Rogano, D.: Assessing public transport reliability of services connecting the major airport of a low density region by using AVL and GIS technologies. In: IEEE 16th International Conference on Environment and Electrical Engineering (2016)
18. Coelho, P., L'os, P., Spina, G.F., Karyda, M.: Assessment of the Urban Mobility System. University of the AEGEAN, Greece (2008)
19. Yatskiv, I., Pticina, I., Savrasovs, M.: Urban public transport system's reliability estimation using microscopic simulation. Transp. Telecommun. **13**(3), 219–228 (2012)
20. Sorratini, J.A., Ronghui, L., Shalini, S.: Assessing bus transport reliability using micro-simulation. Transp. Planning Technol. **31**(3), 303–324 (2008)
21. van Oort, N., Brands, T., de Romph, E., Flores, J.A.: Unreliability effects in public transport modelling. Int. J. Transp. **3**(1), 113–130 (2015)

Modelling Kinetics of Dynamic Crack Propagation in a Gas Mains Pipe as Cyclic Random Process

Iaroslav Lytvynenko[1], Pavlo Maruschak[1(✉)], Olegas Prentkovskis[2], and Andriy Sorochak[1]

[1] Ternopil National Ivan Pul'uj Technical University, Ruska 56 street, Ternopil 46001, Ukraine
maruschak@tu.edu.te.ua

[2] Vilnius Gediminas Technical University, Plytinės g. 27, 10105 Vilnius, Lithuania
olegas.prentkovskis@vgtu.lt

Abstract. The approach to the mathematical analysis of fracture toughness of the gas mains steel is proposed, which allows obtaining informative features of the quantitative evaluation of the dynamic pipe fracture kinetics, taking into account the stochastic and cyclic nature of the process. It creates the possibility to model the kinetics of crack propagation in the gas mains pipe and analyzing individual stages of its growth.

Keywords: Model · Cyclic random process · Dynamic pipe fracture
Crack propagation rate

1 Introduction

At present, the methods for evaluating fracture toughness of the metal of oil and gas mains are actively developing. However, most of them, unfortunately, do not reproduce the real operating conditions of the gas pipeline, since they are usually comparative or biased [1, 2]. Modern oil and gas mains are made of high strength steels, and the relevant methods for evaluating fracture toughness require further development and refinement [2, 3].

The most informative method is carrying out field tests of pipe sections under the conditions of a crack start. Such experiments take into account the impact of temperature, stress-strain state, fracture kinetics, scale factors, and others. A significant limitation of such tests is their high cost, which requires an in-depth study of the results and processing of experimental data [4]. Creation of fracture models of natural objects will allow carrying out theoretical and experimental generalization of the results, as well as proposing measures for the possible prevention or reduction of the probability of accidental fracture of oil and gas pipelines.

It is known that the conditions of gas mains fracture, their test methods and the criteria for evaluating fracture toughness of the pipe steel have certain common features. In particular, they have a significant energy intensity of the fracture process, which is determined by the pipe steel properties. It should be noted that when a crack starts with a high propagation rate, this causes brittle fracture, however, with an increase in the

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_25

crack length and a decrease in the gas pressure, the ductile properties of the pipe material [5] have a significant effect.

As a rule, single tests of pipes do not allow to fully reproduce the test conditions along the entire trajectory of the crack growth in the conditions of its dynamic propagation. This can be achieved by modelling this process and developing relevant approaches to describing the growth patterns of a crack. The methods for describing the crack propagation kinetics must be physically correct and mathematically grounded. Modern methods of strain gauging allow recording the crack propagation rate during the field test of the pipe, which is a prerequisite for an in-depth analysis of its kinetics. However, it remains relevant to create effective computer systems for the automated processing of received signals with the formation of the previous physical and mechanical conclusion about the regulations in fracture of the object. Thus, the development of approaches to the modelling of the dynamic crack propagation kinetics in the pipe is the basis for the creation of new technological approaches to the manufacture of high strength pipes with the preservation of high ductility of the material [6, 7].

The purpose of this work is to use the mathematical model of a cyclic random process and develop on its basis scientific approaches to the analysis of the dynamic fracture process of the gas mains pipe, taking into account the stochastic and cyclic nature of this process.

2 Description of the Crack Propagation Kinetics as a Combination of Two Sections

Several stages of fracture of the gas mains pipe were considered:

– the formation of the initial crack and its start – the rapid propagation of the crack under the influence of internal pressure in the pipe;
– crack growth, at which its length increases and the propagation rate decreases due to changes in the stress-strain state;
– "stabilization" of the crack propagation rate due to "self-organization" of the fracture process of the pipe material;
– loss of stability and ultimate fracture of the pipe.

The curve can be conventionally divided into two sections:

The first section. On the first section (l_{stab}), the propagation rate changes from a certain maximum value $V_{\max}$ and enters the plateau (gets stable) with the value V_{stab}. We can assume that on the first section (l_{stab}) the propagation rate varies according to the exponential regularity. Best of all this pattern is described by the function

$$f(l) = V_{\text{stab}} + (V_{max} - V_{\text{stab}})e^{-l/\text{t}}, \tag{1}$$

where the maximum propagation rate is $V_{\max} = 400$ m/s; the stabilized propagation rate is $V_{stab} = 200$ m/s; $t = 50$ is the parametric coefficient (which adjusts the curvature of the function).

Figure 1 shows the approximation of experimental data by an exponential function

$$f(l) = V_{\text{stab}} e^{b/l}, \tag{2}$$

where $b = 15 \div 20$ is the parametric coefficient.

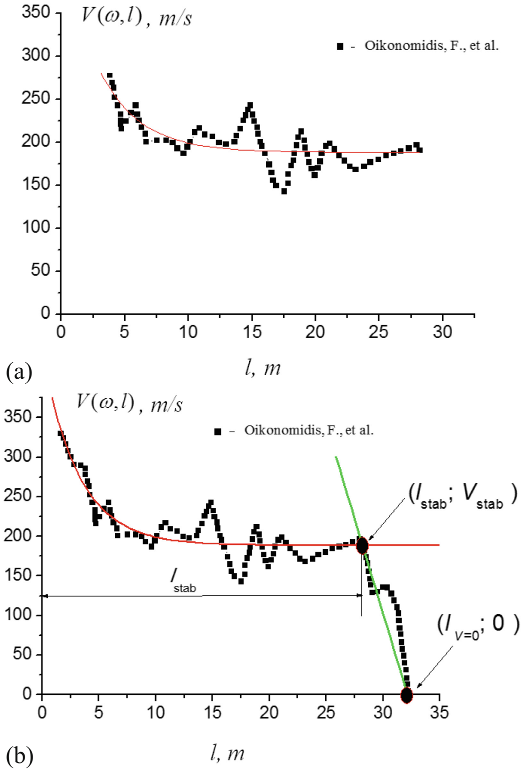


Fig. 1. Experimental dependence of the crack propagation rate on the trajectory of its growth [2, 3] and its approximation: (a) description of the first section only; (b) description of both sections ($V_{max} = 433.18$; $V_{stab} = 188.46$; $t = 3.19$).

Function (2) describes experimental data quite accurately, however, at $l \to 0$ the propagation rate goes to infinity. Thus, at $l \approx 0 \div 30$ formula (2) will yield incorrect results.

The second section. This section precedes the complete stop of the crack. It can be described by a linear function that decreases from a point with coordinates (l_{stab}, V_{stab}) to zero, Fig. 1b.

This section, which describes regularities in the gas mains fracture, was not analyzed in this paper.

3 Approach to the Analysis and Modelling of the Dynamic Crack Propagation

We considered the kinetics of dynamic crack propagation in the gas mains pipe in the form of an additive model, which is the sum of the deterministic exponential function and cyclic random process. This allowed to rationally applying the methods of statistical analysis of the crack propagation kinetics in a gas mains pipe for the tasks of its analysis, modelling and possible further prediction.

An additive mathematical model is used to describe the kinetics of the dynamic crack propagation:

$$V(\omega, l) = f(l) + \xi(\omega, l), \omega \in \mathbf{\Omega},\ l \in \mathbf{R} \tag{3}$$

where $\xi(\omega, l)$ is the cyclic component of the signal, whose model is a cyclic random process; $f(l)$ is the deterministic exponential function, which reflects the trend (trend component) of decreasing the rate of dynamic fracture of the pipe depending on the trajectory of crack growth, formula (1);

At the first stage of processing the experimental dependence "dynamic crack propagation rate – fracture length", the experimental data on the gas mains fracture (Oikonomidis et al.) published in [2, 3] were taken as initial data. Of these, trend component $\{f(l),\ l \in \mathbf{R}\}$ and cyclic component $\{\xi(\omega, l),\ \omega \in \mathbf{\Omega},\ l \in \mathbf{R}\}$ were identified. The procedure for identifying the trend consisted in describing the data by exponential dependence and finding corresponding coefficients.

The cyclic component was determined by subtracting the trend function $\{f(l),\ l \in \mathbf{R}\}$ from the input process $V(\omega, l)$, namely:

$$\xi(\omega, l) = V(\omega, l) - f(l), \omega \in \mathbf{\Omega},\ l \in \mathbf{R}. \tag{4}$$

Some results of statistical processing of the data on the gas mains fracture. $V_\omega(l)$ is the realization of the investigated crack propagation rate process at the input. $\xi_\omega(l)$ is the realization of the cyclic component of the investigated crack propagation rate process.

To analyze this component, it is necessary to use the methods of segmentation [8], define the discrete rhythm function, interpolate it [8, 9] and use the obtained results for the application of statistical estimation methods [10].

The methods of statistical analysis of cyclic random processes are adapted to the problems of analysis of the gas mains fracture rate to evaluate such probabilistic characteristics as the initial moment function of the first order (mathematical expectation) and the central moment function of the second order (dispersion).

The realization of the estimation of mathematical expectation is given in the following form

$$\hat{m}_\xi(l) = \frac{1}{M} \sum_{n=1}^{M} \xi_\omega(l + L(l, n)),\ l \in \mathbf{W}_1 = [L_1, L_2), \tag{5}$$

where $L_1 \neq 0$ in the general case, M is the number of cycles of the cyclic process (cyclic component), $\xi_\omega(l)$ is the realization of the cyclic component of the crack propagation rate (the realization of the cyclic random process), $L(l, n)$ is the rhythm function of the a cyclic random process, $\mathbf{W}_1$ is the area of definition of the first cycle of the process.

The realization of the dispersion estimation was defined as:

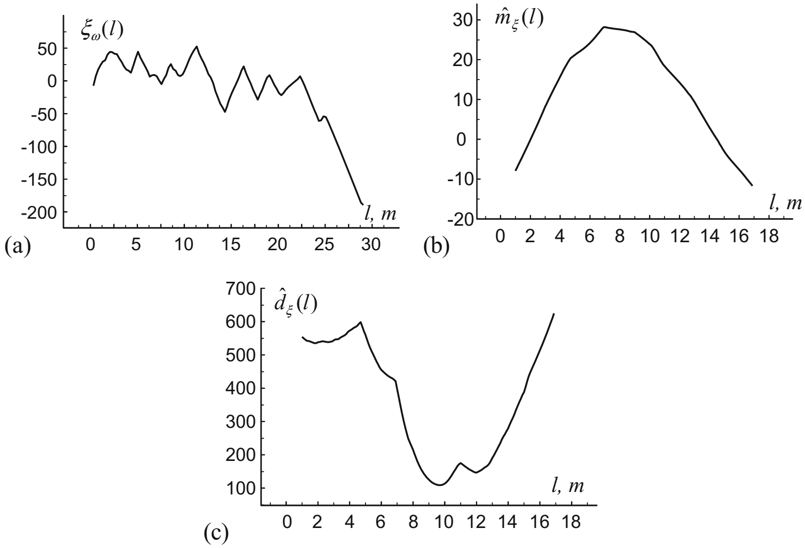


Fig. 2. Cyclic component of crack propagation: (a) statistical evaluation of its mathematical expectation; (b) and dispersion; (c) in the analysis of crack propagation rate in a gas mains pipe on the basis of the proposed model.

A decrease in the intensity of deformation processes in local sections of the material in the vicinity of the crack tip, which propagates under conditions of dynamic fracture of the gas pipeline, causes a decrease in the rate of fluctuations, leading to changes in the parameters presented in Fig. 2a. According to the analysis of the obtained data, it should be noted that the obtained estimates of mathematical expectation and dispersion, Fig. 2b, are sensitive to fluctuations in the fracture rate of the gas mains, which indicates the possibility of using these numerical features in automated systems of technical diagnostics.

In order to test the mathematical model and the method of analysis used in this work, a series of simulation experiments was conducted to compare actual and simulated realizations of fracture rates of the gas mains, taking into account the obtained statistical estimates and the trend component.

4 Simulation of the Dynamic Crack Propagation

The accuracy of computer simulation of the fracture rate of the gas mains by the proposed method was evaluated. As can be seen from the data in Fig. 3a, b, the proposed methods of statistical processing of the fracture rate of the gas mains take into account the changing nature of the main crack propagation rhythm [2, 3].

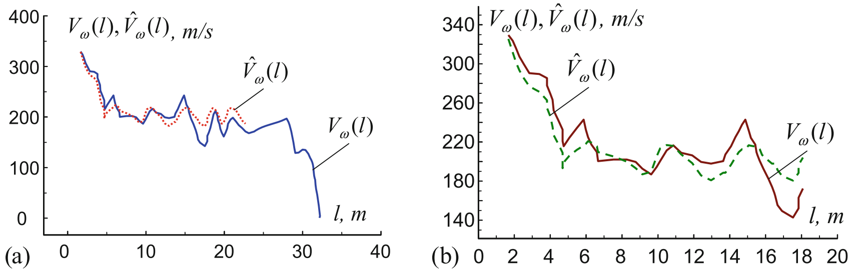


Fig. 3. Comparison of simulation results (a, b) of the crack propagation rate in a gas mains pipe with experimental data: $\hat{V}_\omega(l)$ is the simulated realization of the crack propagation in the gas mains pipe, taking into account statistical estimates of the cyclic component and the trend component, $V_\omega(l)$ is the realization of the crack propagation rate process at the input.

It is established (Fig. 4) that the relative mean square error of modelling does not exceed (in experiments carried out) 0.3%, which is satisfactory for engineering evaluation. It should be noted that the comparative analysis into the accuracy of the reproduced fracture rate of the pipe is individual for each realization of the process of dynamic crack propagation. Figure 4 shows the results of realization of the crack propagation rate in the gas mains pipe.

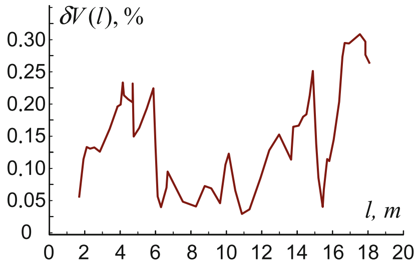


Fig. 4. Determining the relative error in percent simulation of the crack propagation rate in the gas mains pipe and its comparison with experimental data.

In addition, it should be noted that in all the series of statistical experiments conducted on the processes of dynamic fracture of the [11] gas mains pipe, the fact of repeatability (approximate to cyclic repeatability) of the statistical estimates of the probabilistic characteristics of processes and the variability of their rhythm functions is empirically confirmed, indicating the adequacy of the developed mathematical model and method, sufficient substantiation of their use in problems of modelling, analysis and forecasting of dynamic fracture processes of the gas mains pipe.

5 Conclusions

A new mathematical model of the gas mains fracture is developed in the form of the sum of the deterministic exponential function and the cyclic random process. The method of describing the dynamic fracture of the gas mains pipe based on the model of a cyclic random process for the problems of processing and modelling of crack propagation kinetics is proposed. The algorithm is developed and the software package is created, which is based on the proposed mathematical model of the analysis of the two-dimensional crack growth structure along its trajectory.

The methods of statistical analysis of the gas mains fracture process on the basis of the proposed mathematical models are substantiated, and the accuracy of the modelling is estimated. Relative modelling error does not exceed 0.3%.

In further research, it is planned to study the application of the polynomial function as a model of the trend component and compare the obtained results of modelling the crack propagation rate in the gas mains on the basis of two approaches using the polynomial and exponential trend components.

References

1. Scheider, I., Nonn, A., Völling, A., Mondry, A., Kalwa, C.: A damage mechanics based evaluation of dynamic fracture resistance in gas pipelines. Procedia Mater. Sci. **3**, 1956–1964 (2014)
2. Oikonomidis, F., Shterenlikht, A., Truman, C.E.: Prediction of crack propagation and arrest in X100 natural gas transmission pipelines with the strain rate dependent damage model (SRDD). Part 1: a novel specimen for the measurement of high strain rate fracture properties and validation of the SRDD model parameters. Int. J. Press. Vessels Pip. **105–106**, 60–68 (2013)
3. Oikonomidis, F., Shterenlikht, A., Truman, C.E.: Prediction of crack propagation and arrest in X100 natural gas transmission pipelines with a strain rate dependent damage model (SRDD). Part 2: Large scale pipe models with gas depressurization. Int. J. Press. Vessels Pip. **122**, 15–21 (2014)
4. Arabei, A., Pyshmintsev, I., Shtremel, M., Glebov, A., Struin, A., Gervasev, A.: Resistance of X80 steel to ductile-crack propagation in major gas lines. Steel Transl. **39**(9), 719–724 (2009)
5. O'Donoghue, P.E., Kanninen, M.F., Leung, C.P., Demofonti, G., Venzi, S.: The development and validation of a dynamic fracture propagation model for gas transmission pipelines. Int. J. Press. Vessels Pip. **70**(1), 11–25 (1997)
6. Yu, P.S., Ru, C.Q.: Strain rate effects on dynamic fracture of pipeline steels: finite element simulation. Int. J. Press. Vessels Pip. **126**, 1–7 (2015)
7. Botros, K.K., Geerligs, J., Rothwell, B., Carlson, L., Fletcher, L., Venton, P.: Transferability of decompression wave speed measured by a small-diameter shock tube to full size pipelines and implications for determining required fracture propagation resistance. Int. J. Press. Vessels Pip. **87**(12), 681–695 (2010)

8. Lytvynenko, I., Maruschak, P., Lupenko, S., Panin, S.: Segmentation and statistical processing of geometric and spatial data on self-organized surface relief of statically deformed aluminum alloy. Appl. Mech. Mater. **770**, 288–293 (2015)
9. Lytvynenko, I.V.: Method of the quadratic interpolation of the discrete rhythm function of the cyclical signal with a defined segment structure. J. Ternopil Nat. Tech. Univ. **84**(4), 131–138 (2016)
10. Lytvynenko, I.V., Maruschak, P.O., Lupenko, S.A., Popovych, P.V.: Modeling of the ordered surface topography of statically deformed aluminum alloy. Mater. Sci. **52**(1), 113–122 (2016)
11. Lytvynenko, I.V., Maruschak, P.O., Poberezhnyi, L.Y., Baran, D.Y.: Analysis of kinetics of dynamic crack propagation in a gas mains pipe as cyclic random process. In: Proceedings of the International Science and Technology Conference "Oil and Gas Power Engineering – 2017", 15–19 May, Ivano-Frankivsk National Technical University of Oil and Gas, pp. 379–382 (2017). (in Ukrainian)

Multi Criteria Decision Making in Transportation

Multi Criteria Decision Making in Life Cycle Management of Modular Ships with Test System

Igor Kabashkin[1(✉)] and Andrejs Zvaigzne[2]

[1] Transport and Telecommunication Institute, Lomonosova 1, Riga 1019, Latvia
kiv@tsi.lv

[2] Latvian Maritime Academy, 12-k Flotes Street, Riga 1016, Latvia

Abstract. Modular construction is one of the popular methods used in shipbuilding industry. The modular ship construction (MSC) requires better than normal quality assurance that must be provided with special product testing. Test-related information is generated throughout the product life cycle for each level in the system hierarchy. The present paper introduces a multi-level decision-making approach for design of optimal test architecture for achieving efficient system for lowering life-cycle cost of modular ship construction. The effectiveness of such approach is investigated through the optimization of test system architecture for known reliability and economic dependence during life cycle of MSC.

Keywords: Modular ships · Life cycle · Test system

1 Introduction

There are several current challenges to the competitiveness of the maritime sector. The short sea shipping industry is facing tougher requirements on the effective performance of vessel technologies.

Sustainable ship production takes a perspective that extends beyond competitive manufacturing and operations and includes minimizing the environmental burden associated with every aspect in the product's lifecycle, from design, to manufacturing and use.

Construction selection is crucial in many engineering projects as it can determine the durability, cost, and manufacturability of final products. In today's integrated design processes, a systematic selection of the best construction for a given application begins with identifying multiple technical properties, environmental impact factors, and life cycle costs of candidate architectures. When multiple criteria from different disciplines are to be satisfied in a construction selection problem, however, complexities often rise with regards to criteria conflicts and the importance of each criteria.

The process of building a new sea ship requires thorough research and consideration; all factors, criteria and risks connected with the project need to be analysed before designing the ship. As with each project, the available knowledge is quite limited in the beginning, but it is also the time when the most important decisions are made. At the end of the project, the available knowledge is considerably higher, but not many major decisions are left to be made.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_26

Reducing total ownership cost and shortening acquisition cycle time are central goals of the maritime industry efforts to revolutionize business affairs. A key systems engineering initiative the stakeholders have launched to address these goals is simulation-based acquisition, in which product information is reusable and robust enough to support cross-functional needs as an enabler of integrated product/process development during life cycle of ships.

The present paper introduces a general approach and application models for holistic multi-criteria decision and decision support for sustainable life cycle management for the maritime sector. It describes a model-based methodology, specifically the test requirements model, which can be shown to facilitate the transfer of test-related product information between various stages of the life cycle.

The rest of this paper is organized as follows. In Sect. 2 some important works in the area of multi criteria decision making in life cycle management of multifunctional ships are reviewed. In Sect. 3 the main definitions and assumptions are presented. In Sect. 4 a model of multi-level decision-making approach for design of optimal mission modularity of multifunctional special ships embedded integrated diagnostic architecture are proposed and optimal solution is designed. In Sect. 5 the conclusions are presented.

2 Related Works

The problems with ship construction optimization were well covered in [1, 2]. Cost reduction during the construction phase was explained in [3–5]. Sustainable ship production takes a perspective that extends beyond competitive manufacturing and operations and includes minimizing the environmental burden associated with every aspect in the product's lifecycle, from design, to manufacturing and use. In the project [6] decision support for sustainable ship production in global fluctuating markets are studied.

Properly allocating resources to achieve system development objectives has remained a challenge in systems engineering management. Different optimization models have been proposed for system development planning. Paper [7] proposes a multi objective optimization model to simultaneously optimize the maturity of the system's selected functions or capabilities and the consumption of development resources. An evolutionary algorithm is used to obtain a Pareto set of optimal solutions, where each solution represents a development plan informing which system components should be advanced to which maturity levels without exceeding a certain amount of resource consumption with correlation to typical system development lifecycle. In [8] the optimization problem of designing a universal platform is formulated and genetic algorithm to solve the optimization problem is used. Multi-criteria decision analysis on the base of Analytic Hierarchy Process has been applied for the Norwegian maritime sector in [9].

One of the trends in ship building today is modularity [10, 11]. The reason why modular construction process become popular in ship building industry is because there is no other method available that allows you to start and finish a project in a more-timely manner than modular. Modular construction process is one of the popular method uses in shipbuilding industry. Different construction, maintenance, operation benefits and cost models for modular ship construction are discussed in [12]. Multi-criteria decision making for different type of specialization modules for multifunctional special ships is discussed in [13].

The modular ship construction (MSC) requires better than normal quality assurance that must be provided with special product testing [12]. Product testing is a function that is repeated at each phase of the product life cycle and at each level of the system hierarchy. Test-related information is generated throughout the product life cycle for each level in the system hierarchy. Such information is often underutilized, if not discarded, in each subsequent life cycle phase. The result is that engineers are forced to reinvent the test information, significantly increasing the cost and schedule, while reducing the confidence of the testing process.

The present paper introduces a general approach for holistic multi-criteria decision and decision support for sustainable life cycle management for the modular ship construction and describes a model-based method, which can be shown to facilitate the transfer of test-related product information between various stages of the life cycle.

3 The Definitions and Assumptions

The following symbols have been used to develop equations for the models:

$k = \overline{1, l}$	Levels of structural modularity of system architecture
n	Number of modular subsystems
$U = \{\pi_i\}, i = \overline{0, n}$	Architecture of test system (TS) with π_i channels for diagnosis
$C_p(U)$	Cost for the development and production of TS with U architecture
$C_o(U, t)$	Cost for the identification of failures in system with U architecture of TS during time t of system operation
π_0	Channel for diagnosis of system in general
$\pi_i, i = \overline{1, n}$	Channels for diagnosis of subsystems
$z_i, i = 0, n$	Cost of channel development and production for diagnosis of i-th element
$c_i, i = \overline{0, n}$	Diagnostic system cost, if it has a channel for the diagnosis of i-th element
s	The average cost of diagnosis of one failure by TS with U architecture
u	Number of failures in the system during life cycle
λ_i	Failure Rate of i element of system
$\Lambda = \sum_{i=1}^{n} \lambda_i$	Failure rate of subsystem with $i = 1, \ldots n$ elements

In this work, we apply multi-level decision-making to a system with modular hardware and software structure of identical or non-identical deteriorating components. The system has multilevel structure submitted by several levels of k division reflecting the correspondent depth of the search of the defects and failures: for example, level $k = 0$

– level of system in general; $k = 1$ – level of subsystems; $k = 2$ – level of modules for subsystem; $k = 3$ – level of submodules, etc.

The efficiency of TS for above mentioned system is assessed by the total cost of life cycle

$$C(U, t) = C_p(U) + C_o(U, t). \tag{1}$$

The problem of embedded architecture design for integrated diagnostics we formulate as development of TS with optimal architecture U_{opt} with minimal total cost during life cycle T_0:

$$C\left(U_{opt}, t\right) = min\left\{\left(C(U, t)|t = T_0\right)\right\}. \tag{2}$$

4 Model Formulation and Solution

The system design model as a decision-making system is conceptually defined by two developing components: the design object and the design process. As external parameters of the system design model, target, functional, technical and resource requirements and constraints are applied to both the design object and the actual design process (Fig. 1).

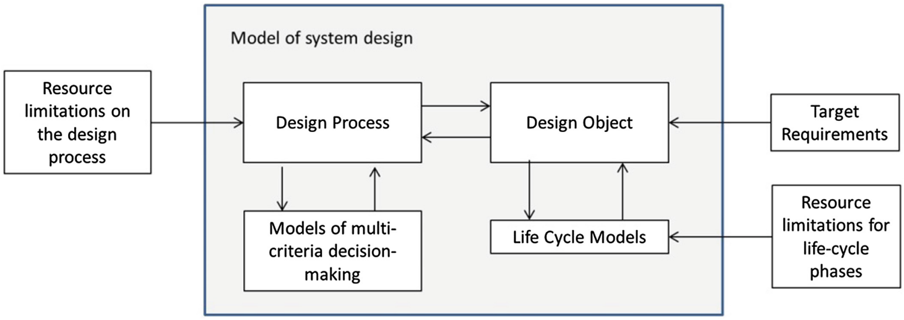


Fig. 1. Model of system design.

In general, the MSC design process can be represented through the evolution of its states:

$$H = \left\{h_i\right\} = \{T, X, W, S, U\}, i = \overline{1, m},$$

where T is the set of time moments during the life cycle in which the system is observed, S is the set of solutions (alternatives, system variants), X is the set of functional system requirements and technical requirements for the system, U is the set of strategies for managing design decisions.

The functional requirements of the X system are provided by the conformity of the designed system to certain technical characteristics of W, which can be determined with the help of the operator ω: W $\rightarrow X$.

The operation of the MSC requires certain costs C to provide its functioning, determined by the evolution of its states, the formation mechanism of which can be specified by the operator z: $H \rightarrow C$.

The purpose function of the MSC can be determined by the correspondence of its states to the set of variants of the structural realization B that characterize the conditions for the correspondence of the MSC to the specified functional and technical parameters, the mechanism of the interconnection of which is described by the operator φ: $H \rightarrow B$ (Fig. 2).

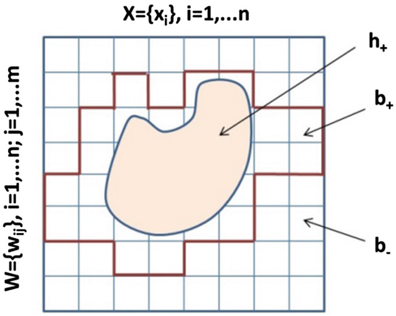


Fig. 2. Pattern of requirements and real project solutions.

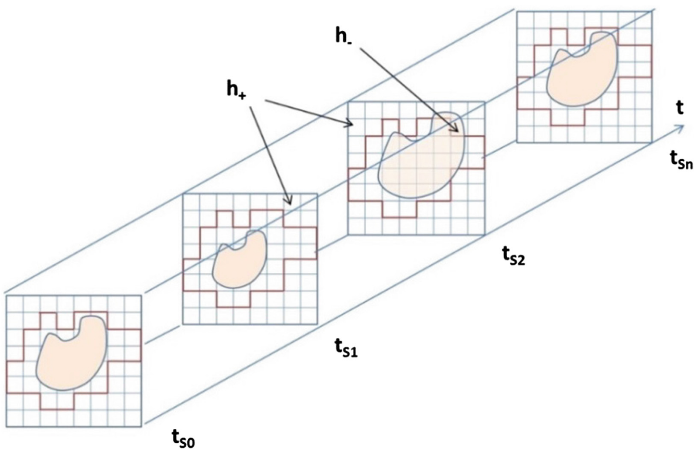


Fig. 3. Evolution of the domain of specified requirements and real project solutions in the MSC life cycle.

In this case, in accordance with the general requirements for the MSC, the set B can be represented by a set of favourable conditions from design point of view states b_+ and the set of inadmissible states b_- in which the specified requirements are not satisfied. In this case $B = b_+ \cup b_-$.

The set of admissible $\alpha: h_+ \to b_+$ and unacceptable $\beta: h_- \to b_-$ states of the system $H = h_+ \cup h_-$ can be associated with the indicated variants of the structural realization (Fig. 3).

The set of states $h_i \in H$ forms a space that corresponds to the chosen strategy $u_i \in U$ of the project solutions management:

$$h_i = h\left(t_j, x_l, s_k, u_i\right):\forall\left(t_j \in T, x_l \in X, s_k \in S\right).$$

The set of MSC states forms a system of patterns which are changed during the life cycle at certain time points t_{Si} (Fig. 3). They can correspond (time moments t_{S0}, t_{S1}, t_{Sn} at Fig. 3) or mismatch (time moment t_{S2} at Fig. 3) to the given functional and technical requirements for the system.

The task of improving the efficiency of MSC design can be formulated in terms of the task of optimal operation planning, as the search for optimal control strategies for designing $u_{opt} \in U$ that meet the criterion

$$u_{opt} = u_i:\left\{(h_i \to \min{}_i\, C)\, \Lambda\, \left(h_i \in h_+ \to b_+\right)\right\}.$$

The application of this objective function to the solution of the problem requires the determination of the type of criterion functional $z: H \to C$ and functionals of the constraints $\varphi: H \to B$, as well as the mechanisms of the influence of the optimized components of the MSC on the evolution of its states.

In this case one of the important tasks for investigation is monitoring of states evolution to define the moment of transition from set of favourable states b_+ to set of inadmissible states b_-. For such monitoring the test system (TS) can be used. It carry out the diagnosis of design object and identify its states h_+ and h_-.

Let us examine the subsystem on the lowest structural level l of the system, which consist of n elements. The diagnosis of the technical condition of the examined subsystem and its elements could be performed with TS, which can have $\pi_k \in \pi_i, i = 0, 1, \ldots n$ diagnostics channels (DC). Diagnostics with π_0 DC to determine the defects arising in any of n elements of the subsystem without indicating certain defect element, and the diagnostics with the $\pi_i, i = \overline{1, n}$ DC can determine the defect arising only in i element of the subsystem. The architecture of such TS is shown in Fig. 4. In this case we can use approach proposed in [14].

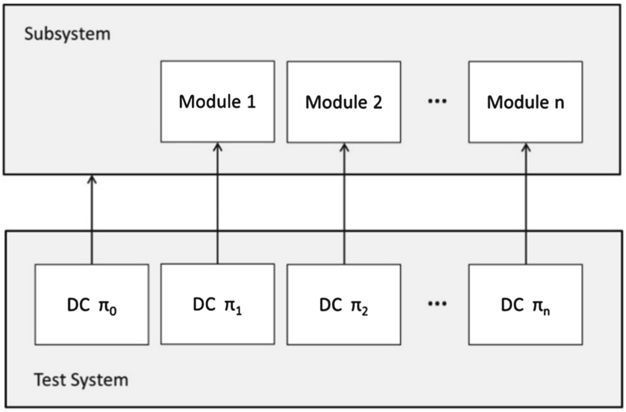


Fig. 4. Architecture of test system.

There are three alternative TS architecture for the examined subsystem:

1. TS with general control and diagnostic system (GCDS), which use only one DC π_0 for failures detecting of all n elements of subsystem. TS architecture is $\in \{\pi_0\}$.
2. TS with separate control and diagnostic system (SCDS), which use individual DC π_i for all $i = 1, \ldots, n$ elements of subsystem. The TS architecture is $U \in \{\pi_i\}, i = 1, \ldots, n$.
3. TS with mixed control and diagnostic system (MCDS), which use $m + 1$ DC, where $m < n$ individual DC are used for the identification of the technical condition of m elements, and the DC π_0 is used for the identification of the technical condition of the rest $n - m$ elements. The TS architecture is $U \in \{\pi_i\}, i = 0, 1, \ldots m, m < n$.

For any of above mentioned TS structures we can use common expressions for the cost of development and production of TS $C(U) = \sum_{i \in U} z_i$ and cost for the identification of failures in system with U architecture of TS during time of operation $tC_o(U, t) = us$.

Cost s for the diagnostics of one failure of the system depends on TS architecture. The detection of the failure on the level of element allows avoiding considerably more cost during the detection of the failure on the level of the subsystem in general. Practical experience of TS design shows that for modular constructions $c_i = c$ for all individual DC $\pi_i, i = 1, \ldots n$, and $c_0 = bc$ for DC π_0 for failures detecting of all n elements of subsystem. Empiric coefficient b characterizes the increase in the cost of failure diagnosis in the implementation of the TS at a higher level of structural hierarchy. In most cases this coefficient is $b \approx 10$ [15].

With taking into account the made remarks the expression for the determination of costs for the diagnostics of one failure in the process of operation of the examined subsystem could be submitted in the following form:

$$s = c[r + (1 - r)b],$$

where $r = \frac{\sum_{i=1}^{m} \lambda_i}{\Lambda}$ is a coefficient of completeness of subsystem diagnostics with the individual diagnostics channels.

In particular case for the structure of TS with general control and diagnostic system ($r = 0$) the value of $s = bc$, and for the structure of TS with separate control and diagnostic system ($r = 1$) the value of $s = c$.

For the exponential model of failures $u = \Lambda t$. In this case the expression (1) takes a form

$$C(U, t) = az_0 + \sum_{i=1}^{m} z_i + \Lambda tc[r + (1 - r)b], \tag{3}$$

where $a = 1$ for $m < n$ and $a = 0$ for $m = n$.

The graphs of $C(U, t)$ form a family of lines with an initial value equal to the cost of development and production of SKD chosen architecture: $C_1(U, t)$ – GCDS, $C_2(U, t)$ – SCDS, $C_3(U, t)$ – MCDS. The slope of the lines is determined by specific operating costs (Fig. 5).

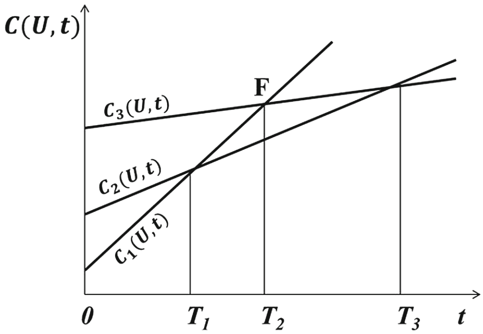


Fig. 5. Cost of life cycle with TS of different architectures.

If only GCDS and SCDS architectures of TS are used, their respective feasible application domains are determined by coordinates of the point F at Fig. 5. Point F is a crossing point of the straight lines $C_1(U, t)$ and $C_2(U, t)$. These lines coorespond to the cost of GSDS and SCDS functioning, respectively. Such coordinates may be determined solving the equation $C_1(U, t) = C_2(U, t)$, incerting into this equation expressions (3) for GCDS ($m = 0, r = 0$) and for SCDS ($m = n, r = 1$) structures of TS:

$$T_2 = \frac{\sum_{i=1}^{n} z_i - z_0}{\Lambda(b - 1)}. \tag{4}$$

In this case the interval $T_0 < T_2$ corresponds to the more efficient application of GCDS architecture of TS, whereas the interval $T_0 > T_2$ describes the more efficient application of SCDS structure.

In terms of cost-saving, MCDS structure application may become more efficient only at the condition that the straight line $C_3(U, t)$ reflecting the cost of MCDS functioning, will be passing below the point F; otherwise diapason t will be completely missing the intervals where MCDS structure would be seen as more efficient compared to GCDS or SCDS structures.

The use of MCDS could be economically more expedient only in case if the straight line $C_3(U, t)$, reflecting the cost of MCDS functioning, will be passing below the point F at Fig. 5. In the opposite case in all range of t the area where MCDS structure would be more efficient than GCDS or SCDS will be absent.

The pre-condition of MCDS application's cost-effectiveness can be estimated solving the inequality $C_3(U, T_{2)} < C_1\left(U, T_2\right)$, incerting into it formulas (3) and (4):

$$r \sum\nolimits_{i=1}^{n} z_i - \sum\nolimits_{i=1}^{m} > rz_0. \tag{5}$$

In the particular case of the identical DC with equal reliability of elements and with equal cost $\left(\lambda_i = \lambda, z_i = z, i = 1, \ldots n\right)$ the condition (5) will be changed into $z_0 < 0$. The last inequality is impossible for any permitted values of its variables. It leads to the conclusion that MCDS architecture of TS is not economically feasible, in the case if the elements are equally reliable and their diagnostic costs are identical.

While condition (5) is observed, the maximum economic feasibility will be demonstrated by GCDS for $T_0 < T_1$, MCDS for $T_1 < T_0 < T_3$, and SCDS for $T_0 > T_3$. The values of T_1 and T_3 for the boundaries of economically feasible MCDS architecture area are be determined from the equation $C_1\left(U, T_1\right) = C_3\left(U, T_1\right)$ and $C_2\left(U, T_3\right) = C_3\left(U, T_3\right)$:

$$T_1 = d^{-1} \sum\nolimits_{i=1}^{n} z_i, \; T_3 = d^{-1}(\sum\nolimits_{i=m+1}^{n} z_i - z_0),$$

where $d = c(b - 1) \sum_{i=m+1}^{n} \lambda_i$.

Using the expression (3) as an objective function (2), the problem of the synthesis of the optimal TS architecture U_{opt} on the examined k-level of structural representation for specified T_0 of life cycle could be solved by the known methods of integral optimization for $T_0 < T_2$ in the class of GCDS and MCDS architectures, and for $T_0 > T_2$ in the class of SCDS and MCDS architecture.

The optimal structure of TS for other subsystems of (k − 1)-level is performed in the same manner.

For hierarchical structure of system in general the problem of synthesis of optimal architecture of TS is led to the stepwise recurrent procedure, on each step of which the algorithm of TS architecture is realized for each of the hierarchy levels, starting with the lowest one.

Numerical example. Let us define the optimal TS architecture of the system with four subsystems, failure rate and costs for the creation of DC of which are submitted in the Table 1. The term of operation $T_0 = 10$ years, $c = 1, b = 10$.

Table 1. The data for the numerical example.

i	0	1	2	3	4
λ_i, 1/year	1	0.1	0.2	0.4	0.3
z_i, conv. unit	10	20	30	40	20

Using formula (4), we can determine that $T_2 = 11, 1$ years. Since $T_2 > T_0$, we should search for the optimal TS architecture within the GCDS and MCDS structures. For GCDS structure $U = \{\pi_0\}$, and in accordance with formula (3), $C_1(U, T_0) = 110$ conv. units. For MCDS architecture (using, for example, brute-force search method of optimization) it is possible to determine the optimal structure of diagnostic channels $U_{opt} = \{\pi_0, \pi_4\}$, which supplies the smallest cost of system operation $C(U_{opt}, T_0) = 103$ conv. unit. Thus, the developed TS should contain DC for the control of the whole system in general and DC of the control of the forth subsystem that supplies the smallest costs for the fulfilment it functions during the life cycle.

5 Conclusions

Modular construction is one of the popular method uses in shipbuilding industry. The modular ship construction requires better than normal quality assurance that must be provided with special product testing. Test-related information is generated throughout the product life cycle for each level in the system hierarchy. The present paper introduces a multi-level decision-making approach for design of optimal diagnostic architecture at the early stage of development that combines maintenance decisions at the k-levels of system and integration with test mechanisms for achieving efficient system level maintenance and lowering life-cycle cost of system operation. The proposed approach can be useful for the express analysis of the effectiveness of the decisions on the use of TS. This approach can be implemented in software or directly in hardware of integrated circuits. The effectiveness of such approach is investigated through the optimization of embedded architecture of diagnostics for known reliability and economic dependence during life cycle of the MSC.

References

1. Deschamps, L., Trumbule, J.: Cost Estimation, Ship Design and Construction. SNAME (2004)
2. Bertram, V., Maisonneuve, J., Caprace, J., Rigo P.: Cost Assessment in Ship Production. RINA (2005)
3. Caprace, J.D., Rigo, P.: Multi-criteria decision support for cost assessment techniques in shipbuilding industry. https://orbi.ulg.ac.be/bitstream/2268/9967/1/03_Caprace.pdf. (Accessed 09 June 2017)
4. Brown, A., Salcedo, J.: Multiple-objective optimization in naval ship design. Naval Eng. J. **115**(4), 49–62 (2003)
5. Landamore, M., Birmingham R., Downie, M.: Establishing the economic and environmental life cycle costs of marine systems: a case study from the recreational craft sector (2007)

6. Project SUSPRO – Decision support for sustainable ship production in global fluctuating markets. https://www.ntnu.no/suspro. (Accessed 09 June 2017)
7. Tan, W., Sauser, B., Ramirez-Marquez, J., Magnaye, R.: Multiobjective optimization in multifunction multicapability system development planning. IEEE Trans. Syst. Man Cybern. Syst. **43**(4), 785–800 (2013)
8. Zvaigzne, A., Bondarenko, O., Boiko, A.: Decision support system on the base of genetic algorithm for optimal design of a specialized maritime platform. Comput. Model. New Technol. **21**(2), 11–18 (2017)
9. Guerran, A., Jenssen, M.: Multi Criteria Decision Analysis in the Norwegian Maritime Sector. Norwegian University of Science and Technology (2014)
10. Azizie, M.: Shipbuilding technology modular construction and lean shipbuilding construction. https://ru.scribd.com/doc/50574298/Shipbuilding-Technology-Modular-Construction-and-Lean-Shipbuilding-Construction. (Accessed 09 June 2017)
11. Anuar, A.: Slipway and modular ship construction. https://ru.scribd.com/document/224405669/Slipway-and-modular-ship-construction. (Accessed 09 June 2017)
12. Deschamps, L. Extended modularization of ship design & build strategy. In: Shipbuilding Opportunities in Short Sea Shipping Workshop, 21–23 October, Charleston, USA (2008)
13. Zvaigzne, A., Pollaks, K., Pavlovics, A.: Multi-criteria decision making for oil pollution recovery module for swath multifunctional special ships. Int. J. Mod. Eng. Res. (IJMER) **7**(3), 41–49 (2017)
14. Kabashkin, I.: Design of embedded architecture for integrated diagnostics in avionics domain. Procedia Eng. **178**, 419–426 (2017)
15. Freise, A.: The value of easy-to-use products, pictures of the future, p. 65. Siemens (2003)

The Analytic Hierarchy Process (AHP): Prospects for Application in Supply Chain Management

Valery Lukinskiy(✉) and Vladislav Lukinskiy

National Research University Higher School of Economics (HSE), St. Petersburg, Russia
lukinskiy@mail.ru, vladas27@mail.ru

Abstract. The most expedient way, in our opinion, to generalize the numerous AHP studies is by the formatting of a morphological table and the corresponding block diagram. On the basis of the table, it is possible to take into account unobvious variants that can be missed with a simple search. Thus, the AHP, despite its obvious benefits, requires further research so as it can be successfully applied in the management of supply chains.

Keywords: Decision making · Analytic hierarchy process (AHP) Supply chain

1 Introduction

An improvement of the efficiency of supply chain management requires rationalization and optimization of all types of logistics functions and operations. One of these logistics functions is the choice of intermediaries (e.g., carriers, suppliers, freight forwarders, etc.). It should be noted that other tasks, such as the choice of the infrastructure object (warehouse), routes, mode of transportation, the location of the distribution center, are also related to this type of tasks.

Based on the analysis of the works [2, 7–9, 18], two approaches can be distinguished on the basis of which the logistics intermediary (LI) is selected:

- Analytical, implying the implementation of choice using models that include a number of parameters, which characterize LI;
- Expert, which is based on the assessments of an expert for the parameters characterizing LI, and describes the procedures for obtaining integrated expert assessments (ratings).

In turn, the expert approach includes at least three main methods: a ball-rating evaluation; the analytic hierarchy process (AHP) and the general mediator selection algorithm.

Let us consider the AHP in more details. It assumes application of the following hierarchical structure: goal-criteria-alternatives.

Table 1 presents the publications of various authors and their brief characteristics. Table 2 is formed from scientific works, authored by T. Saaty or co-authored with a team of researchers.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_27

Table 1. Systematization of literature employed AHP.

Source	Evaluation of AHP
[4]	Successful AHP applications have been reported in marketing, finance, education, public policy, economics, medicine, and sports. AHP overcomes other decision-making methods in many ways. It is a method with large penetration both in academic and professional environment and is implemented by business tools widely tested and validated
[3]	This methodology has been already used in healthcare, with different goals and objectives, such as user needs elicitation, medical decision-making, budget allocation and medical device purchasing. It is possible to assess the coherence of respondent judgments and ask the experts to refine incoherent answers
[1]	The elements of the hierarchy can relate to any aspect of the decision problem – tangible or intangible, carefully measured or roughly estimated, well- or poorly-understood – anything at all that applies to the decision at hand
[19]	This approach appears especially useful when effects cannot be fully monetized, nor even quantified
[6]	AHP was first proposed by the famous American mathematician named "Thomas L-Satty" who sought to provide an appropriate strategy for the ordinary people to make decisions about complex issues that several factors were involved in and his findings was known as the "AHP method"

Table 2. Systematization of scientific works authored or co-authored by T. Saaty.

Source	Considered topic
[17]	This paper is concerned with understanding synthesis of electric signals in the neural system
[14]	Choosing a city to live in using ratings
[13]	This paper has the potential to dramatically change the process of resource allocation in economics
[15]	We have identified some key shortcomings of traditional majority voting
[12]	Criteria used to buy a house for a family
[11]	Judgments are needed in medical diagnosis to determine what tests to perform given certain symptoms

The AHP has been used in various settings to make decisions. The Department of Defense in the US uses it frequently and extensively to allocate their resources to their diverse activities. British Airways used it in 1998 to choose the entertainment system vendor for its entire fleet of airplanes. Xerox Corporation has used the AHP to allocate close to a billion dollars to its research projects. In 1999, the Ford Motor Company used the AHP to establish priorities for criteria that improve customer satisfaction. IBM used the process in 1991 in designing its successful mid-range AS 400 computer [16].

Thus, presented in Tables 1 and 2 systematized results of the analysis of publications allow us to draw the following conclusions:

- It is hard to find areas of human activity for which the AHP method would not be used;

- The authors, on the whole, positively characterize the possibilities of using this approach and, moreover, note its superiority over other approaches;

In most of the reviewed works, the authors do not propose discussing the questionable points related to the application of the AHP. For example, the need to examine the opinions of experts in the case of exceeding the established reference values.

2 Morphological Features and Basic Operations of the Method of Analytic Hierarchy Process

The most expedient way, in our opinion, to generalize the numerous AHP studies is by the formatting of a morphological table and the corresponding block diagram. It is known that the morphological analysis provides for the division of the problem into subsystems (characteristic features of P_i) and elements (alternatives to the realization of the features of A_{ij}). On the basis of the table, it is possible to investigate the significant (conceivable) number of variants, resulting from the construction of the AHP, which will allow, in turn, to take into account unobvious variants that can be missed with a simple search.

The total number of possible M variants, which form a morphological set can be determined by the formula

$$M = \prod_{i=1}^{L} k_i = k_1 \cdot k_2, \ldots, k_L, \quad (1)$$

where k_i – number of ways to realize the i-th attribute.

Table 3 shows the morphological features of the analytic hierarchy process (fragment). According to the formula (1), the maximum number of variants taking into account the different features of the AHP formation is 13824.

To realize the advantages of morphological analysis, it is advisable to supplement it with a block diagram that allows to concentrate the directions of the search for the most promising combinations of features and alternatives (Fig. 1). According to the block diagram, the main features were as follows: the choice of the scale ($A_{1i} = 4$); selection of the processing method ($A_{3i} = 4$) with the subsequent definition of the priority vector, etc.

Despite the optimistic tone of the publications, in which the AHP was highly appreciated, a number of questions remain open.

First, the uniqueness of the restriction of 0.1 is questionable. In particular, [10] indicates that the condition of $C.R. \leq 0.20$ is allowed, but not more. At the same time, [5] states that ‘some circumstances may allow higher values of CR, even up to 0.3’.

Secondly, why is the restriction value 0.1 (or some other), which is calculated and formalized using a randomly generated matrix, defines the boundary for experts with intelligence, intuition and experience?

Table 3. Morphological features of the hierarchy analysis method (fragment).

Subsystems (features) P_i	Alternatives for the implementation of the characteristic A_i	Number of ways to realize K_i
Scale (P_1)	A_{11} – any A_{12} – the most popular (1–9) A_{13} – alternative accounting (1–7, 1–5) A_{14} – the simplest (0–1, yes/no, 0–2)	4
The object (the conditions for constructing the matrix) (P_2)	A_{21} – tangible (quantitative) – certainty A_{22} – intangible (qualitative) – full uncertainty A_{23} – mixed (quantitative/qualitative) – partial uncertainty (probabilistic description of the object)	3
Methods of information processing (P_3)	A_{31} – rationing by the column A_{32} – rationing by rows A_{33} – double rationing (in columns and in rows) A_{34} – the geometric mean of the row	4
The consistency criterion (C.R.) (P_4)	A_{41} – without consistency A_{42} – consistency with a limited scale (0.05–0.14) A_{43} – consistency with the extended scale (up to 0.3)	3
Verification capability (P_5)	A_{51} – for physically measured quantities (analytical dependencies) A_{52} – indirect verification (by analogy) A_{53} – there is no possibility of verification A_{54} – based on criteria (e.g. statistics)	4
The random matrix consistency index (R.I.) (P_6)	A_{61} – based on generation of matrices [10] A_{62} – calculation of the value by the analytical dependence [18]	2
Criteria for verification (P_7)	A_{71} – the eigenvalue of the matrix A_{72} – the consistency index (C.I.) A_{73} – consistency ratio (C.R.)	3

Thus, the AHP, despite its obvious benefits, requires further research so as it can be successfully applied in the management of supply chains.

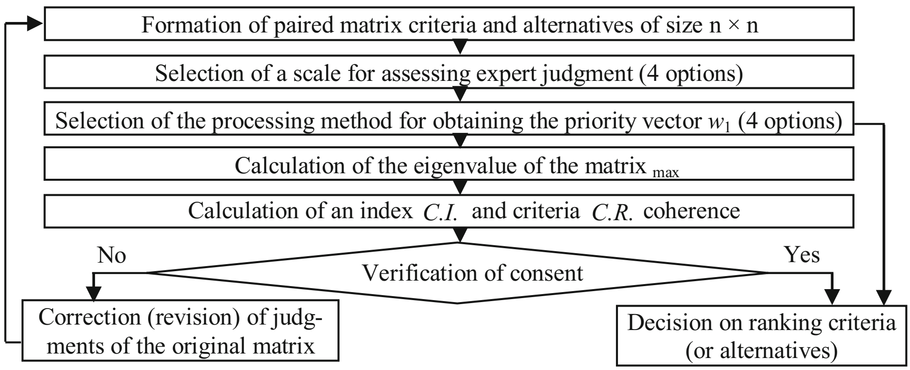


Fig. 1. Generalized block diagram of the basic operations of AHP.

3 Conclusions

The analysis of trends in the development of logistics showed that among a large number of problems in managing supply chains, one of the most important is the issue of decision-making, in particular, the choice of logistics intermediaries.

The generalization of a number of sources devoted to the AHP allowed the formation of a morphological table. According to the table, the total number of possible variants (the so-called ‘morphological set’) is $M = 13824$.

The developed block diagram of decision-making on the basis of AHP makes it possible to carry out a purposeful search for the most promising combinations of features and alternatives that are necessary for improvement and further development of the AHP.

References

1. Beria, P., Maltese, I., Mariotti, I.: Multicriteria versus cost benefit analysis: a comparative perspective in the assessment of sustainable mobility. Eur. Transp. Res. Rev. **4**, 137–152 (2012)
2. Bowersox, D.J., Closs, D.J.: Logistical Management: The Integrated Supply Chain Process. McGraw-Hill Companies Inc., New York (1996)
3. Cancela, J., Fico, G., Waldmeyer, M.: Using the analytic hierarchy process (AHP) to understand the most important factors to design and evaluate a telehealth system for Parkinson’s disease. BMC Med. Inform. Decis. Mak. **15**(Suppl 3), S7 (2015)
4. Dulange, S.R., Pundir, A.K., Ganapathy, L.: Prioritization of factors impacting on performance of power looms using AHP. J. Ind. Eng Int. **10**, 217–227 (2014)
5. Gogas, M., Papoutsis, K., Nathanail, E.: Optimization of decision-making in port logistics terminals: using analytic hierarchy process for the case of port of Thessaloniki. Transp. Telecommun. **15**(4), 255–268 (2014). https://doi.org/10.2478/ttj-2014-0022
6. Hamidy, N., Alipur, H., Nasab, S., Yazdani, A., Shojaei, S.: Spatial evaluation of appropriate areas to collect runoff using analytic hierarchy process (AHP) and geographical information system (GIS). Model. Earth Syst. Environ. **2**, 172 (2016)
7. Kabashkin, I., Lučina, J.: Development of the model of decision support for alternative choice in the transportation transit system. Transp. Telecommun. **16**(1), 61–72 (2015). https://doi.org/10.1515/ttj-2015-0007
8. Lukinskiy, V.S., Lukinskiy, V.V.: Analysis of the logistics intermediaries choice methods in the supply chains. Transp. Telecommun. **16**(4), 288–295 (2015). https://doi.org/10.1515/ttj-2015-0026
9. Saaty, T.L.: The Analytical Hierarchy Process. RWS Publications, Pittsburgh (1990)
10. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process. RWS Publications, Pittsburgh (1996)
11. Saaty, T., Vargas, L.: Diagnosis with dependent symptoms: Bayes theorem and the analytic hierarchy process. Oper. Res. **46**(4), 491–502 (1998)
12. Saaty, T.: Decision-making with the AHP: why is the principal eigenvector necessary. Eur. J. Oper. Res. **145**, 85–91 (2003)
13. Saaty, T., Vargas, L., Dellmann, K.: The allocation of intangible resources: the analytic hierarchy process and linear programming. Soc. Econ. Plan. Sci. **37**, 169–184 (2003)

14. Saaty, T.: Rank from comparisons and from ratings in the analytic hierarchy/network processes. Eur. J. Oper. Res. **168**, 557–570 (2006)
15. Saaty, T., Shang, J.: Group decision-making: head-count versus intensity of preference. Soc. Econ. Plan. Sci. **41**, 22–37 (2007)
16. Saaty, T.: Decision making with the analytic hierarchy process. Int. J. Serv. Sci. **1**(1), 83–98 (2008)
17. Saaty, T.: Neurons the decision makers, Part I: the firing function of a single neuron. Neural Netw. **86**, 102–114 (2017)
18. Taha, H.: Operations Research: An Introduction, 9th edn. Prentice Hall, Upper Saddle River (2011)
19. Wiethoff, M., Brookhuis, K., Waard, D., Marchau, V., Walta, L., Wenzel, G., Brucker, K., Macharis, K.: A methodology for improving road safety by novel infrastructural and invehicle technology combinations. Eur. Transp. Res. Rev. **4**, 67–77 (2012)

Methodology for Assessment of Electronic Payment Systems in Transport Using AHP Method

Ivana Olivková(✉)

Faculty of Mechanical Engineering, VŠB – Technical University of Ostrava, 17. Listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
ivana.olivkova@vsb.cz

Abstract. The article deals with the choice of electronic payment systems in public transport. In this case, one of the expert methods of multicriteria evaluation of variants must be used. The Analytical Hierarchy Process method [1] was selected for evaluating the variants of electronic payment systems. This method is currently one of the most frequently used multicriteria assessment methods thanks its completeness, simplicity and wide range of uses. The method AHP uses the decomposition of a complex unstructured situation into individual criteria, which are arranged in a hierarchical structure. This method was used to determine the weights of criteria that were set based on both carrier and passenger requirements. Paired comparison method was used at each level of the hierarchical structure to compare each criterion with the other criteria. The result of the comparison is the weight of the individual criteria. In the next subchapters of the article partial evaluation of variants of electronic payment systems was made, the synthesis of these evaluations was presented and the most suitable variant was chosen. Finally, the advantages and disadvantages of the AHP method and the recommendations regarding its use in practice are mentioned.

Keywords: Electronic payment system · Public transport
Multicriteria evaluation of variants · AHP method

1 Introduction

There are several possible ways and the multi criteria decision-making methods. Their use depends on the type, completeness and on level of detail of available information. In the transport practice there are often cases when it is necessary to decide on the variants of fundamental importance the earliest opportunity but the availability, quantity and quality of information for decision usually does not correspond to their real needs. For solving such type of the task are usually employed methods based on paired comparison of variants [2]. The common feature of this group of methods of multicriteria evaluation is that the basic information to determine of preferential arrangement of variants makes up results of paired comparisons of these variants with respect to the individual criteria [3].

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_28

2 Methodology

In this paper is described the possibility of using of the Analytic Hierarchy Process (AHP) method for preferential arrangement of variants [1]. The AHP serves as a mathematical solution method for individual or group decision-making with multiple criteria. With the AHP the decision-maker design the hierarchy structure. Then the decision-maker develops priorities for alternatives and criteria used judgments or comparisons on each pair of elements. Finally, these priorities are expressed as weights. The selection of the best variant is then based on the synthesis of the weights throughout the hierarchy.

2.1 Analytical Hierarchy Process

In solving decision-making problems, it is necessary to take into account all the elements that affect the outcome of the analysis and the relationship between them. The decision problem can be represented as a hierarchical structure. It is a linear structure containing s-levels, and each of these levels includes several elements. Layout of individual levels is always from general to specific. For the general role of multicriteria evaluation of variants, the hierarchy may be as follows (see Fig. 1):

Level 1 – the goal of evaluation, which may be the arrangement of variants,
Level 2 – the experts who are involved in the evaluation,
Level 3 – the criteria of evaluation,
Level 4 – the variants of evaluation.

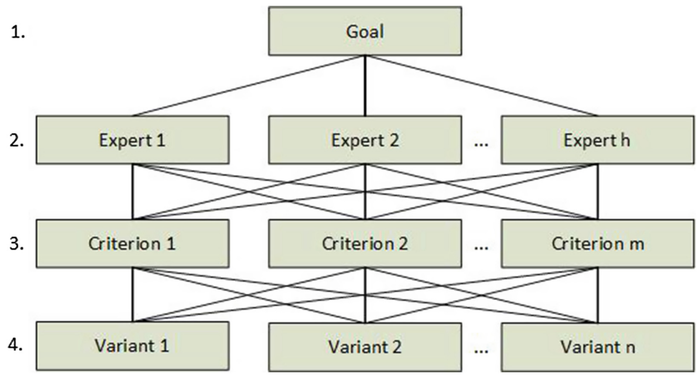


Fig. 1. Hierarchy of multicriteria evaluation of variants.

The links in this hierarchy are evident from the Fig. 1. The figure shows a four-level hierarchy-one goal, *h* experts, *m* criteria, and *n* variants. On the second level of the hierarchy there is one matrix of paired comparison with dimensions $h*h$. On the third level there are *h* matrices with dimensions *m x m* and the fourth level there are m matrices with dimensions $n \times n$. Using calculations (scoring method for calculating criteria weights) in these matrixes, the variants "divide" the value of the weight of the criterion (the criteria then "divide" the weight of the relevant expert). The calculated values are called preferential indexes of variants for each criterion. If the preferential indexes are summarized for all criteria, the variants can be evaluated from the point of view of all the experts and from the point of view of all the criteria.

The Method of multicriteria evaluation of variants by hierarchical depiction was proposed by prof. Saaty in 1980 [7]. It is called Analytical Hierarchy Process (AHP) method. The AHP method is based on a hierarchy defined by the m-criterion and the n-variant.

The fundamental question of the AHP method is how users can determine the intensity of the relationships between the elements of the hierarchy. The AHP method offers a relatively accessible procedure for the user. The user don't have to enter absolute values of weight of experts, weight of criteria, preference indexes of variants. The user only gives information about the relationship between each criterion and evaluates variants through these criteria.

In the case that there are the needed information in the detailed structure and the evaluation are carried out by experts, can be objectified their evaluation. For solving such type of the task can be employed methods based on the thresholds of sensitivity [1]. From this group of methods have been chosen method of approximation of fuzzy relations. This method is, due to its nature, suitable for evaluating variants with a set of qualitative criteria [2].

2.2 Method of Approximation of Fuzzy Relations

For solving the problem was chosen multicriteria discrete decision-making model in which individual variations of fare collection systems are described explicitly by the list of variants assessed according to various criteria. The task of multicriteria decision means the following problem:

Is given set of variants $X = \{x_1, x_2, \ldots, x_n\}$ which are evaluated according to the set criteria $K = \{K_1, K_2, \ldots, K_m\}$. The task is to select from a given set of variants X the variant x^*, which is the best with respect to criteria from the set K. If the elements of the criterial matrix Y are designated y_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, may be criterial matrix written in the following form:

$$Y = \begin{array}{c} \\ x_1 \\ \vdots \\ x_n \end{array} \begin{array}{c} \begin{array}{ccc} K_1 & \cdots & K_m \end{array} \\ \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nm} \end{bmatrix} \end{array}. \tag{1}$$

To determine the optimal variant x*∈ X, it is sufficient to arrange the variants from X on the basis of their overall evaluation with respect to the criteria from K. The variation that occupies the first place in this arrangement is then the optimal variant [2].

Let's assume that is set the relative importance of the criteria in the form of weight:

$$v = (v_1, v_2, \ldots, v_m), \sum_{i=1}^{m} v_i = 1, v_i \geq 0, \tag{2}$$

where $v_1, v_2, \ldots, v_m$ are weights of criteria.

The objective of methods based on the thresholds of sensitivity is creation of matrix of preference variants V. For every pair of variants x_j and x_k is grouped the criteria that:

1. Prefer the variant x_j before the variant x_k, a set of their indexes is defined I_{jk};
2. Prefer the variant x_k before the variant x_j, a set of their indexes is defined I_{kj};
3. Have for both variants x_j and x_k the same values, and in terms of these criteria both variants are indifferent, a set of their indexes is defined $I_{j\sim k}$;
4. Have the variant x_j incomparable with the variant x_k, a set of their indexes is defined $I_{j?k}$.

The degree of preference variant x_j before the variant x_k is defined:

$$V_{jk} = \sum_{i \in I_{jk}} v_i. \tag{3}$$

The degree of preference variant x_k before the variant x_j is defined:

$$V_{kj} = \sum_{i \in I_{kj}} v_i. \tag{4}$$

The degree of indifference variant x_j and x_k is defined:

$$V_{j\sim k} = \sum_{i \in I_{j\sim k}} v_i. \tag{5}$$

The degree of incomparability variant x_j and x_k is defined:

$$V_{j?k} = \sum_{i \in I_{j?k}} v_i. \tag{6}$$

From the sums of the degree of preference is formed matrix of preferences of variants $V = [V_{jk}]$.

Method of approximation of fuzzy relations is based on the idea of sorting a decision variant from the best to the worst according to decreasing row sums of the V^* matrix [2]. Practical procedure begins by modifying the matrix of preferences of variants V using one of the following relations:

$$V_{jk}^* = V_{jk} + \frac{1}{2}\left(V_{j\sim k} + V_{j?\mathrm{k}}\right), \tag{7}$$

$$V_{jk}^* = \frac{1}{2} + \frac{1}{2}\left(V_{jk} - V_{kj}\right). \tag{8}$$

Thus calculated matrix of strict preference is symbolized $V^* = \left[V_{jk}^*\right]$. This step ensures strict asymmetry. The row sums $S_j = \sum_k V_{jk}^*$ of the V^* matrix are calculated.

According of decreasing values of characteristic S_j are gradually rearranged both rows and columns of the matrix V^*. Thus calculated matrix is symbolized by $T = [T_{jk}]$. Then it is necessary to determine whether the matrix T fulfill the condition of transitivity, i.e. if for all (j, k, l), where $j > k > l$ is valid $T_{ji} \geq \max\left(T_{jk}, T_{kl}\right)$. If this is not the case, the matrix T must be approximated by matrix W.

$$W = \frac{1}{2}\left(W^1 + W^2\right), \tag{9}$$

where

$$W^1 = \left[W^1_{jk}\right], \quad W^1_{jk} = \max_{\substack{p \geq j \\ q \leq k}} T_{pq} \quad (j \leq k), \tag{10}$$

$$W^1_{kj} = 1 - W^1_{jk}, \tag{11}$$

$$W^1_{jj} = 0.5, \tag{12}$$

$$W^2 = \left[W^2_{jk}\right], \quad W^2_{jk} = \min_{\substack{p \leq j \\ q \geq k}} T_{pq} \quad (j \leq k), \tag{13}$$

$$W^2_{kj} = 1 - W^2_{jk}, \tag{14}$$

$$W^2_{jj} = 0.5. \tag{15}$$

With the resulting matrix W is also obtained resulting arrangement of variants according to the preference from best to worst on the basis of decreasing values of characteristic U_j.

3 The Application AHP Method

Decision-making method based on hierarchy structure is applied to select the most suitable fare collection systems in public transport. The possibilities of modern technology in the fare collection systems in public transport are discussed for example in publications [4–6]. Based on these sources are identified following five variants of fare collection systems:

- Check in – is based on the principle of login of the passenger through smart cards.
- Check in/Check out – is based on the principle of login and logout of the passenger through smart cards.
- Be-In/Be-out – is system eliminating any further manipulation of the media when getting in and out; the presence of a passenger in the vehicle is detected with Be-In/Be-Out system through a specific personal smart cards or mobile phone with an chip that is active for connection.

- SMS ticket – is a specific type of electronic fare collection system that allows electronic ordering of tickets through SMS sent from a mobile phone.
- NFC (Near Field Communication) – is a new technology that combines smart card (RFID) and mobile phone.

Aim is to find a variant that achieves the best evaluation of all criteria, eventually these variants appropriately arrange.

This problem is solved using the AHP method. Hierarchy is presented in Fig. 2.

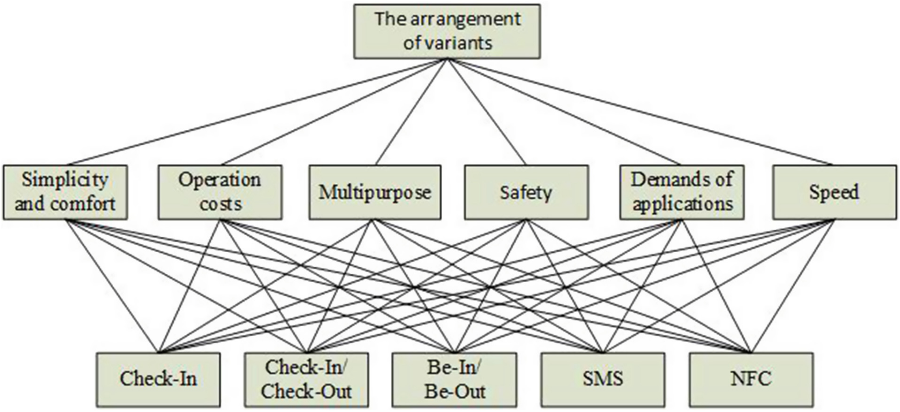


Fig. 2. Hierarchical structure of the problem.

On the first level of the hierarchy is defined the goal of evaluation – the arrangement of variants. This objective depends primarily on the evaluation criteria used. Therefore, the criteria in the hierarchy are immediately on the next, second level. Importance of each criterion is evaluated through the weights of the criteria. At the last level of the hierarchy are showed the evaluated variants. Their arrangement depends on their relation to the evaluation criteria.

The data for the evaluation was obtained by traffic of selected group of evaluators (experts on traffic issues). Were interviewed 128 evaluators from January to April 2017. The evaluators were informed about the variants of fare collection systems and based on their expertise they evaluated variants through criteria.

Table 1 shows the scores of individual variants of fare collection systems according to selected criteria. Rating scale is in the range from 0 to 1 b. The higher is the number, the evaluator better evaluate the variant. Table 1 is the default matrix for solving. There are average values (benefits) of evaluation criteria assigned to the assessed variants.

The second step is to determine the weight of each evaluation criteria. Weights of evaluation criteria reflect their importance numerically; the importance from the point of view of the evaluators. For the determination of weight the Eq. (2) is applied.

Table 1. Matrix Y.

Variants	Criteria					
	Simplicity and comfort	Operation costs	Multi-purpose	Safety	Demands of applications	Speed
Check-In	0.68	0.67	0.59	0.48	0.63	0.76
Check-In/Check-Out	0.62	0.65	0.55	0.43	0.61	0.51
Be-In/Be-Out	0.89	0.51	0.41	0.57	0.42	0.98
SMS	0.72	0.46	0.43	0.67	0.81	0.43
NFC	0.78	0.61	0.92	0.75	0.84	0.81

The weights were calculated on the basis of scoring method that evaluators determined through a questionnaire. Rating scale is in the range from 1 to 10. Due to the requirements of comparability weights to the criteria established by different methods these weights should be normalized (the sum of the standardized set of criteria weights is equal to one). The standardization of criteria weights is done according to:

$$v_i = \frac{b_i}{\sum_{i=1}^{m} b_i}, \qquad (16)$$

where v_i is weight of criteria and b_i is score of criteria.

In the Table 2 there are the average scores and average values of the weights of individual criteria calculated from all the evaluators' values.

Table 2. Weights of criteria.

Calculation	Criteria					
	Simplicity and comfort	Operation costs	Multi-purpose	Safety	Demands of applications	Speed
Score	4.8	4.6	7.8	8.6	5.6	3.6
Weight	0.14	0.13	0.22	0.25	0.16	0.10

In step 3 it is possible to create matrix of preference variants *V*. The procedure is described in Sect. 2.2. From the sums of the degree of preference V_{jk} and V_{kj} is formed matrix of preferences of variants $V = [V_{jk}]$ which is a square with zeros on the main diagonal (Table 3).

Table 3. Matrix of preferences of variants *V*.

Variants	Check-In	Check-In/Check-Out	Be-In/Be-Out	SMS	NFC
Check-In	0	1	0.51	0.45	0.13
Check-In/ Check-Out	0	0	0.51	0.45	0.13
Be-In/ Be-Out	0.49	0.49	0	0.37	0.24
SMS	0.55	0.55	0.63	0	0
NFC	0.87	0.87	0.76	1	0

The initial step of the method of approximation of fuzzy relations is the creation of matrix of strict preference V^* from matrix V and calculation values of row sums S_j. The procedure is described in Sect. 2.2 All is indicated in the Table 4.

Table 4. Matrix V^*.

Variants	Check-In	Check-In/ Check-Out	Be-In/Be-Out	SMS	NFC	S_j
Check-In	0.5	1	0.51	0.45	0.13	2.59
Check-In/ Check-Out	0	0.5	0.51	0.45	0.13	1.59
Be-In/ Be-Out	0.49	0.49	0.5	0.37	0.24	2.09
SMS	0.55	0.55	0.63	0.5	0	2.23
NFC	0.87	0.87	0.76	1	0.5	4

According of decreasing values of row sums S_j are gradually rearranged both rows and columns of the matrix V^*. Thus calculated matrix is symbolized T (Table 5).

Table 5. Matrix T.

Variants	NFC	Check-In	SMS	Be-In/Be-Out	Check-In/Check-Out
NFC	0.5	0.87	1	0.76	0.87
Check-In	0.13	0.5	0.45	0.51	1
SMS	0	0.55	0.5	0.63	0.55
Be-In/Be-Out	0.24	0.49	0.37	0.5	0.49
Check-In/Check-Out	0.13	0	0.45	0.51	0.5

Since the matrix T does not fulfil the condition of transitivity. it must be approximated by the matrix W.

Matrix W was created based on the arithmetic mean of the W^1 and W^2 matrixes. created according to the relations in Sect. 2.2. The matrix W is shown in Table 6 below. including row sums U_j.

Table 6. Matrix W.

Variants	NFC	Check-In	SMS	Be-In/Be-Out	Check-In/ Check-Out	U_j
NFC	0.5	0.82	0.88	0.88	0.94	4.01
Check-In	0.19	0.5	0.50	0.57	0.94	2.69
SMS	0.12	0.50	0.5	0.57	0.59	2.28
Be-In/Be-Out	0.12	0.43	0.43	0.5	0.50	1.98
Check-In/Check-Out	0.07	0.07	0.41	0.50	0.5	1.54

Finally, variants from the best to worst are arranged according to values of row sums. The results of the preferential arrangement of the variants obtained using the method of approximation of fuzzy relations is shown in Table 7.

Table 7. The results of the preferential arrangement of the variants.

Order	Approximation of fuzzy relations
1	NFC
2	Check-In
3	SMS
4	Be-In/Be-Out
5	Check-In/Check-Out

From the resulting arrangement of variants is evident that the best variant of the payment system is evaluated NFC technology. The next in the order is a variant of the Check-In and third is the variant of SMS tickets.

4 Advantages and Disadvantages of AHP Method

The AHP method has its advantages and disadvantages as well as other multicriteria decision-making methods.

Advantages of AHP method:

1. Simplicity for decision-makers (users) – compared with other methods of multicriteria evaluation. This method is not difficult for decision-makers.
2. Simplicity for solvers (mathematics) – the calculation of partial weights of the criteria and partial and total evaluation of variants is not complicated; there are also programs created directly to calculate using the AHP method.
3. Creating hierarchies – due to the division of the decision-making problem into the hierarchy is the decision-making problem clearer; the hierarchical structure allows us to get a comprehensive view of the whole issue and so we can tell whether we have included all the important aspects (criteria) in the model.
4. The method accepts different types of criteria – the AHP method allows to combine the criteria whose values are given objectively and the criteria whose values are defined subjectively by decision-maker.
5. Wide application capabilities – this method has a very wide range of applications; it can be used to select the best variant from a given set. to arrange variants and to evaluate a set of variants.

Disadvantage of AHP is namely number of pair comparisons. We have to make a large number of pair comparisons. We need to compare both the criteria and all the variants in relation to each criterion. A large number of pair comparisons can be time consuming to calculation. The decision-maker may lose attention and make inconsistent decision in the calculation.

5 Conclusions

This article deals with the application of the multicriteria decision making methods in specific decision situation in the public transport. Decision problem with n criteria and *m* variants can be graphically represented as a hierarchical structure. This representation is equivalent to decision matrix and can be effectively used for the decision-making with more criteria.

When we create this form of hierarchical structure we can choose the best variant using AHP method. The main aim of this paper is to use this method on large datasets and verify the accuracy of the method based on method of approximation of fuzzy relations. This method compare variants according to the preferences from best to worst based on the values of characteristics that determines before how many variants is each variant preferred.

Using this method is evaluated the best variant of fare collection systems in public transport – NFC system. Check-In and SMS variants of fare collection systems in public transport are suitable too. The decision-maker could choose one of these variants as an alternative to the NFC variant. But variants Be-In/Be-Out and Check-In/Check-Out appear to be less suitable and in making further decisions would no longer be taken into account.

The AHP needs pair wise comparisons as input information in any form. the decision tree with quantitative as well as qualitative data on weights. evaluation of variants can be solved by AHP method.

Advantage of this approach is the possibility of solving of the decision problems:

– with multiple criteria.
– with more decision variants and
– with different type of all kind of input data.

Acknowledgements. This paper was supported with the internal research project SP2017/138 “Research in the field of optimization of transport system operating technologies”.

References

1. Fiala, P., Jablonský, J., Maňas, M.: Vícekriteriální rozhodování. VŠE Praha (1994)
2. Fotr, J., Píšek, M.: Exaktní metody ekonomického rozhodování. Academia, Praha (1986)
3. Fotr, J., Švecová, L.: Manažerské rozhodování: postupy, metody a nástroje. Praha: Ekopress. Praha (2010)
4. Olivková, I.: Comparison and evaluation of fare collection technologies in the public transport. Procedia Eng. **178**, 515–525 (2017)
5. Olivková, I.: Metodika výběru elektronického platebního systému pro platbu ve veřejné dopravě. Certified methodology (2012)
6. Olivková, I.: Metodika pro výběr nejvhodnější moderní telematické technologie. Certified methodology (2012)
7. Saaty, T.L.: The Analytic Hierarchy Process. Mc Graw Hill Inc., New York (1980)

Intellectualization of the Spare Parts Supplier Selection by the Analysis of Multi-criterial Solutions

Irina Makarova, Ksenia Shubenkova(✉), Polina Buyvol, and Eduard Mukhametdinov

Kazan Federal University, Syuyumbike prosp., 10a, 423822 Naberezhnye Chelny, Russia
kamIVM@mail.ru, ksenia.shubenkova@gmail.com

Abstract. The quality of automobiles efficiency maintaining is largely determined by the effectiveness of providing system of dealer and service centers with spare parts. It is necessary to analyze the characteristics of suppliers and the distribution of flow requests for spare parts, taking into account reliability factor of the supplier by the control center of the firm service system. The authors proposed a methodology for multi-criteria evaluation of the spare parts suppliers, which is determined not only by the reliability of compliance with the terms of supply and the organization of logistics processes, as well as the reliability and quality of the spare parts themselves. Also, in order to assess the probability of fulfillment of obligations by the new supplier, a methodology for forecasting such indicators characterizing the supplier's reliability, as the delivery time, the spare part quality and the level of shortages, is described with usage of logistic regression.

Keywords: Supplier · Estimate · Reliability · Super-criterion

1 Introduction

The growth of motorization and intensifying competition in the automotive market force automotive manufacturers not only to improve the quality of their products, but also to seek new ways to attract customers. In this sense an important role is played by the car buyer's confidence in the possibility of its trouble-free operation throughout the life cycle. It is especially important to provide quality service during the warranty period. Any deviation from the warranty can adversely affect the producer reputation and reduce the brand's trust in clients – customers and owners. The quality of automobiles efficiency maintaining is largely determined by the effectiveness of providing system of dealer and service centers with spare parts. To minimize risks in the logistics system associated with late or substandard delivery of spare parts it is necessary to analyze the characteristics of suppliers and the distribution of flow requests for spare parts, taking into account the received reliability factor of the supplier by the control center of the firm service system. Therefore, the objective of this study was developing a methodology for assessing the existing and valid selection of potential suppliers.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_29

2 Existing Methods for Evaluating and Selecting Suppliers

Vehicles' reliability is determined by the quality and reliability of its components, assemblies and systems. Therefore selecting of suppliers, in the context of economic globalization, with assembly factories emerging in different countries, is by no means unimportant. One of the ways to implement rational management is usage of decision support systems for the intellectualization of the best supplier selection by the analysis of multi-criterial solutions. When evaluating the supplier's reliability, many of the authors highlight the factors coming to the forefront when considering alternatives in decision making. There are different methods proposed for multi-criteria analysis. It should be noted among the most frequently used multi-criteria approaches to the decision making, proposed for the selection of suppliers, the TOPSIS method, Dempster-Shafer evidence theory [1], the analytical hierarchy process (AHP) [2], fuzzy sets theory [3, 4], data envelopment analysis (DEA) [5, 6], goal programming, multi-objective programming, the liner programming, mixed integer programming [7].

These methods have been improved in later studies. Thus, the review [8] reports on work using analytical and empirical methods for the selection of strategic suppliers. In paper [9, 10] the QFD-AHP method is used to evaluate the effectiveness of the strategic supplier. The disadvantage of this method is the mandatory individual approach for its implementation in a particular industry.

Despite the diversity of approaches, all these methods are based on expert assessments regarding the choice of priorities. The main drawback of these approaches is the element of subjectivism, which is due, first of all, to the peculiarities of qualitative methods and, secondly, to the personal characteristics of the experts themselves. The reliability of the supplier should be the main factor in its choice, although when evaluating the company's business operations through multidimensional analysis, this is only one of the areas that affect the effectiveness of its activities.

Considering alternatives when making decisions to assess supplier reliability, many authors identify the factors that are considered most significant in each case. In particular, studies [5, 11] presented a comprehensive approach that takes into account the factor of negative impact on the environment when choosing a supplier.

In our opinion, further research should be aimed at testing the influence of specific factors affecting the choice of the spare parts supplier for automotive vehicle, as well as the development of dynamic assessment methods.

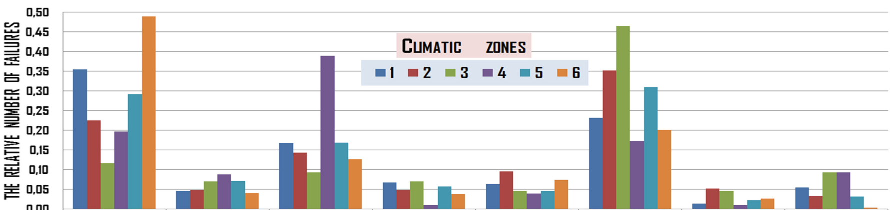
For manufacturers of vehicles, the reliability of suppliers should be determined not only by the reliability of the supply itself, i.e. compliance with its conditions and the speed of the organization of logistics processes, but also the reliability and quality of the spare parts themselves. In connection with the foregoing, in this study, as a priority criterion, one of the indicators of reliability (namely, failure-free operation, was added as a priority criterion in the evaluation of suppliers, since it is the most important during the guarantee period). To identify the likelihood that the supplier will be reliable, a multiple logistic regression model is used.

3 Intellectualization of Suppliers Selection Process: Opportunities and Prospects

3.1 Algorithm for Supplier Selecting by the Criterion of Spare Part Reliability

The effectiveness of technical systems (TS) operation depends to a large extent on the reliability of both the individual devices in the systems and the elements that ensure interaction between these devices. Despite considerable efforts in improving the TS reliability, the level of their reliability remains insufficiently high and does not satisfy ever increasing demands. Insufficient reliability of devices and their components leads to significant downtime of systems, as well as increase in operating costs. In addition, TC failures can lead to emergency situations, the consequences of which can be significant. The parts and units limiting the reliability of the car aggregates are those that fail at least 50% of the total number of failures, and the costs for eliminating these failures (for spare parts and replacement work) are not less than 70% of the total cost [12]. At the same time, it is necessary to take into account that different units, aggregates and systems of the car have different life under different conditions and have a different degree of reliability. The failure statistics are also influenced by the characteristics of the operating region.

Thus, as a result of collected in the entities of the service network of PJSC "KAMAZ" failure statistics studies of various units, mechanisms and aggregates of the KAMAZ vehicle in countries located in different climatic zones, a significant spread was revealed (Fig. 1). The structure of failures and therefore the parts reliability depends to a large extent on the climatic zone of the operating region. Significant differences in failure statistics can also be observed within a single country, which has significant territories that differ in nature of the climate. At the same time, the quality of parts produced by different automotive parts manufacturers can be different, i.e. they can have different resources within the same operating region.



	ENGINE	TRANSMIS-SION	CHASSIS	STEERING SYSTEM	BRAKE SYSTEM	ELECTRICAL EQUIPMENT AND ELEC-TRONICS	CABIN	BODY OF TRUCKS AND SPECIAL VEHICLES
1	0,355	0,045	0,168	0,068	0,064	0,232	0,014	0,055
2	0,225	0,048	0,144	0,048	0,096	0,353	0,052	0,033
3	0,116	0,070	0,093	0,070	0,047	0,465	0,047	0,093
4	0,197	0,089	0,389	0,010	0,039	0,172	0,010	0,094
5	0,292	0,071	0,169	0,057	0,046	0,311	0,022	0,032
6	0,490	0,041	0,127	0,038	0,074	0,201	0,027	0,003

Fig. 1. Distribution of relative number of vehicles KAMAZ aggregates failures in countries of different climatic groups (1–6) according to the Keppen classification [13].

Since the dealer-service network management center determines the deliveries structure, it must have full information in order to make an informed decision as to which supplier or group of suppliers of reliability-limiting parts, it is necessary to send a preliminary application. If the given vehicle model and modifications have a sufficient array of failures, the decision is made using multidimensional analysis methods. In the first stage, those that make up the group of "interchangeable" are selected from the array of spare parts (Fig. 2). The allocated group of spare parts, in turn, is divided into products of foreign and domestic production, and the latter – on the products produced by the manufacturer and the factories-subcontractors. Thus, for each kind of part a number of different manufacturers can be distinguished. The analysis is carried out separately for each of these groups.

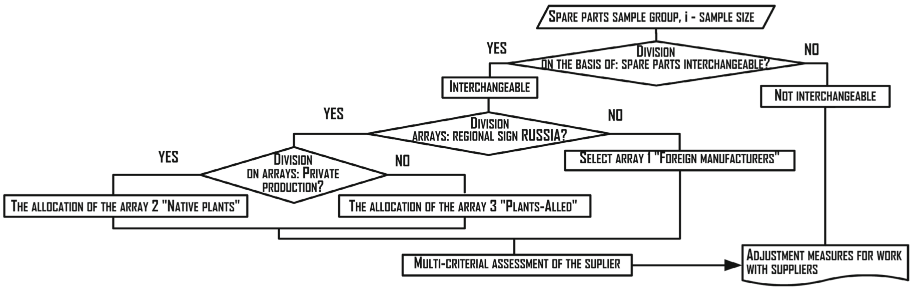


Fig. 2. Supplier evaluation algorithm.

3.2 Criteria Selection for Evaluation of Supplier

The analysis of works in the field of supply logistics has shown that when assessing the supplier, the speed of response to an application is chosen as the main criterion. At the same time, the reliability of the spare part that is important to ensure trouble-free operation is not taken into account. Such approach may negatively affect the loyalty of the customer, the direct owner of the automotive equipment, for whom re-contacting the service center to replace the corresponding part will cause the lost profit due to the car's idle time in repair. Replacement during repairs to a more reliable unit will increase the car reliability and as a result increase customer satisfaction with the service quality.

Therefore, as one of the main criteria that affect the supplier reliability, we chose "Spare part failure-free operation". To determine the numerical value of the "Spare part failure-free operation" characteristic, it is necessary to carry out the processing of the car units and aggregates failure statistics. Another concept of "Deliveries Reliability", which must be taken into account in the analysis, includes minimizing the delivery time in compliance with all requirements and conditions of supply. The delivery time can depend either on the quality of management decisions (choice of the optimal route and traffic schedule, choice of transport mode), and on the characteristics of the logistics chain (the number of overload nodes and the change of transport mode on the route, the need to re-arrange the consignment, the risks of cargo loss and its damage on the route, supplier location settings, etc.). Since the choice of a particular supplier requires considering the ratio of characteristics such as cost and time of delivery, as well as the quality of the spare part, the decision-making process for selecting a new supplier of a specific spare part is a multi-criterial task. This implies the introduction of a super-criterion, i.e. a scalar function of a vector argument, also called a linear convolution:

$$q_0(x) = q_0\left(q_1(x), q_2(x), q_3(x), q_4(x)\right), \tag{1}$$

where $1/q_1(x)$ – delivery time; $q_2(x)$ – spare part failure-free operation; $q_3(x)$ – level of fulfillment of quantitative obligations by this supplier (percentage of delivered spare parts from the total volume of the application); $1/q_4(x)$ – price.

The delivery cost of a spare part depends on the service center and the supplier itself location, as well as on the urgency and method (lot or one) of delivery. If the storage cost exceeds a spare part delivery cost, then storage is preferred on demand. In addition, there are different models for the formation of spare parts stores "for sale" and for warranty service, when they assess not only the supplier, but also the organization of the entire service delivery process. The task of warehouse logistics is complex, so it should be isolated in a separate study. When the task of selecting a supplier is being solved, it is logical to take into account the delivery time as a criterion, since this is a quantitative criterion. In addition, it is necessary to take into account that a vehicle idle time during the warranty period leads to a loss of profit for its owner, and for the manufacturer can lead to penalties and a loss of confidence in the brand.

The ordering of alternatives in magnitude with the help of super-criteria makes it possible to distinguish the best from them (according to this criterion). The form of the function is determined by the method of estimating the contribution of each criterion to the super-criteria. You can use additive or multiplicative functions.

$$q_0(x) = \sum\nolimits_{i=1}^{4} \frac{\alpha_i q_i(x)}{S_i}, \tag{2}$$

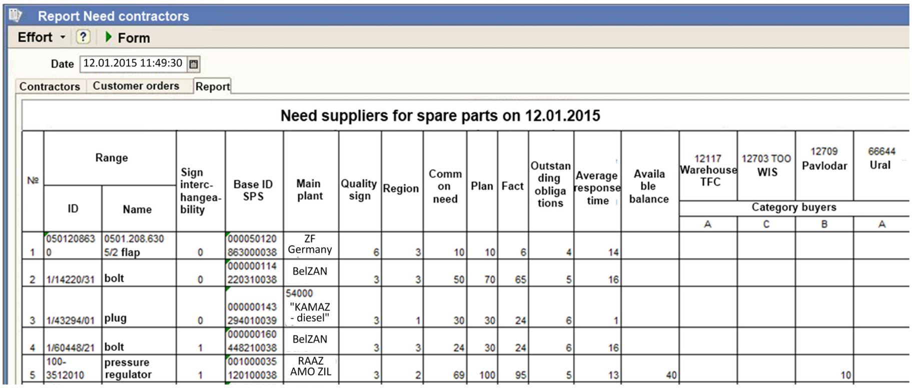
$$1 - q_0(x) = \prod\nolimits_{i=1}^{4} \left(1 - \frac{\beta_i q_i(x)}{S_i}\right). \tag{3}$$

First, the coefficients S_i provide the dimensionless of the quantity $q_i(x)/S_i$, since the partial criteria can have different dimensions and then some arithmetic operations, for example, addition, will not make sense. Secondly, in the necessary cases, with their help, the normalization condition is satisfied. To solve the above problems, we can take as S_i the average value of the corresponding criterion, which was done in this paper. The coefficients α_i, β_i reflect the relative contribution of the partial criteria to the super-criterion, i.e. they are weight coefficients. The values of these coefficients for further calculations have been determined in accordance with corporate regulations of the PJSC "KAMAZ" branded service network. Since for automotive equipment the indicator "spare part reliability" is more important, it has the greatest weight.

Separating from the array of spare parts those that constitute the group "interchangeable", and having solved the task of optimizing the super-criteria, the dealer-service network control center chooses to which supplier or group of suppliers it is necessary to send a preliminary application.

3.3 Practical Implementation of the Methodology for Evaluating Suppliers in the Presence of Failure Statistics

To collect the initial information for the evaluation of suppliers, the authors developed the software module "Report on the needs of counterparties in spare parts" (Fig. 3). This module contains information on the need for DSN in spare parts planned on the basis of preliminary applications, the actual supply to the logistics center by the manufacturer, the average response time of the manufacturer to the logistics center application, the quality characteristic of the spare parts delivered to it, the size of free spare parts left in the warehouse of the logistics center, as well as data on the need for spare parts of both the DSN as a whole and each dealer-service center (DSC) separately.


Report Need contractors
Effort ? Form
Date 12.01.2015 11:49:30
Contractors | Customer orders | Report

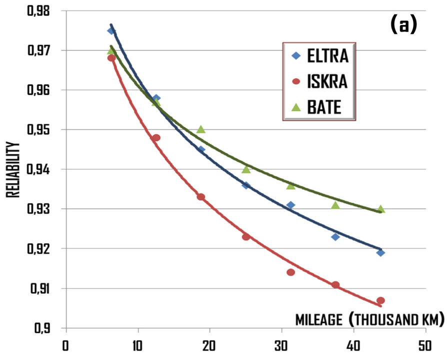
Need suppliers for spare parts on 12.01.2015

№	Range ID	Range Name	Sign interc-hangea-bility	Base ID SPS	Main plant	Quality sign	Region	Comm on need	Plan	Fact	Outstanding obligations	Average response time	Available balance	12117 Warehouse TFC	12703 TOO WIS	12709 Pavlodar	66644 Ural
														Category buyers			
														A	C	B	A
1	0501208630	0501.208.630 5/2 flap	0	000050120863000038	ZF Germany	6	3	10	10	6	4	14					
2	1/14220/31	bolt	0	000000114220310038	BelZAN	3	3	50	70	65	5	16					
3	1/43294/01	plug	0	000000143294010039	54000 "KAMAZ - diesel"	3	1	30	30	24	6	1					
4	1/60448/21	bolt	1	000000160448210038	BelZAN	3	3	24	30	24	6	16					
5	100-3512010	pressure regulator	1	001000035120100038	RAAZ AMO ZIL	3	2	69	100	95	5	13	40			10	

Fig. 3. Report on the needs of counterparties in spare parts.

In the report, the DSCs are divided into categories selected according to the quality of the prepared preliminary applications. In order to minimize risks, including untimely or inadequate supply or shortage of spare parts, the DCN management center should analyze the characteristics of suppliers and distribute the flows of requests for spare parts taking into account the received reliability indicator of the supplier. The availability of information about suppliers, the dynamics of their supply, and the specific number of failures by year allows us to evaluate suppliers by the criterion of the supplied spare parts operational reliability. Statistical analysis of vehicles various systems, units and aggregates failures has shown that the main causes of complaints are failure of power units and electrical equipment. We reviewed the failures statistics of starters of three manufacturers: ELTRA (Samara, Russia), ISKRA (Slovenia), BATE (Borisov, Belarus). The study covered the warranty period of operation, the mileage was up to 45 thousand km.

As shown by the analysis of the reliability functions (Fig. 4a), constructed in the STATISTICA 8 [14], the most reliable (from the point of failure-free operation) starter is BATE (Borisov).



(b)

		ELTRA	ISKRA	BATE
$q_1(x)$, 1/days	$\alpha_1=0{,}2$	1/16	1/14	1/13
$q_2(x)$, %	$\alpha_2=0{,}4$	0,92	0,907	0,93
$q_3(x)$, %	$\alpha_3=0{,}2$	0,86	0,95	0,9
$q_4(x)$, 1/rub	$\alpha_4=0{,}2$	1/18990	1/14590	1/14900
$q0(x)$		0,94	1,01	1,06

Fig. 4. (a) Graphs of the reliability function (probability of failure-free operation) of starters; (b) Values of criteria, weights and estimates of starter suppliers' reliability.

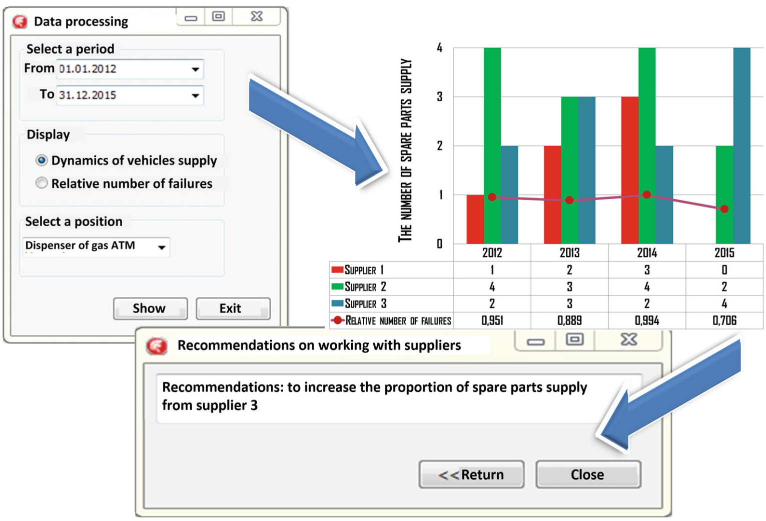


Fig. 5. Implementation of the methodology for developing recommendations for improving supplied spare parts reliability.

Thus, we evaluated suppliers according to formulas (2) and (3). Figure 4b shows the calculation of the additive function (2). As a result of calculation and comparison of super-criteria, it can be concluded that the most reliable supplier is the starter of BATE. Among the vendor estimates calculated for the multiplicative function (3), the BATE rating of the supplier has the maximum value, which is consistent with the results of the additive function application.

We implemented a software module with which it is possible to assess the impact of supplier change on the part reliability. This module allows to choose the valuation method, as well as the period during which the failure statistics are evaluated, after that the analysis will be performed and its result will be provided (Fig. 5). The failures statistics accumulate in the form of reclamation acts for different regions of operation, models of cars, etc. in a single data warehouse.

3.4 Forecasting the Supplier Reliability During Introducing New Vehicles to the Markets

The method of selecting a supplier described above assumes that enough information has been accumulated that can serve as a basis for multidimensional analysis and decision-making on specific situations. That is, the choice is made from those contractors with whom the work has already been conducted or is being conducted.

However, since in the conditions of competition the expansion and updating of the vehicle model range is so fast that the volumes of statistical data are often insufficient for adequate analysis, it is necessary to use slightly different estimation methods.

The following is a description of the developed methodology based on the intellectual analysis of the accumulated data for the search for hidden patterns that allows to predict such indicators that characterize the supplier's reliability, as the delivery time, the quality of the spare part and the level of shortages when choosing a new supplier. The basis of the analysis is information from the developed software module. The sequence of stages of obtaining new knowledge for deciding on the selection of a new supplier for the chosen position of the spare part nomenclature is the following:

1. Data collection in the data warehouse. The collection of data on current and potential suppliers' questionnaires, concluded agreements, information on the DSC operational activities in terms of vehicles failures, service centers' requests for spare parts for replacement and the deliveries made for them, histories of unfulfilled obligations are being collected. After a certain time, when the repository replenishes with new data, a decision is made about the need to update the model.
2. Selection of factors affecting the target reliability of the supplier. The following set of input variables was selected: x_1 – price level; x_2 – the percentage of defects; x_3 – warranty period; x_4 – the age of the company; x_5 – market share; x_6 – coefficient "Net debt/ EBITDA", x_7 – index of profitability.
3. Construction of the model. To search for hidden patterns, a model is calculated using the methods of constructing decision trees and logistic regression. Based on the available data, the search for hidden regularities characterizing the behavior of suppliers is carried out. To construct the decision tree, the ID3 algorithm and its modification C4.5 are chosen.

Accumulated data on interaction with suppliers allow performing analysis of the influence of various factors on the probability that the supplier will prove to be reliable. To forecast the probability of fulfillment of obligations by the supplier, a logistic regression is used [15], with the help of which the relationships between several independent variables and the dependent variable are analyzed. This is achieved by applying the regression equation (logit-transformation):

$$y = \frac{\exp(g(x))}{1 + \exp(g(x))}. \tag{4}$$

It is easy to see that regardless of the regression coefficients or the magnitude of the vector x, the predicted values (predicted y) in this model will always lie in the range from 0 to 1. In other words, y is the probability that an interesting event will occur, and in our case the supplier will be reliable.

The values of the variable g for constructing the logistic regression will be calculated as follows:

$$g(x) = \begin{cases} 1(\mathit{reliable}), \text{ if } \frac{z_1}{\overline{z_1}} * c1 + \frac{z_2}{\overline{z_2}} * c2 + \frac{z_3}{\overline{z_3}} * c3 \geq T \\ 0(\mathit{unreliable}), \text{ if } \frac{z_1}{\overline{z_1}} * c1 + \frac{z_2}{\overline{z_2}} * c2 + \frac{z_3}{\overline{z_3}} * c3 < T \end{cases}, \tag{5}$$

where $1/z_1$ – the delivery time; z_2 – spare part failure-free operation; z_3 – level of fulfillment of quantitative obligations by the given supplier (percentage of delivered spare parts from the total volume of the application), T – classification threshold.

c_1, c_2, c_3 – coefficients determined by the analyst and showing the weight of quantities z_1, z_2, z_3. The values of these coefficients for further calculations have been determined in accordance with corporate regulations of the PJSC "KAMAZ" branded service network.

To identify the probability a multiple logistic regression model with a set of input variables X and output Y defined at the previous step is constructed. The solution will be the probability of the supplier fulfilling obligations.

3.5 Practical Implementation of the Methodology for Predicting the Reliability of New Suppliers

We take $T = 0.6$, $c_1 = 0.25$, $c_2 = 0.5$, $c_3 = 0.25$.

Based on the accumulated data on the activities of suppliers, we obtain a logit transformation for the multiple logistic regression model:

$$g(x) = 0.35 + 0.05 * x_1 - 0.45 * x_2 + 0.78 * x_3 + 0.026 * x_4 + 0.2 * x_5 - 0.082 * x_6 + 0.014 * x_7. \tag{6}$$

Thus, using the transformation f (x), we can estimate the probability that the supplier will be "reliable" for given values of the input variables. For example, the reliability of supply companies with a "higher than average" price level, with a low percentage of defects, a guarantee period equal to the market average, an average age of the company (15 years), a market share of 15 to 30%, a net debt to EBITDA ratio less than 2, and the profitability index is more than 4, will equal f (x) = 0.81. This means that the supplier's reliability is about 81%.

4 Conclusion

Thus, for vehicles models that have long been present in the market and for which a sufficient amount of failures statistics has been accumulated, the choice of suppliers is based on the analysis of available statistical data. For the firm service system, they are strategic or “approved” suppliers. However, when new models or modifications are introduced to the market, “problematic” parts with a low level of reliability may appear during the warranty period of operation. Repeated failures and values of reliability parameters, which do not correspond to the declared by manufacturer, indicate the need to review the supply chain and choose, if possible, a more reliable supplier.

Usage the above described methods for the evaluation and selection of the supplier will enable the acquisition of knowledge, based on available information resources, using methods of data mining and multi-criteria selection, enhancing the effectiveness of the DSN management system.

References

1. Um, W.S.: Supplier evaluation and selection using AHP method and uncertainty theory. In: Proceedings of the 2nd World Congress on Mechanical, Chemical, and Material Engineering, 22–23 August 2016, Paper No. ICMIE 101, Budapest, Hungary (2016)
2. Senyigit, E., Soylemez, I., Atici, U.: Long-term supplier selection problem: a case study. In: 5th World Conference on Business, Economics and Management, 12–14 May 2016, Istanbul, Turkey, vol. 04, pp. 182–189 (2017)
3. Chai, J., Ngai, E.W.T.: Multi-perspective strategic supplier selection in uncertain environments. Int. J. Prod. Econ. **166**, 215–225 (2015)
4. Bruno, G., Esposito, E., Genovese, A., Simpson, M.: Applying supplier selection methodologies in a multi-stakeholder environment: a case study and a critical assessment. Expert Syst. Appl. **43**, 271–285 (2016)
5. Amit, K., Vipul, J., Sameer, K.: A comprehensive environment friendly approach for supplier selection. Omega **42**(1), 109–123 (2014)
6. Kumar, A., Jain, V., Kumar, S., Chandra, C.: Green supplier selection: a new genetic/immune strategy with industrial application. Enterp. Inf. Syst. **10**(8), 911–943 (2016)
7. Molaali, A., Tarokh, M.J.: A novel method for selecting the supplier based on association rule mining. J. Comput. Robot. **10**(1), 39–52 (2017)
8. Ho, W., Dey, P.K., Bhattacharya, A.: Strategic supplier selection using multi-stakeholder and multi-perspective approaches. Int. J. Prod. Econ. **166**, 152–154 (2015)
9. Dey, P.K., Bhattacharya, A., Ho, W.: Strategic supplier performance evaluation: a case-based action research of a UK manufacturing organization. Int. J. Prod. Econ. **166**, 192–214 (2015)
10. Tavana, M., Yazdani, M., Caprio, D.: An application of an integrated ANP–QFD framework for sustainable supplier selection. Int. J. Logistics Res. Appl. **20**(3), 254–275 (2017)
11. Sarkis, J., Dhavale, D.J.: Supplier selection for sustainable operations: a triple-bottom-line approach using a Bayesian framework. Int. J. Prod. Econ. **166**, 177–191 (2015)
12. Надежность автомобиля и пути ее повышения. http://tehnoinfos.ru/obzory/avto/362-nadejnost-avto.html. Accessed 11 April 2017. [In Russian: The vehicle reliability and the ways to improve it]

13. McKnight, T.L., Hess, D.: Physical Geography: A Landscape Appreciation. Prentice Hall, New Jersey (2010)
14. STATISTICA. http://www.statsoft.com/Products/STATISTICA-Features. Accessed 21 Nov 2016
15. Hosmer, D.W., Lemeshow Jr., S., Sturdivant, R.X.: Applied Logistic Regression. Wiley, New York (2013)

Smart Solutions for Supply Chain Management

Models of Inventory Management in Multi-level Distribution Systems

Valery Lukinskiy(✉) and Vladislav Lukinskiy

National Research University Higher School of Economics (HSE), St. Petersburg, Russia
lukinskiy@mail.ru, vladas27@mail.ru

Abstract. The analysis of perspective directions of development of logistical integration allows to establish the importance of the formation of mechanisms of management of integrated logistic functions, and also functional complexes. The complexity of the problem is determined by the variability of the types of interaction between levels in the investigated multi-level distributed systems, which in turn result in the variety of models of inventory management. They can be divided into three main subgroups: the first is with independent processes, the second is with coordination and the third – integrated models.

Keywords: Inventory management · Multi-level distribution systems Integrated models

1 Introduction

Traditionally, the theory of inventory management deals with the case of 'isolated' warehouses. In the supply chain, the central company (warehouse) is allocated, which deals with the processes of replenishment and consumption of reserves, determines the optimal periodicity of supply, the volumes of supplies and the size of the insurance stock. In other words, in such calculation models, it is assumed that decisions in the field of inventory management, taken with respect to the considered or so-called 'isolated' warehouse, will not affect the situation with the stocks of other participants in the supply chain.

For network business structures the following features are typical [2, 5, 7, 9, 11, 12]:

- inventories are formed at several hierarchical levels;
- between the stocks, there is a relationship and interdependence, conditioned by a single technology of organization of production and (or) marketing of finished products.

Thus, the object of inventory management in network structures is no longer 'isolated stocks', but integrated multi-level inventory allocation systems (multi-level inventory management systems).

A fundamentally new method for accounting for integral relationships between elements is the model of Axsäter [1]. Using the example of a two-level system for placing stock of a linear configuration, a dependence was obtained to determine the optimal

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_30

order quantity, which was called the 'echelon EOQ model'. The principle of operation of the echelon models is the same as for the classical model of the optimal order size [6]. The objective function in the models is the total costs associated with orders and the content of stocks. Required variables – the volume of orders to replenish the stock, in which the objective function is minimized. The main difference between the echelon models is that not one, but several allocations of stocks are considered.

The continuation of Axsäter's research was reflected in scientific work of [8], where modified variants were provided. They allow considering the following processes features: correlation variants of holding costs at the different system levels, various strategies of inventory warehousing and order multi-nomenclature.

Thus, despite the obvious progress, for works devoted to the problems of integrated inventory management in logistics systems [3, 4, 10], several drawbacks are typical:

- There are no recommendations for the application of the proposed models and methods. There are no examples of calculations based on the proposed models;
- Only linear multi-level systems and the case of one-off purchases (static task) are considered;
- Transportation costs are accounted for as a linear function of the transportation distance;
- The cost of storage does not depend on the type of warehouse and warehouse processing;
- Demand in the region is distributed evenly across the territory, etc.

2 Multi-variant Description of Supply Chains

A general scheme of the structure of multi-level location of stocks in the logistics system is shown in Fig. 1.

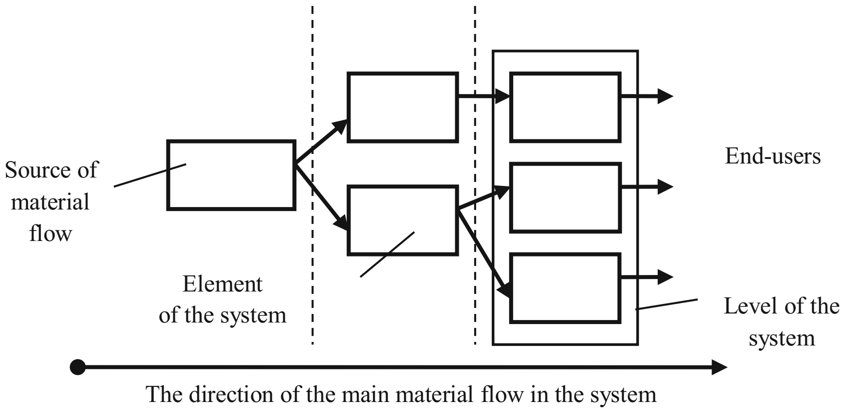


Fig. 1. General scheme of the structure of a multi-level system.

The integration approach in determining these parameters is as follows:

- The optimal periodicity and volume of orders for the elements of the system are calculated in a way, which allows to minimize the total costs in the whole system rather than in individual elements;
- The material flow between the levels of the system is coordinated in the process of deliveries by time and volume due to the formation of multiple shipments (Fig. 2).

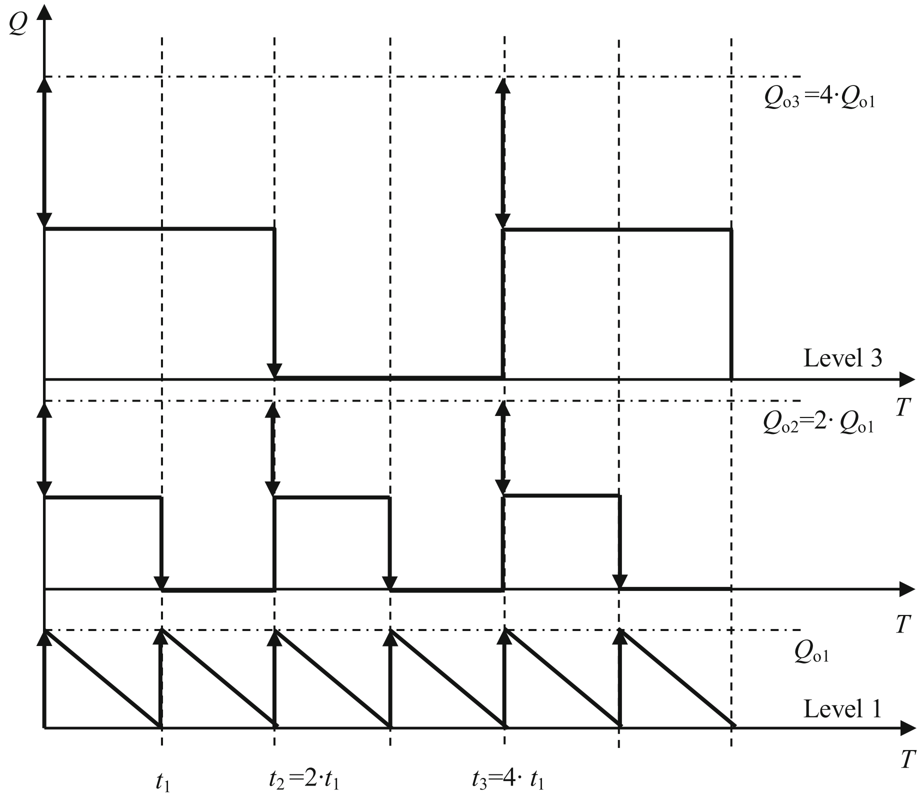


Fig. 2. Coordination of the material flow by the example of stocks allocation in a three-level system with a linear configuration: Q_{0i} — the optimal order size for the i-th level; t_i — optimal periodicity of supplies for i-th level.

The analysis of perspective directions of development of logistical integration within the framework of allocated logistics paradigms allows to establish the importance of the formation of mechanisms of management of integrated logistic functions, and also functional complexes. The carried out studies show that from the point of view of the issue of efficiency increase of logistical systems, the particular role plays the following aspects:

- choice of methods for managing the triad of logistics functions 'inventory management – warehousing – transportation', the central place in which is the inventory management;
- the need to move to the study of multi-level systems within the framework of the concept of supply chain management.

The complexity of the problem is determined by the variability of the types of interaction between levels in the investigated multi-level distributed systems, which in turn result in the variety of models of inventory management (Table 1). They can be divided into three main subgroups: the first is with independent processes, the second is with coordination and the third – integrated models.

Table 1. Models for integrated and coordinated supply chains.

Pe of interaction	Description of the interaction of the Elements of the Logistics System (ELS)	Solution options (for a two-level supply chain)
The first is with independent processes	Lack of coordination and integration (ELS operates independently)	1. Use the EOQ model for each level of the system
The second is the coordination of processes at different levels	Optimization of reserves on one of the levels; Determination of reserves at other levels using proportionality factors k_i	2. Computation of the stock size Q_1 for the lower (first) level based on the EOQ model
		3. The same (Q_2) for the top level
The third is the integration of processes at different levels	Optimization of costs for the supply chain, which includes all levels in the form Total Logistics Costs Model (*TLC*)	4. Calculation of inventory management indicators (Q_i, k, C_{min}, etc.)
		5. The same with allowance of additional corrections (the so-called adjusted or modified version)

3 Integrated Two-Level Distribution System

Nowadays, supply chains, which are represented by the distribution system, are widespread in practice. The most common of them are two-level ones with a central supplier at the second and a certain number of companies at the first level.

The basic options for inventory management in two-level supply chains, taking into account coordination and integration, are presented in Table 2.

Table 2. Variants of the description of two-level distribution supply chains.

Variant	Characteristics
1. Independent elements of SC	Each element of the supply chain is a separate item; Analytical description is provided within the framework of EOQ
2. Coordination between the ELS of different levels	For each element of the level, a link is established in the form of coordination with the element of the second level; Parameters of all elements of the first level are calculated on the basis of EOQ
3. Integration between the ELS of different levels	For each element of the first level, a link is established with the element of the second level in the form of an integrated two-level system
4. Partial integration of the system	All elements of the first level are incorporated into the form of one conventional element; The nominal element is coordinated with the element of the second level
5. Total integration of the system	All elements of the first and second levels are embedded into the form of one calculation model

4 Multi-nomenclature Models of Inventory Management in Multi-level Distribution Systems

The considered models are designed to calculate the optimal stock level and order for single-item deliveries. However, of practical interest are multilevel systems of the distribution configuration in which multi-nomenclature stocks are located.

In Table 3 one of the possible variants of the description of the three-level supply chains, reflecting the relationship between the ELS is presented.

Table 3. Variants of the description of three-level distribution supply chains.

Variant	Characteristics
1. Independent elements of SC	Each element of the supply chain is a separate item; Analytical description is provided within the framework of EOQ
2. Coordination between ELS of different levels	For each element of the first level, a link is established in the form of coordination with the element of the second and the third levels; Parameters of all elements of the first level are calculated on the basis of EOQ
3. Integration between ELS of different levels	For each element of the first level, a link is established with the element of the second and the third levels in the form of an integrated three-level system
4. Partial integration for ELS of the first level	All elements of the first level are incorporated into the form of one conventional element; The nominal element is coordinated with the element of the second and the third level
5. Partial integration for ELS of the first and the second levels	All elements of the first and second levels of the whole system are embedded into the form of one calculation model and coordinated with the element of the third level
6. Total integration of SC	All elements of the of the system are integrated into the form of one calculation model

5 Conclusions

It should be emphasized that, firstly, depending on the degree of integration, parameters of inventory management for warehouses of different levels (stock sizes, the periodicity of shipments, costs in the ELS) vary considerably. Secondly, the considered models can be added by the described modifications.

Further research on the development of the proposed models should be directed to the following areas:

- Focus on automation of calculations and decision-making;
- Expansion of the number of main models by considering concentration (assembly) and combined SC.

References

1. Axsäter, S.: Inventory Control, 2nd edn. Springer, New Jersey (2006)
2. Amiri, A.: Designing a distribution network in a supply chain system: formulation and efficient solution procedure. Eur. J. Oper. Res. **171**, 567–576 (2006)
3. Bowersox, D.J., Closs, D.J.: Logistical Management. The Integrated Supply Chain Process. McGraw-Hill Companies Inc., New York (1996)
4. Buchan, J., Koenigsberg, E.: Scientific Inventory Management. Prentice-Hall Inc., New Jersey (1963)
5. Daskin, M.S., Coullard, C., Shen, Z.-J.M.: An inventory-location model: formulation, solution algorithm and computational results. Ann. Oper. Res. **110**, 83–106 (2002)
6. Harris, F.W.: How many parts to make at once. Mag. Manage. **10**(2), 135–136 (1913)
7. Jang, Y.-J., Jang, S.-Y., Chang, B.-M., Park, J.: A combined model of network design and production/distribution planning for a supply network. Comput. Ind. Eng. **43**, 263–281 (2002)
8. Lukinskiy, V.S., Lukinskiy, V.V., Shulzhenko, T.: Logistic systems efficiency increase based on the supply chains integration. In: Proceedings of the 16th International Conference on "Reliability and Statistics in Transportation and Communication" (RelStat 2016), Riga, Latvia, pp. 19–22, October 2016. https://doi.org/10.1016/j.proeng.2017.01.073. Procedia Engineering, 178, pp. 117–122 (2017)
9. Romeijn, H.E., Shu, J., Teo, C.-P.: Designing two-echelon supply networks. Eur. J. Oper. Res. **178**, 449–462 (2007)
10. Ryzhikov, Y.I.: The theory of queues and inventory management. Nauka, Moscow (In Russian) (1969)
11. Shu, J., Teo, C.-P., Shen, Z.-J.M.: Stochastic transportation-inventory network design problem. Oper. Res. **53**, 48–60 (2005)
12. Vila, D., Martel, A., Beauregard, R.: Designing logistics networks in divergent process industries: a methodology and its application to the lumber industry. Int. J. Prod. Econom. **102**, 358–378 (2006)

The Development of Models of Supply Chain Management in Retailing

Tatyana Odintsova(✉), Nataliya Kocherjagina, and Olga Ryzhova

Saratov State Technical University Named After Yuri Gagarin, Saratov, Russia
Odintsova.tn@mail.ru, sstu_ko4@mail.ru,
helgaryzhova@gmail.com

Abstract. The situation in the retail trade of the Russian Federation is characterized by a number of negative trends: the lack of logistics infrastructure, reduced population activity of the population, active development of multi-format trade. Model development of supply chain management in the retail trade is a necessary condition for the improvement of this industry. Systematization of methods, models and concepts of supply chain management is the result of research of logistics theories and methodologies. The proposed algorithm for modelling supply chains in retail takes into account modern methods and concepts of logistics, but also based on the classification of trade flows by the degree of criticality of resources. The algorithm provides the possibility to control the suppliers, product lines and inventory. Systematic model of inventory management in manufacturing based on strategic management tasks. It is advisable to evaluate the effectiveness of supply chain modelling in retailing on the basis of indicators OOS and OSA.

Keywords: Supply chain management · Retail · Modelling · Algorithm
Methods and concepts of logistics

1 Introduction

The modern stage of functioning of the national retail is characterized by a high level of logistics costs, the absence of a developed logistics infrastructure; decrease the Consumer activity of the population; degradation of the general economic situation in the country; active development of multi-format trading. The Logistics costs level in the Russian economy is high (20% of GDP), and the level of Logistics development (LPI) is low (90th position in the international rating).

The contribution of retail in the Russian economy is estimated by the complex of indicators: the Index of the physical volume of the retail market (IPV); Value of commodity circulation; Average check size.

The study shows that negative trends are observed, contributing to the slowdown in growth and development retail.

The supply chains in a networked retail, as in other branches of the economy, are affected by innovation and new technologies. Consider the most significant factors that define new directions for development supply chain management models.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_31

1. The omnichannel approach that allows of different sales technologies and accounting requirements of all supply chain links.
2. Use of innovative trade equipment: self-service, smart carts, virtual hangs, etc.
3. The increase of the share of products “Private trade mark” in the range of most retail.
4. New technologies for the implementation of logistics functions in retail: Automation of transportation, loading and unloading, warehousing operations.
5. Information technology and new forms of trade: Internet shops, online commerce, “showrooms”, mobile applications for gadgets.
6. The modern digital technologies are promotes “personalization” of product bay reengineering of business processes and improved standards.

The existing supply chain models do not correspond to current trends in retail. Thus, the development of supply chain management models in network retail, ensuring the achievement of competitive advantages and strengthening of market positions is the goal of this study. The study consists of an introduction with justification of relevance, formulation of goal and objectives; theoretical analysis of supply chains management models and concepts and the result presented in the form of supply chains modelling algorithm in retail.

2 Theoretical Analysis

The study and synthesis of theoretical studies in supply chain management made it possible to identify and articulate the following tasks for managing supply chains in the retail:

- Selection of logistics intermediaries;
- Forecasting (indicators, flows, etc.);
- “Make or buy” (decision models);
- Identification of the nomenclature groups (ABC analysis and others);
- Additive Provisional models (“just-in-time”);
- Choice of modes of transport (transport tasks, network methods, etc.);
- Multi-objective optimization in the face of risk and uncertainty;
- Models for the synthesis (design) of logistics systems with “minimizing total costs” or “economic compromises” and others.

In practice, the Supply Chains Management is requires the simultaneous use of different concepts, methods and modelling tools [1]. In Table 1 models, methods and concepts of SCM with the specifics of the flow processes in retail were systematized as a result of a Methodological Logistics Base and Supply Chain Management study.

The organization and management of supply chains is based on an integral approach to the rational movement of all flows, the creation of a unique consumer value, full transparency, the distribution of the risks and benefits of the participants, mutual cooperation, a high degree of sharing of information while generally seeking to minimize the total cost of the entire chain [2].

Table 1. Systematization of retail supply chain management models, methods and concepts.

Model, method, concept	Characteristic
Efficient Consumer Response (ECR)	Strategy to increase the level of services to consumers through cooperation among retailers, wholesalers, and manufacturers
Vendor Managed Inventory (VMI)	Method in which a material supplier holds and manages materials and parts of their customers
Continuous Replenishment Planning (CRP)	The modified RP concept helps to improve the response to changing consumer demand
Collaborative Planning Forecasting and Replenishment (CPFR)	Business strategy that combines the intelligence of multiple trading partners in the planning and fulfillment of customer demand
Forecasting methods	Forecasting for time series using different models: trends, least squares, interval forecast
Select Vendors	Select the optimal vendor based on certain criteria: reliability, delivery time, price, quality, and risks
Game Theory	A modelling tool for consumer behaviour and response in uncertain demand conditions
Make or Buy model	Solution of the feasibility problem to transfer operations of a third-party organization with costs optimization
Models of mass service systems, Markov random processes, simulation models	Establishing a dependency between a trade format and flow parameters; estimate the intensity of the flow of customers; determination of the optimal number of service channels; system state probabilities
Just-in-time (JIT)	Providing products “just in time” provided that the costs are optimized
ABC method	Value management; Ranking of resource types by level of importance
CRM Concept	Building of customer information base
Situation Analysis	Situation analysis and assessment of possible flow options
Compromise theory (method of resolving conflict situations)	Decision-making on a set of issues of harmonizing the economic interests of participants in the supply chain
QR-concept	Quick response to changes in delivery conditions through monitoring, e-commerce and workflow technologies
The concept of TQM (Total Quality Management)	Comprehensive focused and coordinated application of quality management systems and methods with rational use of technical capabilities at all levels
SCOR Models	Integration of reengineering, benchmarking and improvement of business processes

It must be taken into account that retail supply chains are representing a complex network structure with a large area of production, storage and transport capacity, including a large number of vendors and retail sales points. The integration and management of business processes in the supply chain is provided by the focus company.

The specific of integrated flows and management business-processes in retail through the complexity of parameters (multinomenclature, time-limited implementation, seasonality, etc.) requires the development of analytical tools for supply chain modelling [3].

A number of concepts (ECR, VMI, CRP, CPFR) are used to coordinate supply chains consisting of suppliers, wholesalers and retailers to serve the needs of end-users. All of them are based on the technology of information coordination. The value of this technology is in the use of a system approach to managing complex elements of supply chains in retail. This approach is implemented through electronic data interchange (EDI). The introduction of modern methods for forecasting demand can improve the efficiency of supply planning and provide a competitive advantage to a retail enterprise [4].

The consequence of violations in supply chains in retail is the lack of certain goods on shelves in stores. "Lost sale and customers" reduce the loyalty of both the supplier and retailers. The indicator of temporary absence of goods is used to estimate its deficit in the store (Out of Shelf – OOS). Causes of food shortage are the low level of service provider, the lack of inventory in the distribution channels, incorrect ordering, low quality merchandising. The average world index of environmental protection in retail is 8.3%. The increase in this indicator indicates the inefficiency of supply chains. The reverse indicator is – On-Shelf Availability (OSA). An increase of 3% leads to an increase in supply volumes by 1% [5].

The main approaches to the assessment of indicators are physical audit, residue testing and sales analysis. These methods involve regular assurance of the assortment, accounting for stock balances and forecasting of supplies [6].

3 Results

In an environment of high external and demand structure, the Supply Chains Management in retail requires a change in the configuration (reengineering) of the network structure based on the use of combined modelling techniques.

The synthesis and analysis of the various approaches to supply chain modelling has clarified the algorithm for modelling supply chains in retail (see Fig. 1).

In order to improve the Supply Chains Management in retail, the modelling should start with an analysis of the product flows. At this stage, it is advisable to define criteria for classifying product groups, to develop a demand forecast based on product life cycle analysis, to assess customer sensitivity to product scarcity, to analyse the product portfolio and quality standards.

The essence of the second stage is the classification of commodity groups.

The authors developed a classification of resources with a wide range of nomenclature, limited shelf life, seasonality, which should be linked to methods and models of supply chain management in retail trade. The classification of product flows is based on taking into account specific conditions for the delivery, storage and sale of products:

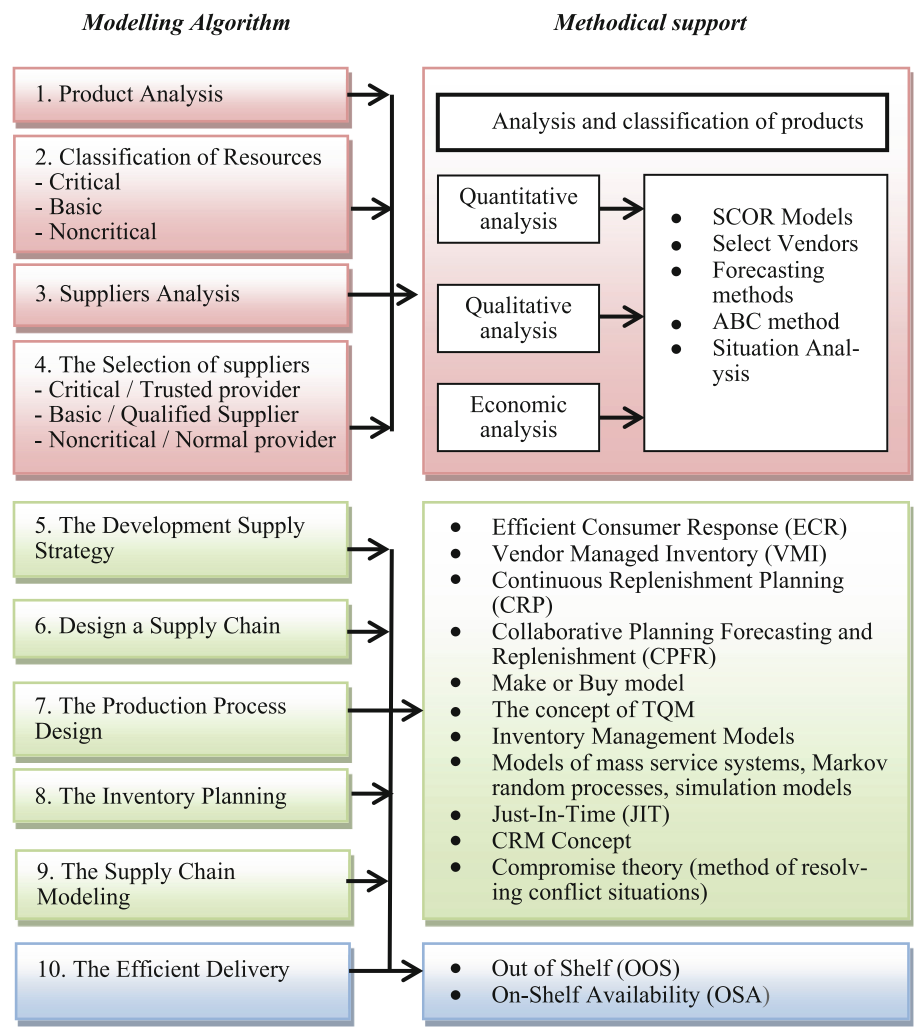


Fig. 1. The algorithm of modelling supply chains in retail.

– Critical flows-products with a high degree of customer sensitivity to the deficit that there are certain difficulties in purchasing, transporting, storing, processing (e.g., scarce, perishable, dangerous, requiring special transport and storage, etc.) for which there is no substitute for an analogy;
– The basic flows are products with the average consumer sensitivity to the deficit, which are relatively complex in terms of procurement, transport, storage, processing;
– Non-critical flows-products with low customer sensitivity to the deficit, which are easy to buy, transport, store, process.

The third and fourth stages of the algorithm involve supplier analysis (database creation, vendor market analysis, comparison and evaluation of suppliers, etc.) and their selection by product flow types. The basic parameters for evaluating suppliers are price, quality and reliability of supply. Taking into account the specifics of retail, the basic parameters should be supplemented with the following indicators: product quality assurance, production capacities, location, technical potential, a wide range of additional services, availability of information communication system and order processing. The information system (database) enables monitoring of the market and allows solving the problems of shortage of goods in the trade network.

At the fifth stage of supply chain modelling, the choice of a supply strategy for each commodity group is justified. Sixth step – design of supply chains is made taking into account the selection of the optimal suppliers depending on the criticality of commodity groups (market research, introduction of modern software for data exchange with chain participants, determining the strategy for working with each supplier depending on product groups, reconciling demand with suppliers' capabilities).

The seventh stage of the algorithm is related to the design of the business processes of the trading enterprise (capacity planning, optimization of sales technologies, production planning according to demand forecasts, supply planning).

The planning stage of stocks, taking into account the criticality of product groups and the creation of insurance stocks, precedes the final stages of the algorithm, connected with the development of the supply chain model and its effective implementation (Table 2).

Table 2. Determination of the insurance stock taking into account the management strategy.

Inventory management strategy	Model of insurance stock calculation
Strategies are preferred with an order point and replenishment to a "baseline" level with continuous monitoring of stock levels in warehouses	$S_c = x_p \cdot \sqrt{(\bar{L} + \frac{\Delta}{2}) \cdot \sigma_d^2 + \bar{d}^2 \cdot \sigma_L^2}$ where X_p – is the option that corresponds to the probability of a deficit $S_c = d(L)_{\max} - d(L) = \bar{d} \cdot \tau$ τ – the possible delay in delivery, days
Strategies of inventory management of any type in combination with the Vetter model for calculating the insurance stock; the demand forecast can be based on the average flow rate	$S_c = x_p \cdot \sqrt{(\bar{L} + \frac{\Delta}{2}) \cdot \sigma_d^2 + \bar{d}^2 \cdot \sigma_L^2}$ X_p – is the option that corresponds to the probability of a deficit
Strategies for inventory management of any type; when calculating the safety stock, only the probabilistic characteristics of the delivery parameters (time and volume of delivery)	$S_c = t_p \cdot \sqrt{\bar{T}_{c3} \cdot \sigma_d^2 + \bar{d}^2 \cdot \sigma_T^2}$ $\bar{T}_{c3}$ – average time period between orders, days
"Periodic" stock management strategies. Preferred is the "uniform delivery" strategy with a constant order size	$S_c = \bar{d} \cdot (\bar{T}' + \tau) - \bar{d} \cdot \bar{T}' = d \cdot \tau$ τ – the possible delay in delivery, days

The theoretical analysis of supply chains and the identification of retail management specific are allowed to develop methodological approaches to improving the supply chains efficiency on the proposed algorithm. Modelling of supply chains in retail trade on the basis of the proposed algorithm will allow enterprises to provide the following competitive advantages: the availability of products in the required quantity and the required quality; increase in the speed of response of warehouse stocks (insurance reserves) to sudden changes in demand; optimization of logistics costs and lost profit.

4 Conclusion

The supply chains modelling algorithm in retail, taking into account the use of analytical models, which allows to identify the necessary direction of changes to ensure the required level of stock in network structures, has been developed. This will make it possible to determine the basic supply indicators and the volume of stocks based on the developed algorithm for various configurations of supply chains.

To further develop this research, we can offer the following directions: expansion of a complex of supply chain management models, taking into account current trends and market requirements, expanding the system of indicators for assessing the effectiveness of supply chains in retail.

The authors are grateful to Professor of the Higher School of Economics (Russia) Lukinsky Valery for valuable advice for planning this study.

References

1. Stock, R., Boyer, S.: Developing a consensus definition of supply chain management: a qualitative study. Int. J. Phys. Distrib. Logistics Manage. **39**(8) (2009)
2. Bowersox, D., Closs, D.: Logistical Management: The Integrated Supply Chain Process. The McGraw-Hill Companies, Inc., New York (1996)
3. Lukinskiy, V., Lukinskiy, V.: Models and Methods of the Logistics Theory. SPbSUEE, St. Petersburg (2012). (in Russian)
4. Tsenina, T., Tsenina, E.: Risks Management: Training. SPbSEU, St. Petersburg (2013). (in Russian)
5. Kotik, S.: On Shelf Availability in Russian retail: do not lose sales. J. Retail & Loyalty **1**(55) (2016). (in Russian)
6. Fernie, J., Sparks, L.: Logistics and Retail Management: Emerging Issues and New Challenges in the Retail Supply Chain, 4th edn. Kogan Page Ltd., London (2014)

Factors Influencing Local Food Sales Through Green Public Procurement in Rezekne Municipality

Anda Zvaigzne[1(✉)], Inita Krivašonoka[2], and Inta Kotāne[1]

[1] Rezekne Academy of Technologies, Atbrivosanas aleja 115, Rezekne, Latvia
anda.zvaigzne@rta.lv

[2] Latvian Rural Advisory and Training Centre, Riga Street 34, Ozolnieki, Latvia

Abstract. Green procurement is a fast-growing trend in the European Union, and it positively affects the development of local territories. In Rezekne municipality, criteria for green public procurement (GPP) of food products are not employed by procuring institutions or it is done infrequently; therefore, the reasons that hinder the application of GPP criteria and motivate the introduction of the criteria in public food procurement operations have to be examined. The present research continues and supplements the previous ones. The authors of the paper have developed a questionnaire for a survey of experts aimed at identifying expert opinions about the implementation of GPP at municipal institutions to purchase food products. The research aim is to examine the factors that hinder local food sales through green public procurement in Rezekne municipality. The research results showed that the experts considered the fact that such "food products are free of genetically modified organisms" or they are healthy to be the most motivating factor after assessing all the factors motivating municipal institutions to implement GPP to purchase food products. Assessing the expected results of use of GPP as a sustainable food system, the experts rated the factor "Fostering environmentally friendly production methods" higher than the other ecological factors. Research methods used: monographic, descriptive, analysis, synthesis, statistical analysis, a sociological method – an expert survey.

Keywords: Local food products · Factors influencing sales
Green public procurement

1 Introduction

In recent years both in the world and in Latvia, food consumption as close to the production site or local food is increasingly discussed among scientists and the general public. Knight [9] has researched trends in the consumption of local products in the world and has found that, according to a number of authors, the demand for local products increases, and localness is one the most latest trends in the global food market. However, Kneafsey together with the co-authors [8] have concluded that local food systems, in which the production, processing, sales and consumption of food products take place within relatively small distances, make a significant positive effect on the local economy. Gómez and Zhang [5] have found that the economic aspects of local food depend on the

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_32

performance of local farmers, and the demand for local and regional food has become the key driver in the farm- and community-based economy, creating new jobs and contributing to economic growth.

Local food systems support the local economy. The purchase of food from local farmers positively affects local entrepreneurship, while providing a significant income source for the local farmers; in this way, the viability of many small local agricultural holdings is maintained [1]. Unlike large industrial-scale farms, small family farms prefer spending their revenues on local products (e.g. seeds, agricultural goods etc.); furthermore, food that has been grown, processes and supplied locally (e.g. for local schools), creates jobs and, in this way, stimulates the local economy [7]. Higher local economic activity and a greater number of jobs generate more tax revenue and create a stronger economic basis to support other enterprises and related institutions.

In the world and in Latvia too research studies [10, 11] have been done on how the introduction of green public procurement (GPP) stimulates the development of a local region. GPP is an instrument that directly contributes to an increase in local food sales.

The research aim is to examine the factors that hinder local food sales through green public procurement in Rezekne municipality. Research methods used: monographic, descriptive, analysis, synthesis, statistical analysis, a sociological method – an expert survey.

To achieve the research aim, the following research questions were raised:

- In what way GPP can affect the sales of local products?
- What trends in the procurement of food products are observed in Latvia as a whole and in Rezekne municipality?
- How an assessment of the motivating factors, basic principles, barriers and sustainability of GPP can contribute to the sales of local products?

The research structure assists in achieving the aim, corresponds to the research questions and includes a theoretical examination of GPP, an assessment of the situation with public procurement operations of the government of Rezekne municipality, an assessment and analysis of expert opinions on the implementation of GPP of food at municipal institutions as well as conclusions.

2 Research Methodology

The research methodology was developed in accordance with the research questions. The research was based on the analysis of scientific and academic publications and the results of the authors' previous research studies; the present paper used information from the Procurement Monitoring Bureau (hereinafter the PMB), as well as the results of the expert survey.

The authors of the paper have developed a questionnaire for a survey of experts aimed at identifying expert opinions about the implementation of GPP at municipal institutions to purchase food products. The expert survey questionnaire was developed based on research studies [3, 12], the legal framework [17] and policy documents [23]. The questionnaire contained four question groups, in which the experts rated the motivating factors for the implementation of GPP of food products at municipal institutions; the importance of GPP basic principles; barriers to the use of GPP and the expected results of use of GPP as a sustainable food system. The experts rated reply options to the questions on a semantically graded five-level scale: not at all important (1 point), unimportant (2 points), neither important nor unimportant (3 points), important (4 points) and very important (5 points). The survey was conducted in June 2017; it involved seven experts.

Expert selection criteria: higher education, at least three-year experience in agriculture or experience in/knowledge of public procurement operations and the management and evaluation of the operations and knowledge of the European Union (EU) legislation on public procurement. The selected experts represented various organisations: local governments, farms, advisory and educational institutions.

The research employed the monographic and descriptive methods to examine the theoretical aspects of green public procurement. Analysis, synthesis and statistical analysis were used to examine empirical data.

3 Research Results

3.1 Theoretical Aspects of Green Public Procurement

Such researchers as Morgan and Sonnino [14], who have extensively examined opportunities to supply high-quality, fresh and local food to schools, believes that an increasing number of individuals begin understanding that healthy food is a strategic priority to improve the health of the population, social justice and environmental integrity, which are the key principles for sustainable development. Michelsen and de Boer [13] have come to a conclusion that sustainable development in general is promoted by contributing to sustainable food consumption and production patterns. Public institutions are considered one of the most influential groups in the manufacturing and production of environmentally friendly products. The indirect impact embodied in purchased products and services is often the major environmental impact caused by public authorities. The way we consume resources in the EU is causing environmental damage at a rate that cannot be sustained. If the world as a whole followed the EU's pattern of consumption, global resource use could quadruple within 20 years. Apart from the resulting environmental and health problems, this trend could threaten economic growth due to rapidly decreasing natural resources and the cost of addressing these issues [6].

GPP is an important means of achieving environmental policy goals with regard to climate change, resource use and sustainable consumption and production, especially taking into consideration how significant expenditures on goods and services by the public sector are [24]. Many public institutions in Europe implement in practice not only GPP but also sustainable public procurement – decisions on purchases take into consideration both environmental and social criteria. Tukker [20] has admitted that GPP becomes a cornerstone of environmental policies both at EU and Member State levels. GPP can help public institutions save funding if considering the total cost of the implementation cycle of any procurement contract, not only the price of purchases. Institutions implementing GPP will be better prepared for facing the growing environmental problems, e.g. reduction of greenhouse gas emissions or transition to a circular economy [24]. GPP is an instrument that directly contributes to and ensures an increase in local food sales. Performing a procurement operation in accordance with the principles of GPP and sustainability means making sure that the goods and services purchased have made the smallest impact on the environment and a positive social effect. Accordingly, the choice of food products also plays a large role in reducing environmental impacts and contributing to the health of the population.

Sustainable procurement can tackle a variety of social and economic problems. It can create new jobs and increase wellbeing in remote regions, increase contractual opportunities for small and medium enterprises and enterprises of ethnic minorities as well as to contribute to adult basic skills and disability, racial and gender equality by means of specific services earmarked by the contracting parties. Sustainable procurement can also encourage a lot of enterprises involved to introduce innovations and produce most cost-efficient, competitive and sustainable goods and services for the public sector, firms and consumers [2]. Walker and Brammer in [22] believe that sustainable procurement is a kind of procurement that, in accordance with the principles of sustainable development, for instance, contributes to a strong, healthy and just society, taking into account environmental constraints as well as promoting good governance.

A summary produced by the EU enumerates the following GPP advantages: achievement of particular goals and targets in the environmental field, e.g. reduction of greenhouse gas emissions, energy efficiency and preservation of natural resources; cost savings; greater trust of residents, entrepreneurs and the public in state administration institutions; support for innovations and the production of environment-friendly goods and services through increasing their sales opportunities in the market; creation of healthy working conditions for employees; increase in the capacity of public organisations to tackle environment-related problems in the future [21]. The alliance for better food and farming has defined that sustainable food is a kind of food that is produced, processed and sold in a way that: contributes to the local economy and sustainable living; protects the diversity of plant and animal species (and the welfare of domestic and wild species); avoids damage to natural resources and increase in climate change; ensures social gains, e.g. good quality food, safe and healthy products and education opportunities [18]. Clement et al. [4] argue that local governments are well suited for introducing green procurement. More than national authorities they have possibilities for more explorative behaviour and can be early movers, also regarding procurement.

Opportunities through sustainable procurement for local governments designed by the EU Member States are as follows: ensuring long-term profitability; reducing the amount of waste, carbon emissions and the consumption of energy and water; promoting recycling and the reuse of packaging etc.; preserving biodiversity; stopping the purchase of unsustainable products or the purchase of products from illegal sources; supporting fair and sustainable economic growth [15]. Inclusion of GPP would also stimulate the innovation capabilities of firms. According to Porter's theory, GPP could represent a "properly designed" environmental policy instrument able to conjugate environmental benefits and competitive improvement in the firm's performance. Indeed, the high impact of GPP on production activities positively influences the probability that firms invest (at both technological and organizational levels) in innovative solutions [19].

The authors believe that GPP and the application of sustainability criteria in the purchase of food products are of great importance. Both environmental and economic considerations, which complement one another in a long-term, are taken into account. A "greener" economy reduces environmental protection costs, as resources are used more effectively, while new, ecological approaches and techniques create new jobs, stimulate the economy and strengthen competitiveness.

3.2 Assessment of the Situation in Public Procurement by the Local Government of Rezekne Municipality

In Latvia, two kinds of public procurement are practised to cater customers of municipal institutions and schoolchildren: procurement of food products and procurement of catering services. In Latvia, public procurement operations and related data are collected and provided by the PMB, which a direct state administration authority subordinate to the Ministry of Finance. In recent years, according to the PMB data, the total amount of public food procurement in Latvia exceeded EUR 40 million, and the amount tented to increase (Fig. 1). In Rezekne municipality, the average annual amount of public food procurement was approximately EUR 300–400 thousand.

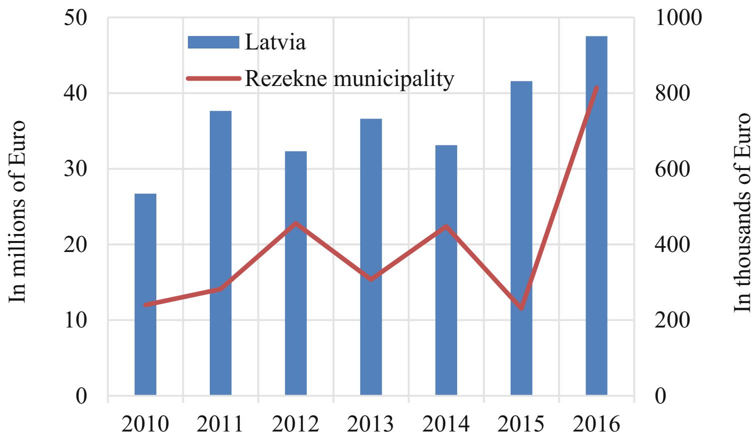


Fig. 1. Amounts of public food procurement in Latvia and Rezekne municipality in the period 2010–2016 (authors' construction based on [16]).

An analysis of the annual amounts of public food procurement shows (Fig. 1) that the expenditures on food were unstable. In view of the fact that the amounts of procurement were not large, any procurement operation considerably affected the total amount. After examining the reasons of the changes, one can conclude that an orphanage and a primary boarding-school are situated in Rezekne municipality, and food procurement by both institutions accounted for approximately 17% of the total amount of food procurement in the municipality; both institutions announced a procurement tender once in two years. For this reason, in a year when no food procurement tender was announced, a decrease in the amount of food procurement was reported in the municipality. An analysis of the amounts of public food procurement in Rezekne municipality allows concluding that procurement operations were carried out in a decentralised way. The board of every civil parish procured food for educational institutions located in its territory, including pre-school educational institutions, while old people's homes and professional educational institutions as well as orphanages carried out procurement operations themselves to provide their customers with food. The dominant principle for identifying the winner of any procurement tender held in Rezekne municipality was the lowest price principle, which did not include sustainability and GPP criteria. This could be explained by the fact that the employees of institutions for whom procurement was only one of the responsibilities lacked knowledge of and experience in how to integrate the criteria in the procurement procedure. Consequently, more than 90% of the procurement tenders were won by wholesale companies and large Latvian food producers.

3.3 Implementation of Green Public Procurement of Food Products at Municipal Institutions: Expert Opinions

The experts rated the motivating factors for the implementation of GPP of food products at municipal institutions (on a five-point scale), and, according to them, the most motivating factor was "Food products free of genetically modified organisms" or healthy food (Table 1).

Table 1. Assessment of the motivating factors for the implementation of GPP of food products at municipal institutions, in points, a survey of experts ($n = 7$), (authors' construction).

Factors	Mean	Mode
Food products free of genetically modified organisms/health	4.71	5
Financial savings	4.29	5
Strengthening public support	4.29	4
Fresh and seasonal	4.00	4
Effective achievement of environmental and health policy goals	4.00	3[a]
Wish to protect the environment (reduction of CO_2 emissions and wastes)	3.71	3[a]
Higher nutritional value	3.57	3
Tasty	3.57	4
Preservation of biodiversity	3.57	5
Effective achievement of social goals	3.57	3
Promotion of innovative solutions	3.14	4
Global contribution	3.00	4

[a.] Multiple modes exist. The smallest value is shown.

The experts rated this factor at 4.71 out of 5 points; the mode was 5, which indicated that most of the experts rated it at 5. The next factors that, according to the experts, motivated the implementation of GPP of food products at municipal institutions to a great extent were "Financial savings" and "Strengthening public support". The experts' average rating for the mentioned factors was 4.29 out of 5 points. The choice of the experts was probably determined by the fact that local food purchases result in more money to the local economy as well as local economic activity increases and additional jobs generate more tax revenue, thereby strengthening the local economy. The experts rated the factor "Global contribution" the lowest, which could be explained by the fact that specialists who do procurement operations mainly consider how their institution, local residents and the local community benefit from it rather than how their procurement action affects vaster territories and global trends.

The promotion of domestic food product sales by applying GPP requirements in public procurement is viewed positively. In the next questionnaire question, the experts had to rate the importance of GPP basic principles. After analysing the ratings of the experts regarding the importance of GPP basic principles, it has to be noted that the average rating of all the GPP basic principles in the five-point system was 4 points and higher. The experts gave the highest rating to the principle that local food products are free of genetically modified organisms. The experts rated this factor, on average, at 4.71 out of 5 points. This rating both in terms of points and in essence is equal to the rating given by the experts to the first question about the motivating factors for the implementation of GPP of food at municipal institutions. The unanimity of the experts regarding using genetically modified organisms (GMO) in food as little as possible confirms the existence of this problem. The reason is that "the great role of livestock feed containing genetically modified organisms in agricultural production is worth global worrying. On the one hand, food containing GMO ingredients is banned in Latvia, while on the other hand there are no restrictions imposed on the use of feed containing GMO ingredients" [25].

Before rating barriers to the use of GPP by the experts in relation to the third questionnaire question, the authors gave characteristics of the barriers [23].

Assessing the potential barriers to the use of GPP, the experts were unanimous that the most important barriers were *economic* ones – the average expert rating was 5 points in the five-point system. The next most important barriers to the use of GPP, according to the experts, were *informational and motivational barriers and legal barriers*. The average expert rating for both above-mentioned factors was 4.29 points. One can conclude that, according to the experts, the use of GPP was hindered by: price policies on "green" products; an insufficient awareness and education of individuals being in charge of procurement operations in respect to environment-friendly goods as well as the environmental impacts of consumption and a lack of clarity regarding the application of legal provisions.

The authors of the paper employed sustainable food system expectations [12] to assess the expected results of use of GPP as a sustainable food system, dividing sustainability dimensions or factors into environmental, economic and social ones. After assessing the expected results of use of GPP as a sustainable food system, the experts rated "Fostering environmentally friendly production methods" the highest among the environmental factors (Table 2). The average expert rating of this factor was 4.86 out of 5 points, the mode was 5, which indicated that 5 points was the most frequent rating given by the experts. Among the economic factors, the experts gave the highest ratings to "Reducing prices of seasonal products" and "Creating employment opportunities".

The experts were unanimous that "Keeping agricultural production and small manufacturing enterprises in the region by creating productivity employment" was the most important social factor with the average rating of 5 points on a five-point scale.

Table 2. Assessment of the expected results of use of GPP as a sustainable food system,in points, a survey of experts ($n = 7$), (authors' construction).

Expected results	Mean	Mode
Ecological factors		
Fostering environmentally friendly production methods (organic production, protection of local biodiversity, reduced chemical inputs)	4.86	5
Reduced resource input	4.00	4
Reducing environmental effects of transportation, e.g. emissions of air pollutants	3.86	4
Reducing specialisation and intensification in agriculture through more diverse local land use	3.86	3
Conserving the traditional agricultural landscape	3.57	3
Economic factors		
Reducing prices of seasonal products	4.43	4
Creating employment opportunities	4.43	4
Increasing regional economic growth	4.29	4
Raising the income for farmers and food manufacturers	4.29	4[a]
Reducing local dependencies on external market forces and reducing market power of processing and distribution businesses	4.14	5
Changing consumption patterns	4.14	4
Social factors		
Keeping agricultural production and small manufacturing enterprises in the region by creating productivity employment	5.00	5
Improving personal wellbeing by providing healthier, more nutritional, fresher and better tasting food	4.71	5
Increasing security and safety of food supply	4.00	4
Increasing awareness about environmental and social effects of consumption (the "embeddedness-effect")	3.86	4
Improving work safety and increasing the job satisfaction of farmers and processors	3.71	3[a]
Conserving traditional production techniques and consumption patterns (cultural identity)	3.57	2[a]
Increasing community power and personal relationships	3.43	3
Raising levels of social justice locally and internationally	3.29	2[a]

[a] Multiple modes exist. The smallest value is shown.

4 Conclusions

1. GPP is an instrument ZPI that contributes to and ensures an increase in the sales of local food products in a direct way as well as tackles various social and economic problems: promotes the economic activity of local small and medium enterprises, increases wellbeing in the regions etc.
2. In recent years, an upward trend in the procurement of food products was observed in Latvia as a whole. In Rezekne municipality, the average amount of food products procured was volatile, as a few government-funded institutions announced a procurement tender once in two years. In Rezekne municipality, food was procured in a decentralised way, and the dominant principle for identifying the winner of any procurement tender was the lowest price principle, which did not include sustainability and GPP criteria.
3. The experts regarded the fact that "*food products are free of genetically modified organisms*" or are healthier as the most motivating factor for the implementation of GPP of food products at municipal institutions, which is also the most important basic principle of GPP. Economic barriers, according to the experts, were the most significant ones in the use of GPP. The assessment of the expected results of use of GPP as a sustainable food system employed the following sustainability factors: environmental, economic and social. The experts gave the highest ratings to the environmental factor "*Fostering environmentally friendly production methods*", the economic factors "*Reducing prices of seasonal products*" and "*Creating employment opportunities*" and the social factor "*Keeping agricultural production and small manufacturing enterprises in the region by creating productivity employment*".
4. In view of the fact that GPP is not used in Rezekne municipality, further research studies are going to focus on the reasons of it, as well as proposals will be made with regard to the integration of GPP criteria in public procurement tenders.

Acknowledgements.

RTA
RĒZEKNES TEHNOLOĢIJU AKADĒMIJA

The research was elaborated with the financial assistance of Rezekne Academy of Technologies provided through a research grant for the "Opportunities for the Green Public Procurement of Food by Municipal Institutions in Rezekne Municipality".

References

1. Brown, C., Miller, S.: The impacts of local markets: a review of research on farmers markets and community supported agriculture (CSA). Am. J. Agr. Econ. **90**(5), 1298–1302 (2008)

2. Burr, T.: Addressing the environmental impacts of Government procurement (2009), http://www.nao.org.uk/publications/0809/addressing_the_environmental.aspx. Accessed 28 June 2017
3. Chakrabarti, S.: Factors influencing organic food purchase in India – expert survey insights. Br. Food J. **112**(8), 902–915 (2010)
4. Clement, S., Plas, G., Erdmenger, C.: Local experiences: green purchasing practices in six European cities. In: Erdmenger, C. (ed.) Buying into the Environment: Experiences, Opportunities and Potential for Eco-procurement. Greenleaf, Sheffield, pp. 69–93 (2003)
5. Gómez, M.I., Zhang, L.: Impacts of concentration in hog production on economic growth in rural illinois: an econometric analysis (2000), http://ageconsearch.umn.edu/bitstream/21846/1/sp00go03.pdf. Accessed 28 June 2017
6. Green public procurement (2012), https://www.etenders.gov.mt/epps/viewInfo.do?section=statistics&selectedItem=viewInfo.do%3Fsection%3Dstatistics. Accessed 28 June 2017
7. Halweil, B.: Home Grown: The Case For Local Food In A Global Market. Worldwatch Paper 163 (2002), http://www.worldwatch.org/node/827. Accessed 28 June 2017
8. Kneafsey, M., Venn, L., Schmutz, U., Balázs, B., Trenchard, L., Eyden-Wood, T., Bos, E., Sutton, G., Blackett, M.: Short food supply chains and local food systems in the EU. A State of Play of their Socio-Economic Characteristics. JRC Scientific and Policy Reports (2013), ftp.jrc.es/EURdoc/JRC80420.pdf. Accessed 28 June 2017
9. Knight, A.J.: Evaluating local food programs: the case of select nova scotia. Eval. Program Plan. **36**(1), 29–39 (2011)
10. Krivašonoka, I., Siliņa, L.: Theoretical aspects of local food distribution. In: Annual 21st International Scientific Conference Proceedings "Research for Rural Development 2015", pp. 276–283. Latvia University of Agriculture, Jelgava (2015)
11. Krivašonoka, I., Zvirbule, A.: Role of municipalities in local food distribution in Latvia. In: Proceedings of the 2016 International Scientific Conference "Economic Science for Rural Development", vol. 41, pp. 96–103. Latvia University of Agriculture, Jelgava (2016)
12. Lehtinen, U.: Sustainability and local food procurement: a case study of Finnish public catering. Br. Food J. **114**(8), 1053–1071 (2012)
13. Michelsen, O., de Boer, L.: Green procurement in norway; a survey of practices at the municipal and county level. J. Environ. Manage. **91**(1), 160–167 (2009)
14. Morgan, K., Sonnino, R.: Rethinking School Food: The Power of the Public Plate (2010)
15. National Audit Office. Sustainable Procurement in Government. Briefing for the House of Commons Environmental Audit Select Committee (2013)
16. PMB data, https://www.iub.gov.lv/. Accessed 28 June 2017
17. Prasības zaļajam publiskajam iepirkumam un to piemērošanas kārtība (in Latvian) (2017), https://likumi.lv/ta/id/291867-prasibas-zalajam-publiskajam-iepirkumam-un-to-piemerosanas-kartiba, Accessed 28 June 2017
18. Sustain Homepage, What is Sustainable Food? https://www.sustainweb.org/sustainablefood/what_is_sustainable_food/. Accessed 28 June 2017
19. Testa, F., Iraldo, F., Frey, M.: The effect of environmental regulation on firms' competitive performance: the case of the building & construction sector in some EU regions. J. Environ. Manage. **92**, 2136–2144 (2011)
20. Tukker, A., Emmert, S., Charter, M., Vezzoli, C., Sto, E., Andersen, M.M., Geerken, T., Tischner, U., Lahlou, S.: Fostering change to sustainable consumption and production: an evidence based view. J. Clean. Prod. **16**, 1218–1225 (2008)

21. Videi nekaitīga iegāde! Zaļais publiskais iepirkums Eiropā (2011) (in Latvian), http://ec.europa.eu/environment/archives/gpp/handbook_summary_2011_lv.pdf. Accessed 28 June 2017
22. Walker, H., Brammer, S.: Sustainable procurement in the United Kingdom public sector. Supply Chain Manage. Int. J. **14**(2), 128–137 (2009)
23. Zaļā iepirkuma veicināšanas plāns 2015.-2017.gadam (2015) (in Latvian), https://likumi.lv/ta/id/272295-par-zala-iepirkuma-veicinasanas-planu-2015-2017-gadam. Accessed 28 June 2017
24. Zaļais iepirkums. Rokasgrāmata par zaļo publisko iepirkumu. 3.izd. (2016) (in Latvian), http://ec.europa.eu/environment/gpp/pdf/handbook_2016_lv.pdf, Accessed 28 June 2017
25. Zaļākās valsts Eiropā dubultā morāle (2017. gada 28. marts). Dienas Bizness (in Latvian). 28 Mar 2017

Determination of Parameters for Forming Right Allocation of Items in Picking Area

Raitis Apsalons and Genady Gromov(✉)

Transport and Telecommunication Institute, Riga, Latvia
raitis.apsalons@inbox.lv, gromovs.g@tsi.lv

Abstract. Forming right allocation of items in picking area is the second approach for diminishing picking distance. The first approach is routing methods for customers' orders picking. In practice, the principle of dividing of orders is used to determine which quantities are going to collect from storing area and which are going to pick from picking area. Also replenishment principles can affect picking distance. Several variants of organizing orders' picking process in the warehouse have proposed by authors in earlier scientific papers. It has been determined parameters for forming right allocation sequence of SKU (stock keeping units) into picking addresses corresponding by picking route. Sometimes the warehouse can use combined parameter which includes two or more parameters for forming right allocation sequence of SKU into picking addresses.

Keywords: Picking process · Replenishment of items · Picking route
Right allocation of items · Picking locations · Parameters of allocation of items

1 Introduction

Nowadays globalization is connected not only by use of different modes of transport or intermodal and multimodal transport, but also by development different methods of logistics in warehousing area. Right planning, organizing and controlling of picking process becomes vitally important. The key indicator for choose of any picking technology in the warehouse appears velocity of order lines picked per paid man hour. If number of order lines picked per paid man hour is relatively small, usually primitive picking technologies are used. Such picking technologies support physical picking system: walk and pick [1]. Picking technologies here are: the paper picking, RFID (Radio-Frequency Identification) picking or more developed picking technologies such as: visual picking, picking by voice [2]. In this paper it is discussed picking area (PA) which is located into storing area (SA). This means that one row rack storing system available in the definite warehouse. Picking process will be realized by picking handling units (HU) and customer units (CU). The ground level and first level of pallet racks are used as PA. The one picking location of each item consists of 2 pallets: 1 pallet on ground level and second one on the first level of rack. The replenishment is appropriated for moving the items from SA to PA to avoid stock – outs in picking time interval. Therefore if any single item in picking location achieves critical level, replenishment starts by the

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_33

signal in warehouse management system (WMS). This approach is called as Red Card principle (RCP) [3].

The main purpose of paper is to determine one general parameter or more parameters for forming right allocation sequence of items for replacement picking addresses corresponding by picking route. This approach helps to decrease total picking time and total picking cost. These results could receive by diminishing total picking travel distance valued in a day, per month or during single season. The definition of the scientific problem is to approve that those parameters for forming right allocation sequence of items into picking addresses can diminish total picking travel distance. The object of the research concerns the picking process. The subject of the research is allocation sequence of items into picking addresses.

2 The Meaning of Principle of Dividing of Orders

At first for each item (SKU) at least one picking address has planned in PA [4]. The replenishment process can be realized by use of approaches of the layout of items in PA: either by the single picking location for each single item, when replenishment is realised in picking process or by various picking locations for each single item, when replenishment is realised just only before picking process or after it. Here main emphasis is on picking process without any interruptions, so that stock – out situations for any SKU would not occur.

The principle of dividing of orders (PDO) states that quantities for definite SKU of each order are divided into 2 parts:

- For picking full pallets (FPLL) from SA – for single order usually expressed in customer units from SA.
- For picking HU and CU from PA – for single order usually expressed in customer units from PA.

Necessity of use of PDO depends on several variants of organizing orders' picking process in the warehouse, proposed by authors:

- Variant 1: Picking from the single area – only the SA exists in the warehouse, in picking process for reaching highest levels of racks either pallet forklifts are used for extracting FPLL, and high level order pickers (HLOP) are used for extracting HU and CU.
- Variant 2: Picking from two separate zones, additionally using PDO, from SA are collected FPLL, but from PA – HU and CU [3].
- Variant 3: Picking from two separate zones, but PA is allocated in SA, additionally using PDO, from SA are collected FPLL (second and higher levels of pallet racks are used), but from PA – HU and CU (the first two levels of pallet racks are used).
- Variant 4: Picking from three separate zones, additionally using PDO, from SA are collected FPLL, from the first PA (PA1) – HU and from the second PA (PA2) are picked CU.
- Variant 5: Automated picking systems, for example the AS/RS – Automated Storage and Retrieval System, expressed as the G2M – goods to man picking system.

The M2G – man to goods picking system is revised in this paper.

At the second there are several routing strategies for “pick and walk” approach, but not all of them are suitable for any situation [5]. The third, there are several picking methods for “pick and walk” approach [6]: S – shape method; largest gap method; return method; midpoint method; combined method of previous. And, finally evaluation of optimal picking method should be done.

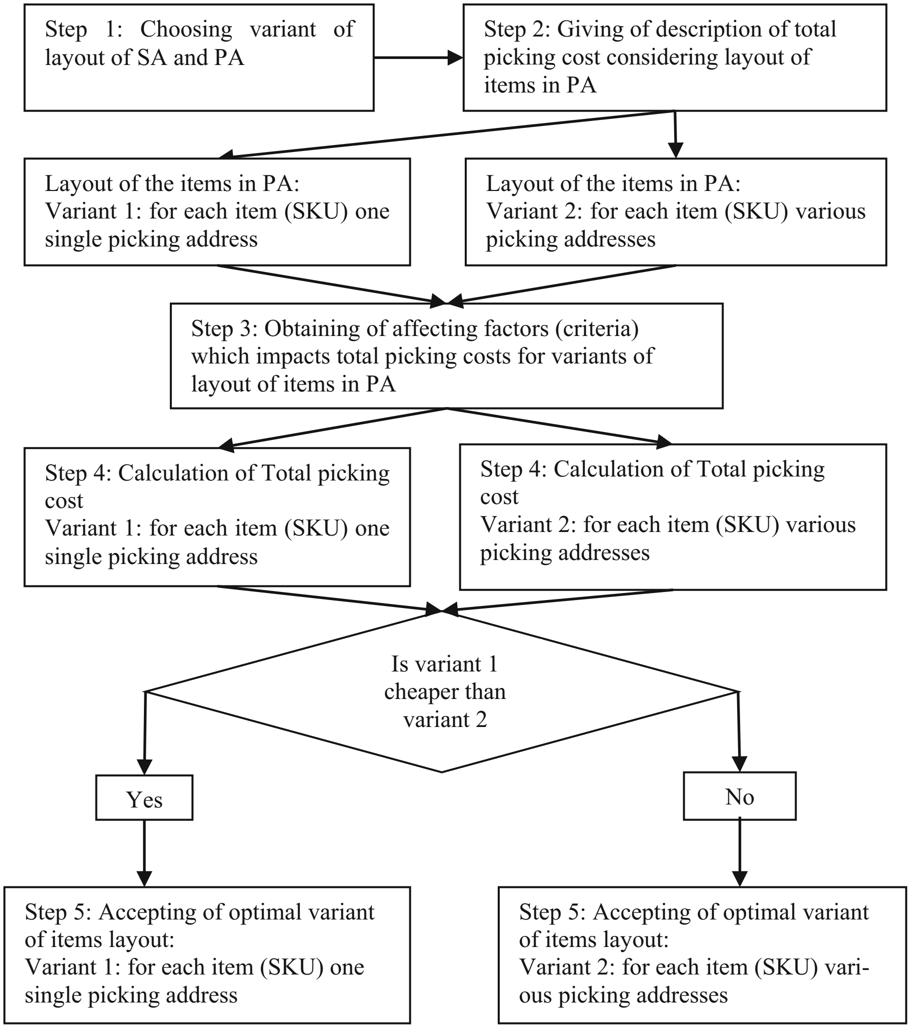


Fig. 1. The methodology of evaluation of total picking cost depends on different variants of layout of SA and PA and different variants of layout of the items in PA (developed by authors).

The additional decision refers to choice of layout of picking addresses for each SKU corresponding with the picking route [7]. The methodology of evaluation of total picking cost consists of 5 main steps presented in Fig. 1:

- Choosing (an optimal) variant of layout of SA and PA.
- Giving of description of total picking cost considering layout of items in PA.
- Obtaining of affecting factors which impacts total picking costs for variants of layout of items in PA.
- Calculation of total handling cost for variants of layout of items in PA.
- Accepting of the decision of optimal variant of layout of SA and PA, and optimal for variants of layout of items in PA.

We can further explore the total handling cost for variant 2, for variant 3 and for variant 4 by comparing two approaches of the layout of items in PA:

- Single picking location for each single item – the replenishment is realised in picking process (see Fig. 2).
- Various picking locations for each single item – the replenishment is realised just only before picking process or after it.

In case of various picking locations for each single item one single or various locations are allocated for each SKU [3]. Fast moving SKU often require multiple pick face locations to ensure sufficient picking stock is available to meet forecasted demand, while slower moving SKU may have only the one pick face location or, in some cases, may only be picked from their reserve location [8].

The first approach defines that the replenishment of definite address is provided once the stock of definite items is below the critical level [3], and concrete inventory stock control method for procurement of goods is used in order to calculate this critical level. For this purpose re-order point (ROP) has been planned, and it depends on safety stock

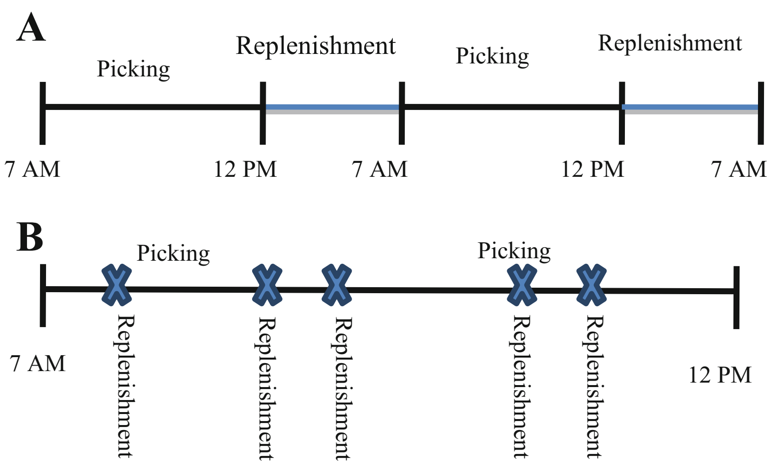


Fig. 2. Approaches of the layout of items in PA: (A) various picking locations for each single SKU; (B) one single picking location for each single SKU.

calculation, considering elimination of the stock-out [9]. For example, the Min/Max stock controlling method can be adjusted for this purpose. It means that replenishment has being realized in uninterrupted picking process. Average replenishment cost could be expensive than for the next approach.

The second approach eliminates replenishment in picking time. If it is not allowed then quantities (stock) of each item in uninterrupted picking process have to be sufficient from the beginning of picking process till the end of it. Automatically it demands various picking addresses for each item, especially, if quantities to be picked exceeds two full pallets. Therefore the replenishment is realized outside of picking process: replenishment of picking addresses has to be before picking process has started.

3 Proposed Parameters for Forming Right Allocation Sequence of SKU

The main emphasis why routing strategies and methods are popular in the practice is diminishing total picking distance either for one single order or batch of orders [10]. Additionally we can aware conditions referring to work by different SKU's. It means the allocation sequence of these SKU, assuming that they are going to pick from PA as HU and CU. These conditions can be expressed by parameters for forming right allocation sequence: as statistical value or measurement metrics.

Authors have proposed parameters for forming right allocation sequence of SKU into picking addresses corresponding by picking route [7]. The parameter 1: the number of orders for each SKU. The parameter 2: the average volume of each SKU per picked order. The parameter 3: the total revenue of each SKU per quarter (forming of quarters depends on seasonal aspect): group A: 80% of total revenue; group B: 15% of total revenue and group C: 5% of total revenue. The parameter 4: the group of brands of SKU. The parameter 5: the key accounts of consumers. The parameter 6: the key accounts of suppliers. The parameter 7: the average turnover of each SKU in a day. The parameter 8: the size of each SKU. The parameter 9: the gross weight of HU or CU.

We would like to stress that this list of parameters does not mean that in warehouse for planning right allocation sequence of items into picking addresses corresponding by picking route only one single parameter could be used. The logical algorithm of forming right allocation sequence of items contains various parameters. It has been proposed by authors.

Next one idea is to develop combined parameter which includes two or more parameters. It could give possibility to choose adequate picking technique referring to its capacity. Also the question of dividing of PA to smaller picking zones could be solved. As result less picking route could be achieved and less total picking cost.

4 Example of Right Allocation of SKU by Number of Orders

Optimisation criteria for right allocation of SKU in the picking addresses in PA is the total travel distance made by picker (warehouse worker) in a day. If picking demand fluctuates in different days of a week then the total travel distance of picker can be measured per month or by seasonal changes – in a quarter. As total travel distance has been diminished the total picking cost and time also could be decreased.

In our example, suppose, that the warehouse has PA which consists of pallet locations as zone included into SA. The 3rd variant of organizing orders' picking process in the warehouse is chosen. In order to simplify process, we are going to assume that M2G picking system is used. Single order is picked by single picker. Additional assumption that picker has to collect 7 orders. Each order contains different SKU's in picking list. Picking zone for this picker consists of following aisles and racks: aisle 01: racks A and B; aisle 02: racks C and D; aisle 03: racks E and F; aisle 04: racks G and H. Information of each order are shown in Table 1. As we see each order demands to collect different handling units (HU) of each SKU. If some SKU is not included in some order then quantity is equal zero. Total picked HU is no criteria for right allocation of SKU's. It means that before right SKU allocation in picking addresses (see column Location in Table 1), the L_1 for this picker equals 1 869 m. Length of rack is 66 m. If we are going to analyze information of orders of each SKU, then some SKU's appears in each order but other ones in each second order or rare.

Table 1. Total travel distance before right SKU allocation in picking addresses (L_1).

Location	SKU	Order_1	Order_2	Order_3	Order_4	Order_5	Order_6	Order_7	Total, HU
A-**0**3	SKU_01	0	0	4	9	12	10	0	35
B-05	SKU_02	1	0	0	0	0	5	0	6
A-17	SKU_03	4	4	0	0	0	0	0	8
B-34	SKU_04	7	5	12	7	8	0	12	51
C-48	SKU_05	3	5	7	3	0	0	5	23
D-23	SKU_06	0	0	0	1	2	0	0	3
D-02	SKU_07	5	7	13	0	0	9	7	41
E-08	SKU_08	0	3	15	8	6	14	5	51
F-13	SKU_09	4	5	7	8	2	2	0	28
F-33	SKU_10	5	0	6	0	4	0	3	18
E-42	SKU_11	0	3	0	8	8	0	2	21
E-66	SKU_12	10	4	14	0	0	16	0	44
G-55	SKU_13	8	7	4	2	0	0	1	22
H-52	SKU_14	0	0	0	4	0	2	0	6
G-21	SKU_15	8	0	0	0	3	1	1	13
G-11	SKU_16	5	0	0	0	0	4	1	10
G-04	SKU_17	0	1	4	5	0	0	1	11
H-01	SKU_18	8	3	12	3	2	4	14	46
L_1, meters		**267**	**267**	**267**	**267**	**267**	**267**	**267**	**1869**

Authors prefer to use parameter for optimizing total travel distance of single picker. The parameter is parameter 1: the number of orders for each SKU. As it is mentioned above – in other situations we can use other parameters. For example, if picking are realized in CU, and simple racks of carton boxes are located in PA, sizes of cartons of SKU's are used as parameter for right allocation of SKU's.

In order to get right sequence of allocation of these items at first we need to sort all SKU's by the number of orders. Then first address in the picking route will be engaged by the SKU with the biggest number of orders, but in the end of picking row we will find SKU which is less popular in all orders.

As result we receive new allocation sequence for all SKU's which take participation in picking process. In Table 2 is shown new addresses of SKU's of 7 orders. Off course, we are not going to show new picking locations for 1 000 SKU's or maybe more, but we can simple make analysis for one single picker for picking 7 orders. In Table 2 also is presented all SKU's by number of orders (O_num). It means that after right SKU allocation in picking addresses (see column Location in Table 2), the L_2 for this picker equals 917 m.

Table 2. Total travel distance after right SKU allocation in picking addresses (L_2).

Location	SKU	Order_1	Order_2	Order_3	Order_4	Order_5	Order_6	Order_7	O_num
A-03	SKU_18	8	3	12	3	2	4	14	7
B-05	SKU_04	7	5	12	7	8	0	12	6
A-17	SKU_08	0	3	15	8	6	14	5	6
B-34	SKU_09	4	5	7	8	2	2	0	6
A-44	SKU_05	3	5	7	3	0	0	5	5
B-46	SKU_07	5	7	13	0	0	9	7	5
B-52	SKU_13	8	7	4	2	0	0	1	5
A-53	SKU_01	0	0	4	9	12	10	0	4
A-65	SKU_10	5	0	6	0	4	0	3	4
B-65	SKU_11	0	3	0	8	8	0	2	4
C-64	SKU_12	10	4	14	0	0	16	0	4
D-64	SKU_15	8	0	0	0	3	1	1	4
D-43	SKU_17	0	1	4	5	0	0	1	4
C-39	SKU_16	5	0	0	0	0	4	1	3
C-37	SKU_02	1	0	0	0	0	5	0	2
D-12	SKU_03	4	4	0	0	0	0	0	2
D-09	SKU_06	0	0	0	1	2	0	0	2
C-01	SKU_14	0	0	0	4	0	2	0	2
L_2, meters		**131**	**131**	**131**	**131**	**131**	**131**	**131**	**917**

5 Conclusions

A choice of appropriate parameters and forming logical algorithm is unequivocally. It depends on picking systems, storing systems, from the speed of the turnover of each SKU, etc. Therefore initially stage is logistics audit of definite warehouse. The results of each warehouse could be different.

Our example of right allocation of SKU in the picking addresses is based on analysis of the number of orders for each SKU. As optimisation criterion the total travel distances before and after right SKU allocation in picking addresses have been evaluated. Results of research showed that the total travel distance for one single picker could be diminished if items will be replaced in sequence of number of orders. Before right allocation the L1 equals 1869 m, but after replacement L_2 equals 917 m. The result after right allocation of SKU's in picking addresses could be different in different warehouses, and it depends of replenishment process as well as.

Also replenishment is one of important processes in the warehouse for realizing pallet transportation process from SA to PA. Very essential is provision of stock availability in uninterrupted picking process. Replenishment process can be affected by approaches of the layout of items in PA, by replenishment heuristics, by choice of several variants of organizing orders' picking process in the warehouse, proposed by authors, by several methodologies for warehouse replenishment systems and, finally, by routing of picking process.

There are some serious questions which will be as guidelines for building right allocation sequence of items into picking addresses corresponding by picking route:

- What is level of similarity of SKU?
- Are items intended or ordered by few clients or delivery points?
- What kind of warehouse systems and racks will be planned in PA?
- What is the impact of size, weight and volume of each SKU?
- What is capacity of picking cars used in picking process?
- What is the impact of turnover of each SKU?

References

1. Tompkins, J.A., White, J.A., Bozer, Y.A., Tanchoco, J.M.A: Facilities planning. 4th edn. Wiley, UK (2010)
2. Harper, J.: Distribution centre order picking technologies compared, http://www.mwpvl.com/html/order_pick_technologies.html
3. Apsalons, R.: The Management of Logistics Centres (in Latvian). Burtene, Latvia (2012)
4. Rushton, A., Walker, S.: International Logistics and Supply Chain Outsourcing. From Local to Global (2007)
5. ERIM. Routing strategies, https://www.erim.eur.nl/centres/material-handling-forum/research-education/tools/calc-order-picking-time/what-to-do/routing-strategies/
6. Dukic, G., Oluic, C.: Order picking methods: improving order picking efficiency, https://www.fing.edu.uy/inco/eventos/icil05/03-wed/F1-Dukic.pdf
7. Apsalons, R., Gromov, G.: Using of logistics' principles for picking items in railway warehouses. In: Development of Infrastructure and Logistics Technologies of Transport Systems, International Scientific – Practical Conference, pp. 93–100. Sankt-Petersburg (2015) (in Russian)
8. Dematic Global Website. Replenishment, http://www.dematic.com/en/supply-chain-solutions/by-supply-chain-function/replenishment

9. Lukinskiy, V., Lukinskiy, V.I.: Evaluation of the influence of logistic operations reliability on the total costs of supply chain. Transport and Telecommun. **17**(4), 307–313 (2016)
10. Won, J., Olafson, S.: Joint order batching and order picking in warehouse operations. Int. J. Prod. Res. **43**(7), 1427–1442 (2005)

Intelligent Transport Systems

The Main Challenges of Winter Road Service to be Solved Within the Framework of Intelligent Transportation System

Boriss Jelisejevs(✉)

Institute of Transportation Structures, Riga Technical University, Kipsalas 6B/6A, Riga 1048, Latvia
b.jelisejevs@gmail.com

Abstract. Winter road maintenance as a set of routine network operations has a huge influence on mobility in snowy regions, providing certain safety level for road users and even possibility of several sections in critical weather conditions. There are still a lot of challenges to optimise the performance of winter road works, as well as to set more appropriate levels of service, defining and controlling the balanced scope of key performance indicators (KPI). Intelligent transportation system (ITS), as a distinctively innovative, but at the same time already quite mature sector, might bring a complex cost effective solutions to the field. Some of them could be distinctively opportunistic, thus to be introduced and operated in low-cost manner, based on reuse of the existing data (integration) and extensive feedback from road users. Traffic information, road weather information system (RWIS), management decision support systems (MDSS) for winter road operations and sectoral subprocesses are in the scope of the proceeding. Although this topic has a global meaning, the discussion is oriented on the actual needs of Latvian road industry with scope of national road network.

Keywords: Intelligent transportation systems (ITS) · Traffic information
Road winter service

1 Introduction

Winter road service is a set of seasonal activities with the main task to ensure the usability of road network, affected by weather. These works (snow clearing from road, spreading with anti-skid materials a.o.) are included in the routine maintenance programme of network operations and covers the period from November to March in Latvia. Meteorological conditions and territories, affected by them are dynamically variable factors, making actual driving conditions changeable, as well. Comparing with another types of road works, interventions of winter service have vast and short-term effect on mobility and traffic safety. Winter service is not an investment in infrastructure at all, or some operations (use of deicing chemicals) even foster deterioration of road structures. Mobilization of resources (readiness to act), ultra short time for provision and near real-time decision making, make winter road service similar to emergent actions also.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_34

Taking into account the absolute necessity of winter road works and willingness to spent for them as less, as possible and also mitigate any related by-effects (f.i. environmental and corrosive impact of road de-icers), lead to a broad scope of optimization tasks, to be solved within the integrated model of service provision. Functional breakdown is needed, to deal with such aspects, as: level of service (requirements, technological scope, available resources); business model (type of contract, institutional responsibilities), KPI.

2 The Existing Model of Service Provision

Winter works' budget of Latvian national roads for the last 5 years annually is about 20.5 milj. Eur or around 30% from all the allocations, aimed for road routine maintenance [1]. This is still in-house service (not contracted out for the market), that's why it is strongly related to legacy approach, where innovations are not widely used yet. Another road operators (mainly local municipalities) take care about their networks and especially big cities are facing with even higher and intensively customized winter service requirements, than Latvian road industry in general. In spite of quite well developed regulations and management processes, provisional criteria of service here still might be dealt subjective in comparison with such thematically close, but much more discrete process, as construction works. The latest legislative incentives introduced proactive strategy (preventive anti-skid treatment of TEN-T roads) in 2014, however there is not enough knowledge and tools to bring such method to practice, e.g. to be sure that decision-making is optimal for all the valid cases. The service is provided by state owned company "Latvian road maintainer" on the base of 7-year contract (2014–2020). Institutional model is quite unusual in the field, because this commercial agreement is a straightforward delegation from the Ministry of transportation, which co-exists with delegation of national road administration to manage the road network.

Actual technical regulations are adopted from Scandinavian experience of 1990-ties. The main differences between service levels are marginal requirements for the state of the pavement, time limit to eliminate incompliance, caused by weather and daytime period of level to be in force (Table 1). In practice it means, that for A and A1 classes "bare pavement policy" will be provided with intensive use of snow melting agents, while B class allows cheapest anti-skid treatment, but C and D classes guarantee only physical throughput, without anti-skid treatment (Fig. 1). Snow clearing measures for roads of D class are generally performed not often than 4 times per winter. So, generally existing provisional strategy is reactive, however some administrative incentives bring also proactive operations (preventive salting all the daytime for certain weather conditions) to the field. As this one is a top-level strategy, that for proper application is to be based on complex supportive measures (diversification of technologies, use of highly reliable MDSS in forecasting mode etc.), that can't be effectively initiated only by administrative acts. Visual observations and evidence data of performed actions are the main indicators of service's performance control, however they have no uniform formal methods for objective fixation and evaluation.

Table 1. List of the main requirements for classes of road winter service in Latvia.

Requirement	A (A 1)	B	C (D)
Max. snow depth on carriageway (in stable weather)	not allowed	4 cm	10 cm (no limit)
Max. depth of fresh snow on carriageway (in changeable weather)	6 cm	8 cm	10 cm (no limit)
Max. height of snowbanks on shoulder	40 cm (60 cm)	80 cm	100 cm (no limit)
Time limit for snow clearing	3 h (4 h)	6 h.	18 h (no limit)
Time limit for deicing	3 h (4 h)	6 h	no limit
Daytime period, when class is in force	06–22 (06–20)	06–18	06–18 (no limit)

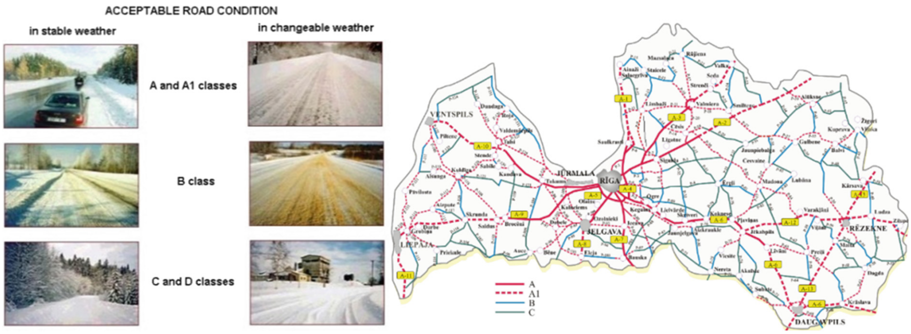


Fig. 1. Visual description of winter service levels (left) and network coverage (right).

The main points to be considered for the discussion here are:

- the present level of service is adequate enough, because seasonal distribution of traffic accidents is quite even;
- road budget is restricted and there are no indications to stable growth (unfair public pricing scheme of industry, demographic decrease, growing difference in regional development, finish of EU co-financing till 2020. and bad overall condition of the network, asking primarily to invest, not operate), so allocations for winter service have indications to decrease till the minimal acceptable level further;
- public road network must be re-arranged, changing historical jurisdiction to the most appropriate managemental scope now, giving up to half of local national roads to the municipalities (however, this will not solve financial problems, because a majority of countryside municipalities have no own resources for these roads);
- provisional methods (strategies and their supportive technologies) should be extensively combined and covered by qualitative MDSS for more effective use within the network and control procedures;
- level of services should be widened to cover all the needs within the network, making more gradual changes, when necessary (f.i. so called Adaytime superclass is proper only for some stretches of strategic meaning, not for all the TEN-T network).

3 Operational Methods

According to the service levels, a set of technologies are in use there (Table 2). Patrol operations for A and B levels are to be done by truck mounted equipment, which covers snow cleaning and universal spreading (pre-wetted salt or abrasive) abilities.

Table 2. De-icing technologies specified for Latvian state road network.

De-icing technology (by used material)	Consumption of material at a single treatment, g/m^2	Effective in climatic conditions
sand/salt mixture (9/1)	190–320	–6 °C > t > –10 °C, continued precipitations
sand or crushed aggregate	320	t < –10 °C
Prewetted salt (NaCl)	5–30	t > –10 °C, black ice, freezing rain, frost, snow
Solution(NaCl, $CaCl_2$)*	15	t > –3 °C, black ice, frost
Ice grinding **	–	t < –8 °C, snowpack on the road

Remarks:
* – de-icing with liquids is specified, but not really experienced in Latvia;
** – ice grinding is in very limited use on local roads due to unstable winter conditions.

The spreaders working for A/A1 class roads are equipped with industrial fleet management devices, where GPS tracking data and equipment status are automatically captured and transmitted to the central system. There are still problems to use this system wider (also to count actions for billing and optimization of spreading routes) due to some existing barriers. First of all, available geospatial network's metadata (required also by INSPIRE directive) is not enough precise to built in routing analytics. The other point is quite broad deviations for spreaders' automatically captured data (f.i. snow showel's status) experienced in practice, that's why such automation still is combined or even doubled with formal documenting procedures.

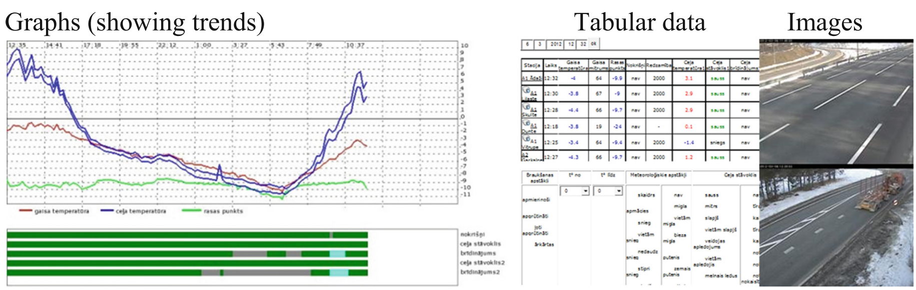


Fig. 2. Screenshots from winter staff's webpage on RWIS data.

RWIS consists of 55 stations along the national main roads with a step of 20–40 km (Fig. 2). It is not enough to cover sensitive meteorological locations and to provide measurements for network-wide reference points, which is necessary for future RWIS → MDSS functionality (thermography and sections' monitoring, integrated forecast of road conditions etc.). That's why actual plans consider to double number of stations and widen range of data gathered there [2].

Actually, there are basic alarm indications on hazardous situations, when pavement tends (by linear extrapolation) or are already slippery (stations' embedded) and no RWIS data technical integration with other sources of meteorological information, to provide automated accurate forecast of changeable road situation in short-term perspective (up to 6 h). However in recent years some pilot projects were initiated and prototype of MDSS, based on open METRO model (forecast of road conditions here is based on heat flow, therefore direct measurements beneath the road surface are crucial here) was introduced in partnership with national meteorological agency, now these R&D activities are interrupted, as lack of coordination between the bodies and critical problems of RWIS data quality still exist [3].

The formal base of decision making in road winter service is a tour of duty, but RWIS and other external data give additional input there (Fig. 3). The detailed route-oriented description of actual driving conditions is also public traffic information, produced here.

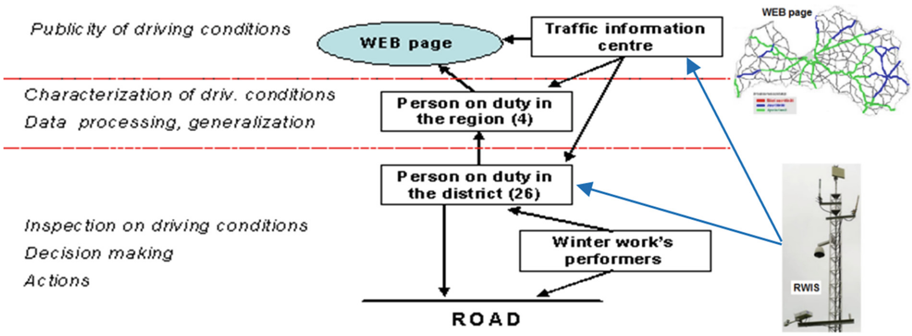


Fig. 3. The structure of the tour of duty for road winter works.

Looking for environmental issues, climate is definitely changing, with direct influence on road winter service. However winters become milder in general, meteorological fluctuations are higher than before: more precipitations and their uneven distribution, grow in marginal cases, when air temperature is going around zero degrees, causing road slippery more often. This also affect drivers' awareness, because amount of "black ice" cases is growing, which appearance is distinctively sudden, randomly localized and not obvious for road users. Accordingly to RCA4 model simulations, period of turbulence, where forecasted climate warming don't mean decrease in road slippery cases in the region may continue up to 30 years [4]. Only later climatic trends will definitely result in significant and consequent reduction of winter road service (Fig. 4).

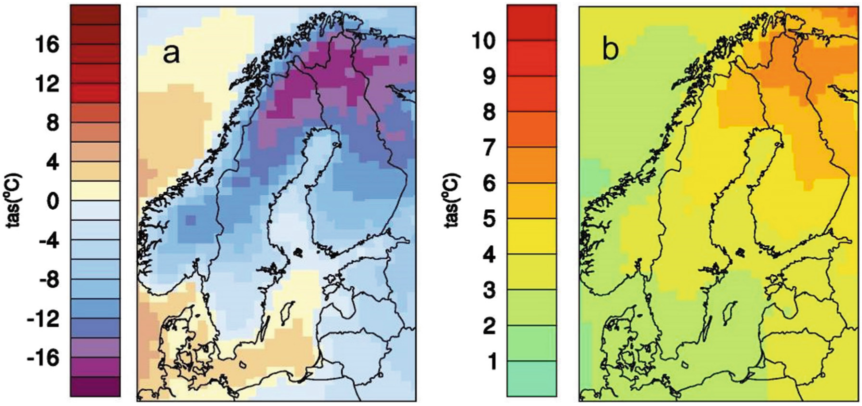


Fig. 4. Data on average air temperature in winter, a - observations 1971–2000, b – RCA4 simulations for 2100 (temp. growth to the reference period). Source: [4].

The main aspects of actual operational methodology are:

- methods and especially decision making tools are not levelled to the latest administrative incentives, bringing to the field top-level proactive service's strategy;
- there is still modest accumulation and attempts for multifunctional use of the related data (as it should be);
- climate changes currently bring additional variables to winter road service, however the situation is much far from the simplification, as: warmer winters mean less works to do.

4 ITS Use for Road Winter Service

ITS already has a stable scope and clear development paradigm on pan European level (action plan and Directive 2010/40/EU), that meets the abovementioned challenges of road winter service also, bringing additional data and analytics to the field. As ITS services are directly aimed to support road users and provide adaptive traffic management firstly, auxiliary processes may derive useful items, arisen from such principal topics as: big data, open data, indirect detection methods and industrial automation. Up to now, ITS is characterized with a group of isolated projects with weak mutual synergy in Latvia [5]. Lack of common view and national strategy (will be introduced in 10.2017.) is the main barrier for fruitful multi institutional collaboration in the field. Being transportation subsector, ITS has a state specific hierarchy of tasks and services. Looking from the global perspective, ITS as ICT driven domain, has clear vector of development and the foreseen schedule to bring road transport up to the completely new mobility level (Fig. 5).

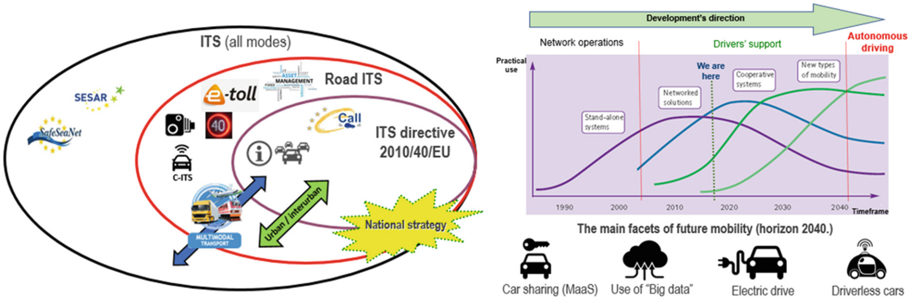


Fig. 5. General ITS framework (left) and global trends in road transport, directly linked to ITS domain (right).

The following ITS application areas were officially set up for the harmonized deployment within the EU: traveler information, traffic management, electronic pricing and payment, freight and logistics, vehicle safety systems and ICT infrastructure [6]. Uniform architecture FRAME was established there to help in avoidance of any misunderstanding and overlapping, when principal managing decisions are to be made. According to that, ITS service areas have mutual relations (including crossing nodes) and are linked to the main types of system's interfaces (Fig. 6, left). FRAME scope is technology independent to be feasible for different kinds of system design (legacy base, priorities, business models), that are actually exist or possible in future (Fig. 6, right).

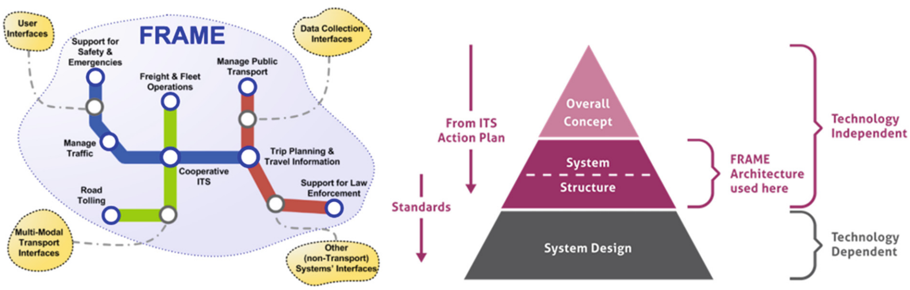


Fig. 6. EU ITS architecture FRAME, source: Frame forum.

Taking into account budgetary constraints, low cost approach to ITS is highly feasible, where total cost ownership is mainly bearing by service provider, as revenue is generated indirectly. In general, such service is more oriented on drivers, who give back personalized input (let to track themselves and put any related notifications). Road authorities, in their turn, should share their own business data to the service platform, getting processed (anonymised. accumulated etc.) data, useful for network operations. One such example in Latvia is partnership of national road administration with Waze service (connected citizen program's agreement), started in 2014 (Fig. 7). The main aspect of success here is high popularity of Waze in Latvia, whereas up to 70 000 drivers use this service on daily basis, simultaneously enriching it with their own data [7]. There are on-going discussions to add there live data on active spreaders' location, however methodology of presentation such data still is to be well-considered.

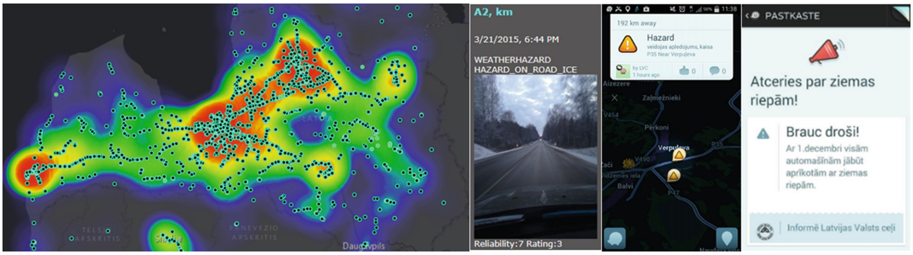


Fig. 7. Waze as an example of a platform to share winter road data between road authorities and users.

The main pillars are to be clearly identified for proper and long-lasting service provision throughout the low cost business model, are following:

- provision of wide access to existing public transport and other mobility related geospatial data, according the concept of digital community;
- definition of KPI to set up basic quality requirements of service;
- definition and control of quality to be sure on service's provisional quality (stability of broadcasting, recovery issues etc.);
- extensive cost sharing within the established partnership (hardware, liability, data fusion, etc.) to make possible multifunctional use of the existing infrastructure and data and reach the users through the existing interfaces (f.i. smartphones or on-board devices);
- provision of service, binding to low cost approach (often with backward data input from users);
- commercialization node, which must not discriminate the users' rights and to be in accordance with data privacy and other legal issues.

From technological point of view, actually IoT (internet of things) leads to broad availability of solutions, based on "Big data", providing even more thankful landscape for cross-sectoral data integration and development of low cost ITS services. A simple smartphone nowadays provides a quite wide combination of various sensors and automation technics, to be used as feasible data input to ITS. With respect of mobility needs, IoT scope might generally provides such set of captured data from in-vehicle systems or nomadic devices:

- tracking of device location is the first one practical IoT input for any ITS applications (FCD, data aggregation to define traffic flows etc.);
- device operational status by itself, which can signalize about certain situation to be interpreted (big potential here is for use of vehicular intrinsic data, to be converted into characterization of traffic conditions);

- mobile probes, which make add-on to device status data, when are available and active (big potential of collect additional target inputs from customers along the business model of a service).

Looking for more users' input to the system, as the main added value and a compensation for service's use, they are able to give also direct multimedia feedback to the applications on any classified cases (f.i. hazardous conditions). Moving towards future mobility, development of underlying ICT will definitely lead to even growing opportunities to put in place low cost approach. That means, if ITS services or their basic technologies will need even more assignations to develop, their core is to be rather software, than hardware oriented, making less costly support (thus also use) of such products during their lifecycle.

Table 3. Use of low cost approach for ITS service areas of EU ITS action plan.

Service area	Low cost approach	Technologies behind	Examples of services
Traveller information	Highly feasible along the whole value chain	FCD (floating car data), direct data input from users, broadcasting to the existing user interfaces (devices)	Waze – interactive navigation with integrated elements of drivers' social network
Traffic management	Feasible, when multifunctional use (cost sharing) of provisional resources is effectively applied	Virtualization of traffic control (f.i. VMS); indirect benefit from low cost traveler information services or mixing both together	In vehicle traffic signage (C-ITS); use of roadside commercial digital billboards
Electronic pricing and payment	Feasible at the several stages, mainly for supportive actions, evidence control	FCD of winter fleet, in-depth data integration with GIS and cross-sectoral information	Novasib (GE, full functionality) – automated register of winter fleet operations.
Freight and logistics	Feasible for many needs (also client's specific)	Virtualization of formal procedures; integration of data from different sources along the certain processes	GoSfift – remote reservation, virtual queue and additional info for EE/RU border crossings
Vehicle safety systems	Economically feasible as an on-board fusion of detection and driving assistance methods	Multifunctional detection (primarily intelligent video image processing), avoiding excessive use of various sensors. Wide range of proprietary developments by car makers	Way towards autonomous driving, where R@D actions are resource consumptive, but potentially lead to nonexpensive in-built industrial systems
ICT infrastructure	Highly feasible, when transform legacy systems and bring new detection methods to road management	Cheap (opportunistic) sensing methods, mixed with geotagging and decision-making algorithms	Different types of MDSS: Vionice (FI, simplified road inventory, using smartph/); Teconer (FI, mobile measurements of road conditions)

Data aggregation from various sensors and extremely high need of customers' feedback (or any specific active input) directly influences the final performance of traffic information services, that's why low cost provisional schemes and avoidance of direct pricing is a logic step towards viable business models, based on "win-win" result for all the involved parties. As it was noticed before, low cost approach is feasible, when added value of new functionality is generated from synergy of the existing processes rather, than from straightforward investments. That disclosure and potential for the main ITS service areas related to road winter works is summarized in Table 3 (on example EU defined ITS service areas).

For sure, any related opportunities are highly feasible, to bring added value to road winter service. Thus, Latvian and Estonian road administrations is implementing EU co-financed (ERAF, Interreg Central Baltic program) ITS project SmartE67 [8]. It is oriented on traffic information and management, firstly, but bring to the field multi-functional solutions also, to be used for winter road service also (installation of new and modernization of existing road weather stations, pilot installations of variable message signs, improvement of traffic information centres a.o.).

5 Conclusions

There are a lot of cost effective opportunities to bring more functionality for road winter service, using ITS. Besides classic measures, aimed to level up existing RWIS and making core MDSS for the further customization, the main directions can be summarized as following:

- as wide as possible open access to public transportation data, making possible it's intensive integration an provision of added values;
- development of the existing fleet register systems and implementation of mobile measurements (probe data) to rise evidence of the performed actions and quality control;
- provision of users' real-time feedback on driving conditions, as additional source of information for decision making.

References

1. SJSC Latvian State roads: The annual report of national road network. LSR, Riga (2016)
2. Jelisejevs, B., Brivulis, I.: Further development of RWIS and it's new potential marketplaces in Latvia. In: Proceedings of 16th SIRWEC international road weather conference, Helsinki, pp. 1–7. e-edition (2012)
3. Jeļisejevs, B.: ITS tools for winter road maintenance in Latvia. In: Scientific Proceedings of RTU, vol. 11(2), Riga, pp. 83–87 (2010)
4. Kjellstrom, E., et al.: Production and use of regional climate model projections – A Swedish perspective on building climate services. Clim. serv. 189–198 (2016)
5. Yatskiv, I., et al.: Review of intelligent transport solutions in Latvia. Transp. Res. Procedia. **24**, 33–40 (2016)
6. European commission: ITS action plan. EC, Bruxelles (2008)

7. Guevara, G., Jelisejevs, B., et al.: Advanced technology for data collection and information to users and operators. Paris, PIARC report, pp. 1–45 (2016)
8. Jelisejevs, B.: Implementation of SMART E67 project. In: Proceedings of 29th International Baltic road conference, Tallinn, pp. 1–8. e-edition (2017)

Effective Wireless Communications for V2G Applications and Objects in Motion

Aleksandr Krivchenkov[1(✉)], Alexander Krainyukov[1], and Rodion Saltanovs[2]

[1] Transportation and Telecommunication Institute, Lomonosova Str. 1, Riga 1019, Latvia
krivchenkovs.a@tsi.lv
[2] TransfoElectric SIA, Riga, Latvia

Abstract. The communication network structure for Vehicle to Grid (V2G) systems is derived on the bases of ISO/IEC 15118 standard. If Wireless Power Transfer (WPT) technology is used and vehicles are in motion wireless data transfer network must present and restrictions for data packets delivery exist. We analyse requirements for efficient compensation process and its influence on requirements of data transfer. For wireless technologies 802.11, the estimations of delays of packets are made on the basis of analytical and simulation models.

Keywords: V2G · WPT · Wireless communication · Packet delay · 802.11

1 Introduction

When Wireless Power Transfer (WPT) technology is used for contactless energy transfer to vehicles the energy transfer efficiency is not constant, since the air gap between the EV chassis and the ground and horizontally misalignment of energy transfer system coils are changed [1]. To ensure a constant of the energy transfer efficiency special designs of coils are created [2]. Another approach is to change the parameters of the transmitting part of the system so that the energy transfer efficiency is constant when the vehicle is moving [3]. In this case, wireless data transfer channel between vehicle and grid is necessary.

The nature of links based on the radio channel and the access to this shared resource cause variable available bandwidth, variable packet delay and packets loss rate. Real-time control processes in WPT-V2G system bring forward the demands to wireless links parameters for data transfer. A good estimation of the network latency together with network bandwidth will facilitate robust network designs.

In this paper compensation process for WPT efficiency give the restrictions for delay of packets in the network. The approach in estimations follows the one described in [4] and simulation model for data link essentially includes the influence of contention mechanism for radio recourse in 802.11 links and different possible data rates for different signal to noise ratio (SNR).

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_35

2 Real-Time Control Processes in WPT-V2G System

2.1 Compensation Processes and Energy Transfer Efficiency

For WPT and ICPT (Induction Coupled Power Transfer) "compensation" mechanism can be used. In [5] due to the load resistance changes with a time the compensation process is introduced. The process is as follows: periodically measuring voltages on the load in secondary side and transfer these values in digital form (sampling) wirelessly to the primary side of the system to adjust the generated pulses (Fig. 1).

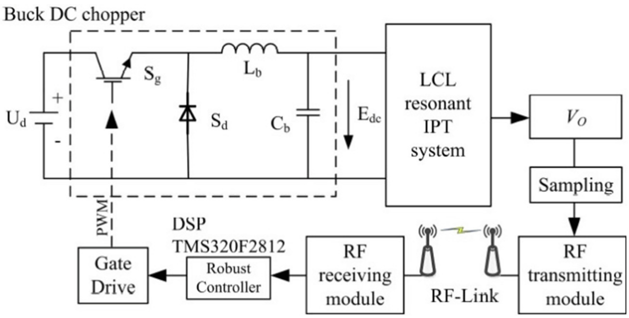


Fig. 1. The functional diagram of the system in which ICPT, the load resistance compensation loop and the wireless transmission channel are used [5].

For ICPT systems the power transfer efficiency is calculated on the base of two ports network with one port being input fed by the source and the other fed by the load.

The function of power transfer efficiency is defined as $\left|\eta = \dot{S}_{21}(\omega)^2\right|$, where the scattering parameter $\left|\dot{S}_{21}(\omega)\right|$ is given by:

$$S_{21} = 2\frac{V_L}{V_S}\sqrt{\frac{R_S}{R_L}}, \tag{1}$$

where R_S is the source impedance, R_L is the impedance of the load, U_L is voltage on load and U_S source voltage. Scattering parameter $S_{21}(\omega)$ is a function of frequency, the elements parameters of the primary and secondary sides, as well as the coupling coefficient of the transmitter and receiver coils. For series – series topology transfer function is represented by the following equation:

$$\dot{S}_{21}(\omega) = \frac{2j\omega k\sqrt{L_1L_2}\sqrt{R_LR_S}}{\left(j\omega L_1 + \dfrac{1 + j\omega C_1R_s}{j\omega C_1}\right)\left(j\omega L_2 + \dfrac{1 + j\omega C_2R_L}{j\omega C_2R_{L1}}\right) + \omega^2k^2L_1L_2}, \tag{2}$$

where L_1 and L_2 are the self-inductances of the primary and secondary side, C_1 and C_2 are the capacitances, and k is the coupling coefficient of the primary and secondary coils

[6]. The coupling coefficient is expressed as $k = M/\sqrt{L_1 L_2}$, where M is mutual inductance of the transmitter and receiver coils. The value of M and thereby k depends on the physical parameters. Circular planar coils are most frequently used for electric vehicles charging, the coupling coefficient of such coils can be written as:

$$k = \left[1 + 2^{2/3}\frac{h^2}{r_t r_r}\right]^{-3/2}, \tag{3}$$

h denotes the distance between two coils (air gap), r_t and r_r are the radii of the transmitter and receiver coils.

In our numeric experiments [4] the transfer function $|\dot{S}_{21}(\omega)|$ was plotted (Fig. 2) around 85 kHz for variable distance between coils. In this case coils inductances were 40 μH (L_1) and 100 μH (L_2); capacitances were 90 nF (C_1) and 36 nF (C_2). Such values provide the same self-resonant frequency of the loops 83.882 kHz. The source impedance R_S was 0.5 Ω, the load impedance R_L was 7 Ω. The radius of the coils was 200 mm and coupling coefficient was changed as (3).

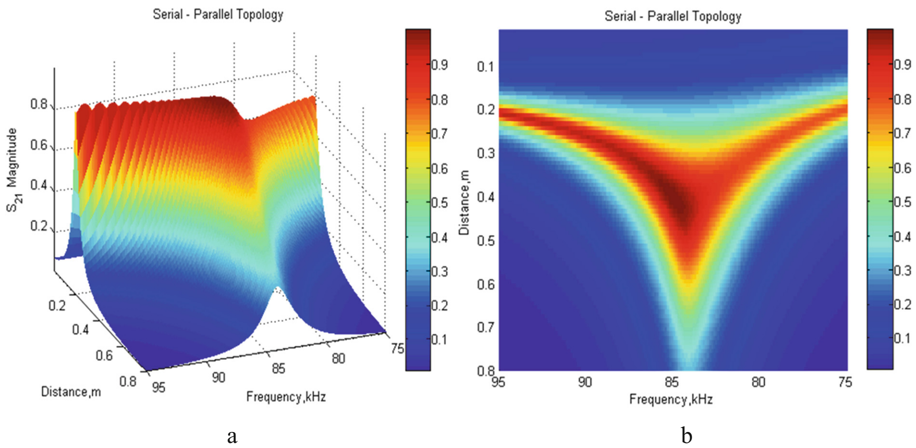
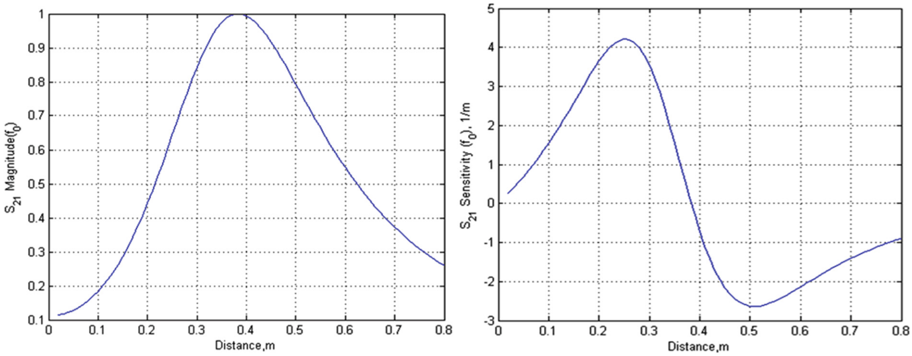


Fig. 2. $|\dot{S}_{21}(\omega)|$ as a function of frequency and distance between coils (a – 3D, b – 2D-view).

In Fig. 2a frequency splitting is clearly visible when the air gap is decreased. As the coupling between the coils decreases, the frequency separation also decreases until the two modes converge at f_0 (83.882 kHz). This point is called the critical coupling point k_{cr} and it represents the optimal distance $h_{opt} = 0.38$ m ($k_{cr} = 0.06$) at which maximum power efficiency is still achievable (see Fig. 2b). This is clearly seen in Fig. 3.



(a) (b)

Fig. 3. $|\dot{S}_{21}(\omega)|$ magnitude (a) and sensitivity (b) at the resonant frequency f_0.

Figure 3a shows the cross-section of the three-dimensional image shown in Fig. 2a when the frequency is 83.882 kHz because the principal condition for maximum power transfer in ICPT system is operating on the resonant frequency.

When we consider changes in the distance between two inductive coils (it leads to coupling coefficient changes) appropriate changes in operational frequency must be done to save the efficiency in energy transfer. Figure 3b depicts the variation of sensitivity of WPT efficiency for series – series topology at the resonant frequency f_0 as a function of the variable distance between two coils at resonant frequency f_0. This dependence is defined by differentiating of (1) and (2) expressions by the distance between the transmitter and receiver coils [4].

2.2 Examples Clearance and Misalignment Influence on Efficiency

When the EV is moving and charging the air gap between the EV chassis and the ground, horizontally and longitudinal misalignment of coils may change the mutual inductance and thereby coupling coefficient, consequently it would affect the power transfer efficiency.

As an example when compensation is necessary in [4] we consider moving EV clearance changes:

$$h(t) = h + h_m \sin\left(2\pi f_g t\right), \tag{4}$$

h is not moving vehicle clearance, h_m is maximum amplitude in clearance, f_g is the cyclic frequency of changes.

If f_g and h_m are known and the allowable change in the energy transfer efficiency ΔS is specified, the time interval required to adjust the operating frequency of the WPT according to [4] is:

$$\Delta T_{cycle} \leq \frac{\Delta S}{S_{max}(h_m) \cdot h_m \cdot \omega_g}, \tag{5}$$

where $S_{max}\,(h_m) = \max(|S(h_{opt} - h_m)|, |S(h_{opt} + h_m)|)$.

For $h_m = 0.05$ m and simulated WPT system: $h_{opt} = 0.38$ m, $S(h_{opt} - h_m) = 2.7\ \mathrm{m}^{-1}$, $|S(h_{opt} + h_m)| = 1.7\ \mathrm{m}^{-1}$, and consequently $S_{max}\,(h_m)$ is equal to $2.7\ \mathrm{m}^{-1}$. The recommended value ΔS is equal 0.01. For these values and $f_g = 1$ Hz, the upper limit of ΔT_{cycle} is equal to 11.8 ms. This value essentially defines the upper limit for the duration of compensation cycle in compensation process that we need to have if the power transfer efficiency needs to be saved even when vehicle clearance changes.

Misalignment between the transmitting and receiving coils also decreases the coils coupling coefficient. The mutual inductance for circular coils with lateral misalignment is given by Eq. [7]:

$$M = \mu_0 \pi r_t r_r \int_0^{\infty} J_0(sd) J_1(sr_r) J_1(sr_t) exp(-s|h|) ds, \tag{6}$$

where J_0 is the Bessel function of the first kind of zeroth order, J_1 is the Bessel function of the first kind of first order, s is the integration variable, r_t is transmitting coil radius, r_r is receiving coil radius, h is coils distance, d is coils misalignment.

To obtain an analytical expression describing the effect of the relative coil misalignment on the coupling coefficient, coupling coefficients were calculated under the following conditions: h was equal to the optimum $h_{opt} = 0.38$ m at the critical coupling coefficient $k_{cr} = 0.06$ (Fig. 3) in the absence of coil misalignment; coils radii were the same $r = r_t = r_s$ and equal h_{opt}; relative coil misalignment $d_r = d\,/\,r$ was variated from 0 to 1 with a step equal to 0.1.

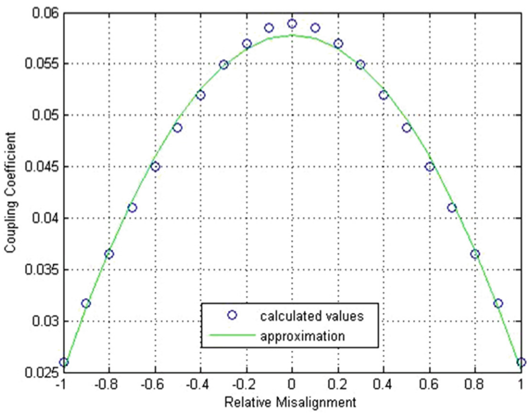


Fig. 4. The effect of the coils relative misalignment d_r upon coupling coefficient k.

Figure 4 depicts the variation of coupling coefficient as a function of discrete relative coil misalignment d_r (this is shown by circles). The Matlab's "polyfit" function was used to approximate this set of values. The "polyfit" uses a data set to determine the coefficients of the polynomial $p\ (y)$ of degree n, which approximates a data set in the sense of the least-squares method. For a polynomial of the second degree ($n = 2$):

$$k(d_r) = a_2 d_r^2 + a_1 d_r + a_0, -1 \le d_r \le 1. \tag{7}$$

Coefficients were determined as: $a_2 = -0.0327$, $a_1 = 0$, $a_0 = 0.0578$.

The function $|\dot{S}_{21}(\omega)|$ was plotted (Fig. 5a) around 85 kHz for the above conditions. It can be seen that the transfer function $|\dot{S}_{21}(\omega)|$ is slightly less than 1 when the coils relative misalignment d_r does not exceed $|0.2|$.

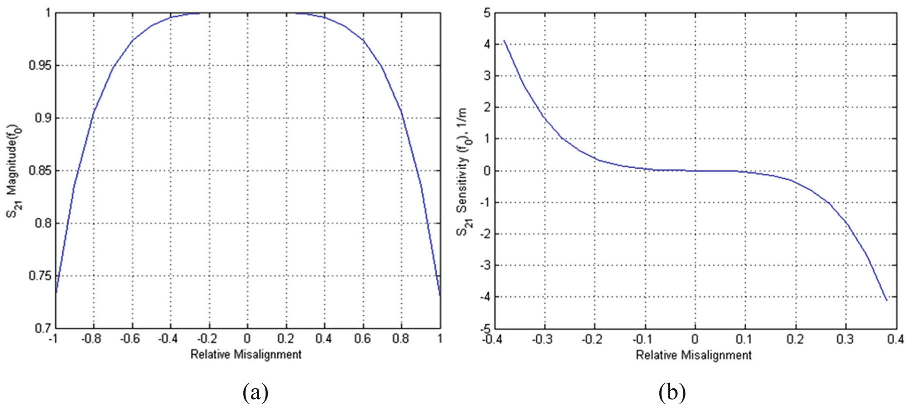


Fig. 5. $|\dot{S}_{21}(\omega)|$ magnitude (a) and sensitivity (b) as a functions of misalignment.

When the EV is moving the relative coil misalignment change can be described by the following expression:

$$d_r(t) = \frac{d(t)}{r} = \frac{Vt}{r}, \ t_c/2 \le t \le t_c/2, \tag{8}$$

where V is EV speed, t_c – charging time (the time when the coil overlap area is not zero) and $(V\ t_c)\ /\ 2$ equally r.

When the EV misalignment changes, the change of the WPT efficiency at the resonant frequency f_0 can be defined as:

$$\frac{dS_{21}}{dt} = \frac{dS_{21}}{dk} \cdot \frac{dk}{d(d_r)} \cdot \frac{d(d_r)}{dt} = S(k)\frac{2a_2 V}{r} = S_{mis}(x) \cdot V, \tag{9}$$

where

$$S(k) = \frac{2\omega_0 \sqrt{L_1 L_2}\sqrt{R_L L_s}\left(R_L L_s - \omega^2 k^2 L_1 L_2\right)}{\left(R_L L_s - \omega^2 k^2 L_1 L_2\right)^2} \tag{10}$$

is the differentiating of (1) expression by the coupling coefficient k, $(2 \cdot a_2 \cdot V/r)$ is the multiplication of differentiating of (8) and (9) expressions, x – the absolute coil misalignment. $S_{mis}(x)$ characterizes the sensitivity of WPT efficiency for series – series topology at the resonant frequency f_0 with known coils radii r and the coupling coefficient k is a function of the absolute coils misalignment. Figure 4b depicts the variation of sensitivity of WPT as a function of the absolute coil misalignment x on frequency f_0 for $r = h_{opt} = 0.38$ m when x varies from $-r$ to r.

If r and V are known and the allowable change in the energy transfer efficiency ΔS is specified, the time interval required to adjust the transferred power of the WPT system can be defined as:

$$\Delta T_{cycle} \leq \frac{\Delta S}{S_{mis}(x)_{max} \cdot V}, \tag{11}$$

where $S_{mis}(x)_{max}$ – maximum sensitivity of efficiency when the coils misalignment is changed.

For simulated WPT system $S_{mis}(x)_{max}$ max is equals 0.41 m^{-1} at $x = |r|$. The recommended value ΔS is equal 0.01. For these values of the parameters and two values of EV travelling speed: 1 m/s and 10 m/s the upper limit of ΔT_{cycle} is equal to 24.4 ms and 2.44 ms, consequently. These values are upper limits for the duration of compensation cycle for coils misalignment when EV is moving and charging.

2.3 Requirements for Packets Transfer Delay in Compensation Process

Control cycle in compensation process has the following time intervals: for measurement parameters and their sampling in data packets Δt_{msr}; delay in transmission of packets over wireless channel D; adjustment of the corresponding parameter Δt_{adj}; for transition process in the system to establish its parameters in accordance with new values Δt_{trn}.

Duration of one cycle ΔT_{cycle} should be greater than the sum of time intervals, hence, as a requirement for packets transfer delay we have [4]:

$$\text{D} \leq \Delta T_{cycle} - \Delta t_{msr} - \Delta t_{adj} - \Delta t_{trn}. \tag{12}$$

In [1] it was shown that Δt_{msr} is several tens of μs. And Δt_{adj} is of the order of several tens μs. Time of transition Δt_{trn} strongly depends on the parameters of the ICPT system and is characterized by the parameter $\tau = 2(\text{L} + \text{M}) / \text{R}$. This time should be evaluated as (3–6) τ and it is from several hundred till thousand μs.

To estimate the minimal value for ΔT_{cycle} in (12) we need to calculate packets delay D on their path through the network. To the definition of this time is devoted analysis presented in the next section.

Analyzing the obtained results, one can come to the conclusion that using the real-time compensation process from (5) or (11) and (12) we have quite strict requirements for transmission delays of packets through the wireless channel.

3 Packets Delay in Wireless Channel

3.1 Model of Network for Packets Delay Estimations

The process of packets transfer is characterized by parameters: T_{Frame} – the time of frame transfer through the link and R_{Frame} – the number of frames generated in nodes and transmitted through link per unit time. A dimensionless parameter that characterizes "busyness" of the channel is defined: $\rho = R_{Frame} \cdot T_{Frame}$. This quantity is in the range $0 \leq \rho \leq 1$. When ρ = 0 no frames are transmitted, when $\rho = 1$ the channel is busy at every moment of time (100% utilized) and an infinite queue of packets is created at the channel. These packets wait for service of transmission.

According to the Queueing Theory, when requests with intensity λ are coming on the sequence of serving nodes where service is made with intensity μ and when time intervals between requests and time of request's service are arbitrary distributed (so-called G/G/1model) the average service time in one node will be [4]:

$$D = T_{Frame} \cdot \left[1 + \frac{\rho \cdot \left(c_a^2 + c_b^2\right)}{2(1-\rho)} \cdot f\left(c_a, c_b\right) \right], \tag{13}$$

c_a and c_b are the coefficients of variation for random variables of inter-arrival time of requests and service time intervals for requests, respectively. Function f is defined by the following expression:

$$f\left(c_a, c_b\right) = \begin{cases} exp^{-\frac{2(1-\rho)}{3\rho} \cdot \frac{\left(1-c_a^2\right)^2}{c_a^2 + c_b^2}}, \; c_a^2 \leq 1 \\ exp^{-(1-\rho) \cdot \frac{c_a^2 - 1}{c_a^2 + 4c_b^2}}, \; c_a^2 \geq 1 \end{cases}, \tag{14}$$

c – the ratio of the square root of the variance to the mean value of the random variable (for exponential distribution of time intervals $c = 1$). The expression (13) is an approximation [8], but its accuracy is sufficient for our estimations.

We will consider the examples: Ex1 and Ex2. For Ex1 Node1 and Node2 are in the same BSS (Basic Service Set for 802.11) and Node2 transmits to Node1. We have 2 wireless links for this case in the path and for packets delays (13), (14) when $c_a = 1$ give [4]:

$$D = 2T_{Frame} \cdot \left[1 + \frac{2R_{Frame}T_{Frame}\left(1 + c^2\right)}{2\left(1 - 2R_{Frame}T_{Frame}\right)} \right]. \tag{15}$$

For Ex2 Node1 transmits to Node2 and Node2 transmits to Node1. If characteristics of application traffic for Node1 and Node2 are the same, the delay of packets in both directions will be:

$$D = 2T_{Frame} \cdot \left[1 + \frac{4R_{Frame}T_{Frame}\left(1 + c^2\right)}{2\left(1 - 4R_{Frame}T_{Frame}\right)}\right]. \quad (16)$$

For different architectures, there are common features. There is a limit of application bitrate and near this limit, the delays of packets are sharply increased. So, the wireless channel has a bandwidth. The Minimal possible delay is defined by average frame transfer time.

3.2 Frame Transfer Time

Following publication [4] estimations for 802.11 and different specifications (a/b/g/n) are presented in Table 1. The estimations give us minimal possible $T_{Frame\ min}$. These values for frames transfer time are achieved when there is no contention for channel access. In Table 1 the values of $T_{Frame\ min}$ are calculated for maximum PHY transmission bit rate R_{raw}. For every 802.11 specifications there is a set of possible bitrates. The calculations for 802.11n use bitrate 108 Mbps and one data stream (no MIMO – multiple input multiple output) is used.

Table 1. Relationships for T_{Frame} estimations and different specifications. UDP packets, application packet size l [bytes], R_{raw} – bitrate on PHY layer [Mbps].

	802.11b	802.11a 802.11 g	802.11 g Protection RTS/CTS	802.11n
$T_{Frame\ min}$ [μs]	$444 + \frac{8(l+78)}{R_{raw}}$	$94 + \frac{8(l+64)+6}{R_{raw}}$	$520 + \frac{8(l+234)+6}{R_{raw}}$	$100 + \frac{8(l+234)+6}{R_{raw}}$
$T_{Frame\ min}$ [μs]	508	105	556	106
$T_{Frame\ min} + T_{BP}$ [μs]		172	623	172

There are several factors that increase frame transfer time. One factor increasing this value is a concurrent mechanism for access to the radio channel. The frame transmission time should be considered taking into account Back off Period (BP) – an additional certain random number of time slots for waiting before starting the transmission. A simple approximation for the increase due to this factor is:

$$T_{Frame} = T_{Frame\,min} + T_{BP}. \quad (17)$$

For 802.11a/g/n it means [4]: 67.5 μs and 139.5 μs (depends on the realization). The simulation experiments on wireless network model [9] give the results that if the probability of errors in packets (PER) is small, then T_{BP} is negligible, and, if PER is sufficient, the previous estimation for T_{BP} is acceptable.

The next reason for frame transfer time increase is the factors of conditions in the wireless channel. Calculating such increase we consider only AWGN channel characterized by SNR. Simulation results for ARF (Auto Rate Fallback) mechanism (most popular for 802.11 devices) show that the relative increase in frame transmission time is less than 3 times in a wide working range of SNR. In [4] we have presented results of simulation when communicating objects are moving from each other with speed V.

Due to changes of SNR in the channel appropriate decrease in bandwidth and hence increase in T_{Frame} was observed. When changes in SNR were in the range 3–25 dB the observed increase in T_{Frame} is not exceeded 10%.

So, for T_{Frame} upper limit estimations (802.11 specifications) in a working range of SNR we have:

$$T_{Frame\ min} \leq T_{Frame} \leq 3\left(T_{Frame\ min} + T_{BP}\right). \tag{18}$$

This relationship and Table 1 data give the range in what one can find frames (also small packets) delay for 802.11 specifications when acceptance of wireless technology for compensation process is analysed.

4 Conclusions

The main results of this work are following:

- estimates were obtained for the variation of sensitivity of WPT efficiency for series – series topology at the resonant frequency when the distance and misalignment between the transmitter and receiver coils are changing;
- time intervals are defined to perform the processes of compensation of the WPT system so that the energy transfer efficiency does not decrease by more than 1.0%;
- for these time intervals we have an estimation for acceptable packets transfer delay in a wireless data channel; upper limit of ΔT_{cycle} is equal to 2.44 ms for the speed of vehicle 10 m/s;
- the simulation of frame transfer was carried out taken into account access to channel method, PHY layer specifications (802.11a/b/g/n);
- for communication in control applications the lowest possible level for packets delay for 802.11 is about 0.2 ms for a/g/n specifications and when MIMO technology is not used;
- in the channel with Gaussian noise, upper packets delay limit may be about 3 times greater (~ 0.6 ms) in working range of SNR changes.

References

1. Kalwar, K.A., Aamir, M., Mekhilef, S.: Inductively coupled power transfer (ICPT) for electric vehicle charging. Renew. Sustain. Energy Rev. **47**, 462–475 (2015). www.elsevier.com/locate/rser
2. Budhia, M., Covic, G.A., Boys, J.T.: Design and optimization of magnetic structures for lumped inductive power transfer systems. In: Energy Conversion Congress and Exposition, ECCE 2009, pp. 2081–2088. IEEE (2009)
3. Ramos, R.R., Biel, D., Fossas, E., Guinjoan, F.: A fixed-frequency quasi-sliding control algorithm: application to power inverters design by means of FPGA implementation. IEEE Trans. Power Electron. **18**, 344–355 (2003)
4. Krainyukov, A., Krivchenkov, A., Saltanovs, R.: Performance analysis of wireless communications for V2G applications using WPT technology in energy transfer. Procedia Eng. **178**, 172–181 (2017). http://www.sciencedirect.com/science/article/pii/S1877705817300851

5. Dai, X., Zou, Y., Sun, Y.: Uncertainty modeling and robust control for LCL resonant inductive power transfer system. J. Power Electron. **13**(5), 814–828 (2013). https://doi.org/10.6113/JPE.2013.13.5.814
6. Sample, A., David, A., Meyer, D., Smith, J.: Analysis, experimental results, and range adaptation of magnetically coupled resonators for wireless power transfer. IEEE Trans. Ind. Electron. **58**(2), 544–554 (2011)
7. Conway, J.: Inductance calculations for noncoaxial coils using Bessel functions. IEEE Trans. Magn. **43**(3), 1023–1034 (2007)
8. Kramer, W., Langenbach-Beiz, M.: Approximate formulae for the delay in the queuing system GI/G/l. In: International Teletraffic Congress, ITC Bibliographic Database (1976). http://m.itc-conference.org/nc/bibliographic-database/?tx_sevenpack_pi1%5Byear%5D=#c796
9. Krivchenkov, A.: Wireless network Model description. IEEE 802.11 specifications (2016). https://docs.google.com/document/d/1maIdb9ooLXAV9yis6gRCM4YH_jhAeJyEuIQ2xj75J3A/edit?usp=sharing

Weigh-in-Motion by Fibre-Optic Sensors: Problem of Measurement Errors Compensation for Longitudinal Oscillations of a Truck

Alexander Grakovski(✉) and Alexey Pilipovecs

Transport and Telecommunication Institute, Lomonosova Street 1, 1019 Riga, Latvia
avg@tsi.lv

Abstract. Recorded signals from a group of fibre-optic sensors of a passing truck with various speeds and known weight of preliminary weighed reference vehicle are used as an input data. Moving vehicle's dynamics model of "inverted lever pendulum" and the impacts of external conditions (speed of vehicle, temperature, tyre width etc.) as well as the longitudinal and transverse oscillations as the main source of measurement errors are in the focus of this research. The shapes of tyre footprint form, pressure and weight distribution along the footprint length are being estimated and discussed in order to extract and compensate the longitudinal and transverse oscillations of tractor's, semitrailer's, each axle's and wheel's "gravity centre" with the aim to decrease the estimation errors to the level till 1–2% of each axle's real weight.

Keywords: Weigh-in-motion · Fibre-optic sensors · Measurement errors Longitudinal oscillations

1 Introduction

Methods and systems for weighing in motion (WIM) are practically used for automatic control of the weight of vehicles passing by the road from the end of the 20th century [1]. Modern types of sensors (piezoelectric, quartz), as well as fibre-optic sensors (FOS) allow to measure the total weight of the machine with relatively high accuracy of about 1–2% as the sum of the weights of all axes [2, 3].

However, when measuring the weight of each axis separately (as it is required to check the compliance of the load values to each axis with the law), measurement errors can reach 10–20%, which is not acceptable and introduces limitations of WIM's capabilities to pre-selection aims only (allocating potentially overloaded transport for subsequent verification by official stationary weighing).

The functionality of new generation fibre-optic sensors (FOS) is based on changes in the parameters of an optical signal due to deformation of the optical fibre under the weight of a passing vehicle [4]. Recorded signals from a group of FOS of a passing truck with various speeds and known weight of preliminary weighed reference vehicle are

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_36

used as an input data. The results of a truck tyre surface contact patch reconstruction allows to the axle weight-in-motion estimation with the accuracy of less than 10% [5].

Moving vehicle's dynamics model of "inverted lever pendulum" and the impacts of external conditions (speed of vehicle, temperature, tyre width etc.) as well as the longitudinal and transverse oscillations as the main source of measurement errors are in the focus of this research. The shapes of tyre footprint form, pressure and weight distribution along the footprint length are being estimated and discussed in order to extract and compensate the longitudinal and transverse oscillations of tractor's, trailer's, each axle's and wheel's "gravity centre" with the aim to decrease the estimation errors to the level till 1–2% of each axle's real weight.

2 Experimental Data and Vehicle's Weight Estimation by FOS

To gather experimental data for research 2-axle tractor Volvo FH12 and 3-axle semi-trailer was used. It was loaded with weight slightly above the legal limit (11 500 kg) for 2nd axle (Table 1). 18 sets of FOS sensor track recordings were performed. The test data is from April 2012 and air temperature was 12 °C.

Table 1. Reference vehicle axle's weights (static platform weights, accuracy < 1%).

Reference wheel weight (tons)	Date: 20/04/2012 (Air temperature +12 °C)				
	1st axle	2nd axle	3rd axle	4th axle	5th axle
LEFT wheel	3.666	6.203	2.541	2.833	2.924
RIGHR wheel	3.630	6.416	2.968	2.808	2.920
FULL axle	7.296	12.619	5.509	5.641	5.844

Totally 180 signals were received (for each 5-axis wheel) with signal sampling rate at 20 μs. Various driving modes (uniform motion, acceleration, braking) at speeds of 10 to 90 km/h were acquired. Experimental truck was previously weighed on a stationary scale with accuracy less than 1% and a standard deviation of the weighting factors of 100 kg per wheel (Table 1). This data was used as a known mass per wheel and per axle.

Total weight of each wheel was calculated using basic method. Air pressure inside the tyre that was calculated based on recorded and linearized signal (according to FOS properties [5]) was multiplied by the dynamic area of tyre footprint. In order to obtain dynamic area of the tyre footprint normalized and filtered FOS signal needs to be split into two components: even (symmetric) component related to sensor's pressure of weight, and odd (asymmetric) component corresponds to friction force pressure [7].

Footprint width of each wheel, speed measurements and vehicle's wheelbase are required to calculate tyre footprint area and weight. Measurement station configuration and calculations are not in scope of this research [5, 6]. After calculations estimated wheel weights can be compared to reference weights (Table 1). Relative measurement errors are calculated using $\varepsilon_w = W_{ha}/W_{ref} \cdot 100\%$. Error examples at different driving modes and speeds are listed in Table 2.

Table 2. An example of errors of wheel's and axle's weights (measured by WIM station).

Error of weight measurement	Calculated errors (%)				
	1st axle	2nd axle	3rd axle	4th axle	5th axle
Signal s2_90 km (speed 85 km/h)					
Left wheel	–10.449	–13.855	2.431	2.562	2.661
Right wheel	13.789	0.297	1.261	9.127	7.130
Full axle	1.610	–6.659	1.801	5.831	4.894
Tractor vs. semitrailer	–3.630		4.202		
Full lorry	–0.0235				

As it is clearly seen, largest error in axle load values varies in the range of 6–10%, despite the error in total weight measurement of vehicle is almost zero. Such errors for each wheel can reach 14% or more (in the case of sudden braking or acceleration, especially at low speeds). In such scenario WIM systems accuracy limits its application to overloaded transport pre-selection task at moving speed greater than 50 km/h, that corresponds to WIM B+(7) class (COST 323) for the high speeds and D2 (OIML R134 [9]).

Calculated measurement errors of each wheel, axle, tractor and semitrailer weights are presented (Fig. 1) in order to illustrate the effect of longitudinal and transverse oscillations in measurements and estimation of weight (Table 2).

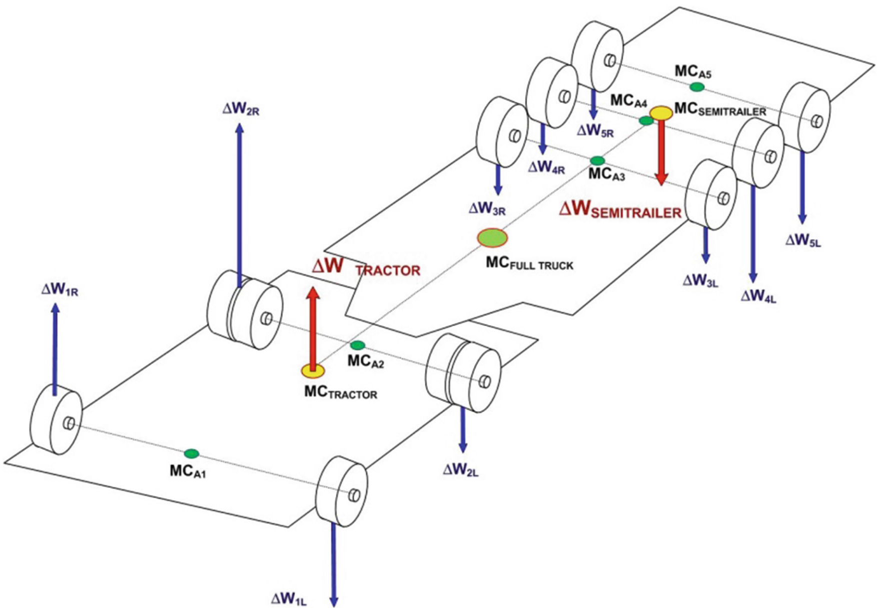


Fig. 1. Visualization of vehicle's wheels weight errors in comparison with estimated and reference values, due to longitudinal and transverse oscillations impact in weighing-in-motion [10].

Tractor's and semitrailer's weight errors compensate each other, forming a minimal error of a whole truck. Same result cannot be observed for axles or each wheel because it was caused by effect of longitudinal and transverse oscillation.

3 Model of "Inverse Pendulum" for Longitudinal Oscillations Errors Compensation

To estimate impact of longitudinal oscillations, a vehicle can be represented as a model of a composite "inverse pendulum" [10]. Its supports (equilibrium points) are geometrically defined in "gravity centres" (Fig. 2), which can be estimated by the weights of the axles and geometrical dimensions of the vehicle.

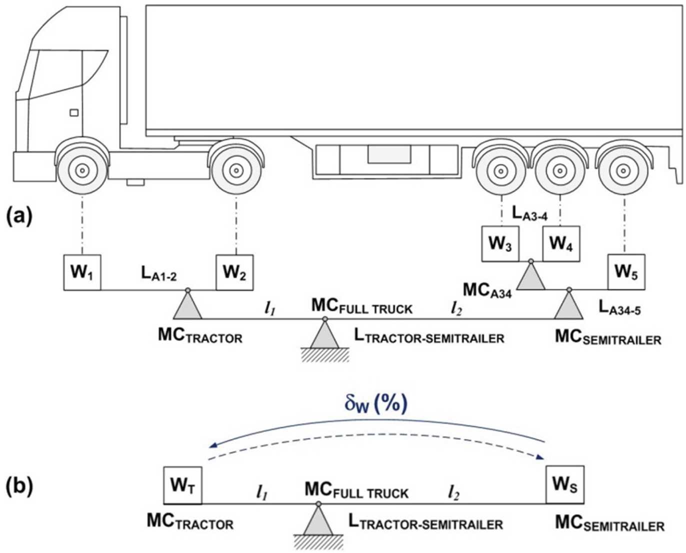


Fig. 2. (a) Complicated "inverse lever" (pendulum) model of vehicles dynamics for longitudinal oscillations description and compensation [10], and (b) inverse lever model for the system "tractor-semitrailer" and transfer of part of full weight δ_W from one side to another.

Having time intervals between the peaks of the FOS signals from each wheel and the speed [10] it is possible to calculate: distance between axles (wheelbase), speed of each axle and axle's tracks. For example, tractor-semitrailer system has position of the support point $MC_{FULL\ TRUCK}$ (Figs. 1 and 2a). It will be determined by traditional mechanical "rule of the lever":

$$W_T \cdot l_1 = W_S \cdot l_2, \tag{1}$$

where W_T and W_S are measured weight of tractor or semitrailer respectively, and l_1 or l_2 are the arms of lever (Fig. 2b).

If value of the tractor W_T will be decreased by half of motor second axle's weight of tractor, and replace the modified weight of the tractor W_T with the semitrailer W_S weight in expression (1), i.e. change the lever to the reverse of "inversive lever", then obviously equilibrium of such system will be violated.

An amazing phenomenon was observed for all analyzed records of the same truck and its load on 20-04-2012, it was the fact that the relative imbalance of equilibrium (disequilibrium) that can be calculated using:

$$\varepsilon = \frac{W_S \cdot l_1 - W_T \cdot l_2}{W_S \cdot l_1 + W_T \cdot l_2} \cdot 100\%, \tag{2}$$

will be approximately equal to the double peak value $\left(\varepsilon_S - \varepsilon_T\right)$ or magnitude of weight measurement errors of a tractor ε_T and semitrailer ε_S from Table 2. The comparison results are shown in Table 3 for all signals of this date (the last two columns).

Table 3. Examples of vehicle's "tractor-semitrailer" weight error comparison with the "inverted pendulum" disequilibrium for records of the same truck and its load (20/04/2012).

Signal no. (speed, mode)	Tractor's weight error ε_T (%)	Semitrailer's weight error ε_S (%)	Magnitude of weight errors $\left(\varepsilon_S - \varepsilon_T\right)$ (%)	Disequilibrium ε of "inversive pendulum" (%)
s3_90 km (85 km/h)	–3.63%	4.20%	7.83%	7.71%
s2_70 km (74 km/h)	–4.36%	5.32%	9.68%	10.33%
s1_50 km (53 km/h)	–4.13%	4.86%	8.99%	9.81%
s3_20 km, (17 km/h, acceleration)	–2.24%	2.61%	4.85%	9.50%
s2_10 km, (12 km/h, braking)	–4.47%	5.04%	9.51%	10.79%

The same effect can be observed, if we transfer approximately $\delta_W \approx 7.92\%$ of full weight of the truck (Fig. 2(b)) from W_T to W_S value. If the magnitude $(\varepsilon S - \varepsilon T)$ of weight measurement errors is known, and in assumption that absolute errors in weight measurements of the tractor ΔT and semitrailer ΔS must be with different signs and equal in magnitude, we can easy reduce the to the usual quadratic equation with respect to unknown variables $\Delta_S = -\Delta_T$ [10].

This error can be subtracted from the results of weight estimation. Distributing this approach to each wheel of tractor or semitrailer respectively, will causes a change in the positions of the geometric "gravity centre" of all system elements (Fig. 2a).

Similar operations for error estimation of longitudinal oscillations can be performed successively for the constituent elements of the model: tractor (1st and 2nd axle), then for semitrailer group (3rd, 4th and 5th axle). After eliminating the errors of vehicle longitudinal oscillations, the resulting errors in weight measurements were significantly reduced to <1–2% for each axle [10].

Unfortunately the problem on supposed approach consists of the fact, that the value of characteristic indicator number $\delta_W \approx 7.92\%$ isn't appropriate for another experiment series (days) with different types of semitrailer and load volume. Experimentally defined characteristic indicator numbers δ_W for another experiment series are presented in Table 4.

Table 4. Examples of value of characteristic indicator number δ_W for separate experiment series (days) with different types of semitrailer and load volume.

Experiment (date)	Full weight (tons)	Lever length (m)	δ_W (%)
14/03/2013	34.6666	8.7969	6.692
17/11/2011	18.8750	9.5320	7.285
20/04/2012	36.9087	8.7895	7.915
08/02/2012	39.8000	8.8123	–9.447
13/03/2013	14.0550	9.3787	16.542
15/05/2012	14.3250	9.4063	22.373
17/05/2012	14.4900	9.4348	24.172

Provided analysis of all experimental data allows to dependence of characteristic indicator number δ_W on the relation of reference (correct) weight of tractor $\overline{W}_T$ and semitrailer $\overline{W}_S$, i.e. $\delta_W = function\left(\overline{W}_T \big/ \overline{W}_S\right)$. The shape of its dependence is presented on Fig. 3.

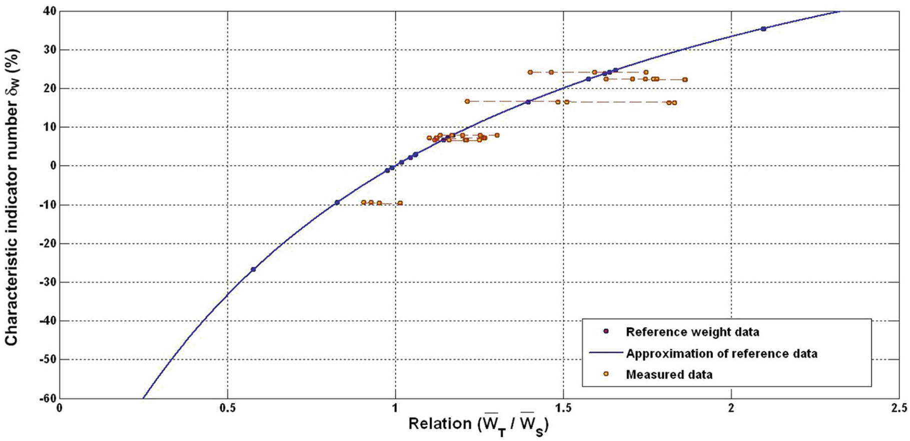


Fig. 3. Dependence of characteristic indicator number δ_W on the relation of reference (correct) weight of tractor and semitrailer $\overline{W}_T \big/ \overline{W}_S$.

Unfortunately, but these results allow calculating the errors of longitudinal oscillations and the correct weights $\overline{W}_T$ and $\overline{W}_S$ only if the right reference $\overline{W}_T \big/ \overline{W}_S$ is known to us, i.e. the problem has already been solved. If we use measurement data for calculations W_T and W_S that contain systems oscillations errors, then determine the error of the characteristic indicator number δ_W may have 20% error, which is unacceptable

(Fig. 3, red points). Therefore, we will consider a different approach based on the calculation of the resonant frequencies of the system.

4 Natural Resonant Frequencies of Truck Model

Like any mechanical system, a heavy truck has its own resonant oscillation frequency, for a serviceable air suspension usually within 1.5–4 Hz, corresponding to bounce, pitch and roll mode resonant frequencies of the sprung masses [11]. The intrinsic resonant frequency depends on the total mass and the geometric characteristics of its distribution along the machine.

To establish the relation between the characteristic indicator number and natural resonant frequencies, approach that considers the analogy between the oscillation system of the "inverse pendulum" model of truck (Fig. 4a) and the system of sprung mass oscillation (Fig. 4b) will be used.

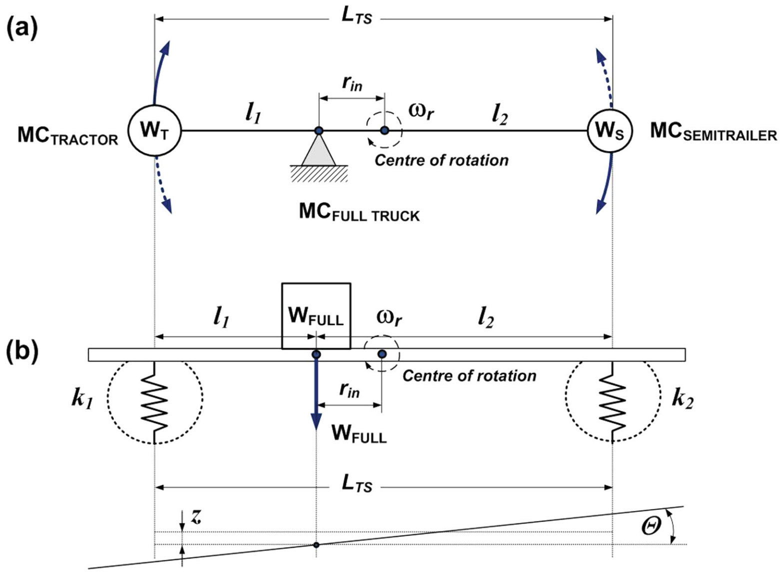


Fig. 4. (a) Oscillation system of "inverse pendulum" model of truck and (b) the system of sprung mass oscillation [8].

The problem of determine the free vibrations of a model with a sprung mass [8] reduce to a system of two linear ordinary differential equations with the respect to variables: the vertical displacement of the centre of gravity $z(t)$ and the change in the angle of rotation $\Theta(t)$ (Fig. 4b):

$$\begin{cases} \ddot{z} + a \cdot z + b \cdot \Theta = 0, \\ \ddot{\Theta} + \frac{b}{r_{in}^2} \cdot z + \frac{c}{r_{in}^2} \cdot \Theta = 0, \end{cases} \quad (3)$$

where

$$a = \frac{(k_1 + k_2) \cdot g}{W_{FULL}}, \quad b = \frac{(-k_1 \cdot l_1 + k_2 \cdot l_2) \cdot g}{W_{FULL}}, \quad c = \frac{(k_1 \cdot l_1^2 + k_2 \cdot l_2^2) \cdot g}{W_{FULL}} \quad (4)$$

are coefficients of the model, k_1 or k_2 are coefficients of stiffness of left and right spring on Fig. 4b, $r_{in} = 0.5 \cdot L_{TS} - \min(l_1, l_2) = 0.5 \cdot |l_1 - l_2|$ is the radius of inertia of the system, and $g = 9.81\ \text{m/s}^2$ is the acceleration of gravity. Under the condition of harmonic oscillations, this problem has two real solutions for resonant frequencies: $f_{0(low)}$ and $f_{0(high)}$:

$$f_{0(low,high)} = \sqrt{\frac{1}{2} \cdot \left(\frac{c}{r_{in}^2} + a \right) \pm \sqrt{\frac{1}{4} \cdot \left(\frac{c}{r_{in}^2} - a \right)^2 + \left(\frac{b}{r_{in}^2} \right)^2}}. \quad (5)$$

If restrictions will be entered, for example, the upper frequency limit $f_{0(low)} = 5$ Hz (<4 Hz according to [11]), and also condition of independence of variables $z(t)$ and $\Theta(t)$ from [8] as $k_1 \cdot l_1 = k_2 \cdot l_2$, i.e. $b = 0$, then such system can be solved by first determining coefficients of spring stiffness k_1 and k_2, and then the lower resonant frequency $f_{0(low)} = \sqrt{a}$ for reference data (Table 5).

Table 5. Natural resonant frequency (low), calculated from reference data by (5) and (6).

Date	Full weight (t)	Length (m)	Ratio $\overline{W_T}/\overline{W_S}$	[a]$\lvert\delta_W\rvert$ (%)	$f_{0(low)}$ (Hz)
14/03/2013	34.667	8.79	1.1434	6.69%	0.3354
17/11/2011	18.875	9.53	1.1571	7.28%	0.3652
20/04/2012	36.909	8.79	1.1719	7.92%	0.3970
08/02/2012	39.800	8.81	0.8274	9.44%	0.4745
13/03/2013	14.055	9.38	1.3964	16.54%	0.8387
15/05/2012	14.325	9.41	1.5764	22.37%	1.1478
17/05/2012	14.490	9.43	1.6242	23.79%	1.2246

[a]We consider absolute volume of characteristic indicator number |δW| because of independence of resonant frequency on the direction of weight changes in W_T and W_S.

As in the dependence of the characteristic indicator number on the ratio of the tractor-semitrailer weights (Fig. 3), we see that for real measurements in motion, a relatively large spread of estimates is obtained (Fig. 5). This is explained by the fact that the calculation of the self-resonant frequencies (5) essentially depends on the correct measurement of the inertia radius rin, and it depends on the position of the centres of mass and varies greatly in the dynamics of motion from signal to signal (Fig. 6).

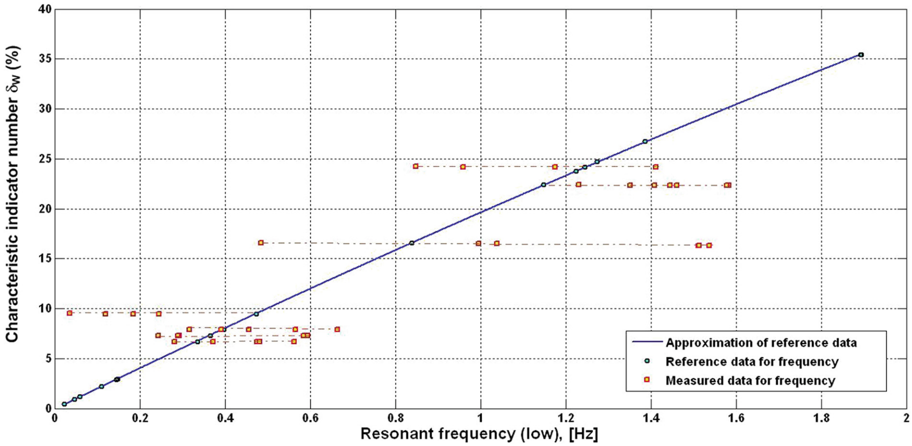


Fig. 5. Dependence of characteristic indicator number δ_W on natural resonant frequency (low) of the system of sprung mass oscillation for reference and measured data.

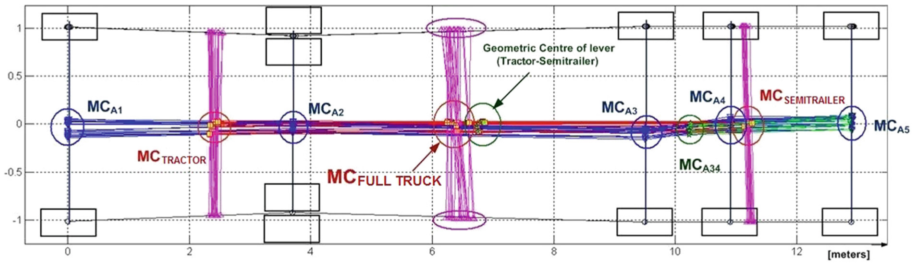


Fig. 6. Tracking of full truck's, axle's 1–5, tractor's and semitrailer's mass (or gravity) centres position shifts during 18 different speed experiments on 20/04/2012 for the same load of truck.

Stable algorithm for calculating the radius of inertia r_{in} for single measurements can be considered to be a further research for success in solving the problem of eliminating the error of longitudinal oscillations on the system.

5 Conclusions

WIM measuring stations that are based on fibre-optic sensors (FOS) are reasonable compared to conventional piezoelectric and other sensors for economy and reliability purposes, unfortunately but cannot have measuring accuracy better than 10%, which limits their use only for automatic pre-selection tasks of overloaded vehicles.

The application of the proposed method for compensation longitudinal and transverse oscillation errors based on the phenomenon of inverse pendulum dependence can significantly reduce errors in measuring the load per axle to 1–2% at a speed of more than 50 km/h, even if there are driver irregularities. Therefore automatic measuring station WIM based on FOS is fully capable of competing in accuracy, reliability and cost with a stationary platform weight scale.

Its usage in WIM system requires additional research for determination of characteristic indicator number using resonant frequency of inverse pendulum model and stable algorithm for determining of the system's inertia radius.

Acknowledgments. This research was granted by state of Latvia project "The next generation of information and communication technologies (NexIT)" (2014–2017).

References

1. Malla, R.B., Sen, A., Garrick, N.W.: A special fiber optic sensor for measuring wheel loads of vehicles on highways. Sensors **8**, 2551–2568 (2008)
2. Teral, S.R.: Fiber optic weigh-in-motion: looking back and ahead. Opt. Eng. **3326**, 129–137 (1998)
3. Kistler Group AG Lineas® Quartz Sensor for Weigh-In-Motion (WIM) (2015). https://www.kistler.com. Accessed 21 Mar 2017
4. SENSORLINE GmbH (© Sensor Line). PUR Series Fiber Optic Traffic Sensor (2014). http://sensorline.de/pages/downloads.ph. Accessed 20 Mar 2017
5. Grakovski, A., Pilipovecs, A., Kabashkin, I., Petersons, E.: Tyre footprint geometric form reconstruction by fibre-optic sensor's data in the vehicle weight-in-motion estimation problem. In: Informatics in Control, Automation and Robotics, 10th International Conference, ICINCO 2013 Reykjavík, Iceland, 29–31 July 2013. Lecture Notes in Electrical Engineering, vol. 325, pp. 123–137 (2015)
6. Grakovski, A., Pilipovecs, A.: The problem of tyre footprint width estimation by fibre optic WIM sensors in condition of geometric complexity. In: Dependability Engineering and Complex Systems, 11th International Conference, DepCoS-RELCOMEX, 27 June–1 July 2016, Brunow, Poland. Revised Selected Papers, Series: Advances in Intelligent Systems and Computing, vol. 470, pp. 219–227 (2016)
7. Krasnitsky, Y.: Transient response of a small-buried seismic sensor. Comput. Model. New Technol. **16**(4), 33–39 (2012). TTI, Riga
8. Timoshenko, S.: Vibration Problems in Engineering, 3rd edn. D. Van Nostrand, New York (1955). (with the collaboration of Young, D.H.)
9. Jacob, B., O'Brien, E.J., Newton, W.: Assessment of the accuracy and classification of weigh-in-motion systems: part 2 European specification. Int. J. Veh. Des. – Heavy Veh. Syst. **7**(2/3), 153–168 (1989)
10. Grakovski, A., Pilipovecs, A.: Multi-purpose fibre optic system for automated vehicle's weighting-in-motion and classification in applications of intelligent transport systems. In: 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 611–616. IEEE Press, Naples, Italy (2017). https://doi.org/10.1109/mtits.2017.8005586
11. Cebon, D.: Vehicle generated road damage: a review. Veh. Syst. Dyn. **18**, 107–150 (1989)

Experimental Research on Introduction of Distributed Road Tracking System for Road Traffic Registration in Latvia

Alexander Dudko[1], Irina Yatskiv[2(✉)], and Yasushi Kiyoki[1]

[1] Keio University, Endo 5322, Tokyo, Japan
aleksandrsdudko@gmail.com, kiyoki@sfc.keio.ac.jp
[2] Transport and Telecommunication Institute, Lomonosova street 1, Riga, Latvia
jackiva.i@tsi.lv

Abstract. Worldwide the number of people killed in road traffic accidents is estimated at 1.2 million a year. Although, road traffic injuries have been a leading cause of mortality for many years, most traffic crashes are both predictable and preventable. There is considerable evidence on interventions that are effective at making roads safer. Countries that have successfully implemented these interventions have seen corresponding reductions in road traffic deaths. One of the important factors for safety is cars movement speeds. In this paper a distributed road tracking system is proposed. This system has multiple devices installed on city roads equipped with cameras and automatic car number recognition module. Two experiments demonstrated that spot speed cameras have very limited local impact on driving speed – drivers are slowing down exactly near the speed camera and accelerate straight after passing it, which means that drivers afraid of fines and slow down when they are checked on speed. Based on the experiments' results it was suggested that proposed system introduction on the roads of Latvia will have a significant effect of reducing the average speed of cars and decrease road fatality rate on the roads of Latvia by at least 25%.

Keywords: Traffic safety · Average speed · Road tracking system Experiments

1 Introduction

According to the global status report on road safety 2015 [1], in the last three years, 17 countries have aligned at least one of their laws with best practice on speed, seat-belts, drink-driving, motorcycle helmets and child restraints. While there has been progress towards improving road safety legislation and in making vehicles safer, the pace of change, as shown in [1], is too slow. Urgent action is needed to achieve the ambitious target for road safety and one of the most important factors for road safety is cars movement speed. An increase of average traffic speed not only influences to the likelihood of a crash, but also raises the risk of death and serious injury, especially for pedestrians,

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_37

cyclists and motorcyclists. Speeding has been a factor in about one-third of crash deaths since 2006 according to [7].

Altintasi et al. in [4] are using floating car data (FCD) for detecting pattern. The authors used 1-min interval FCD experiment for an urban arterial in Ankara data which has been collected during the morning peak hour for 2 months. Average speed values were transformed into a qualitative 4-scale state parameter based on the Level of Service definitions for urban roads. Pattern searches over consecutive segment states using different search length showed that FCD is capable to detect recurrent congestion or bottleneck locations, and even have an idea about the length of queue formed before the bottlenecks. Their results indicate that even though only one parameter of average travel speed from FCD is used, it is still possible to detect critical patterns along urban roads, if data is continuous and extensive [4]. Critical patterns which are able to be detected include persistent congestion due to interchanges, speed enforcement points, bus stops etc.

Demirel and Shoman in [5] presented an approach for integrating various spatial-temporal data to aid dynamic urban road network management. In the proposed methodology, the speed information retrieved from detectors was associated with the OpenStreetMap and quickest paths were simulated within determined time intervals. Different shortest routes are proposed to users in various time spans via incorporated speed information and network analyses. For the same origin and destination, 2.5 min difference is observed within the study area.

In the current research, a distributed road tracking system on the basis of multiple sensors installed on city roads equipped with cameras and automatic car number recognition module is proposed. The system should be a part of a Smart City Roads System and used for transport registration, road laws enforcement system, and smart traffic flows analysis to advance urban transportation. The goal of system is to advance the existing city transport by means of raising urban innovations to maximize road laws enforcement and minimizing the negative effects like traffic accidents. The research involves two experiments on data collection of environment and traffic flow data on the roads of Riga city and in Latvia to estimate real-time road traffic capacity, average traffic speed, individual drivers' speeds, and route source-destination information statistics. The goal of experiments is to estimate the benefits of the proposed system introduction on the roads of Latvia and analyze the possible effect of reducing the average speed of cars and decrease road fatality rate on the roads of Latvia.

The paper describes in Sect. 2 the architecture of distributed road tracking system. Road experiments and their results are presented in Sect. 3. In Sect. 4 we discuss the results and propose a system introduction plan on main roads of rural areas and in Riga city, as well as give the estimated contribution of the system to the roads safety in Latvia.

2 Proposed Distributed Road Tracking System Architecture

The proposed Distributed Road Tracking System consists of two main parts: (1) road traffic sensor with automatic car number plate recognition module, and (2) centralized processing and storage server, which connects all road sensors together and collects all data for analysis.

The possible variant of road sensor architecture is presented on Fig. 1. Road traffic sensor with automatic car number plate recognition module is a small and inexpensive device consisting of a video camera, basic video processing module, and small storage for images with detected vehicles. All stored images are assigned with a predefined time-to-live (TTL) which is depended on distances to other nearest road traffic sensor points. In general, it could be less than 1 h for urban areas and around 4 h for countryside roads with less amount of road traffic sensors. During a normal operation old images are cyclically overwritten with the new ones to reuse the limited storage space. In case of detected speeding violation, which is determined at the central server, specific image can be marked for a longer storage by extending its time-to-live setting and uploading to a server.

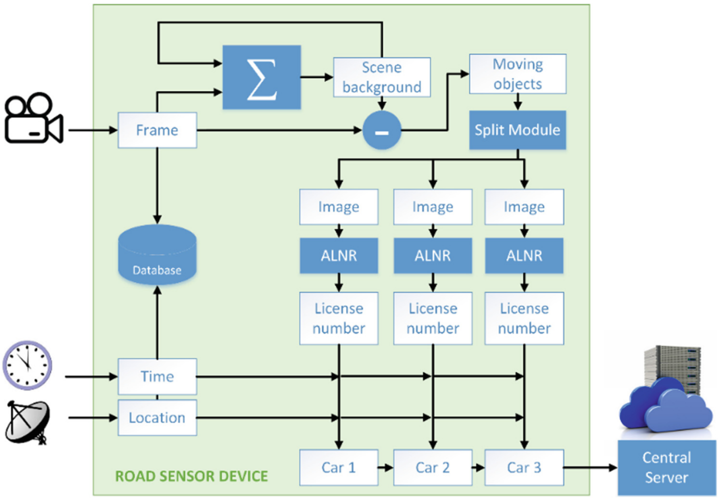


Fig. 1. Road sensor architecture.

Car number plate detection is performed in real-time and due to limited performance capabilities of the device, it must be very light although precise. One of the important steps of video processing is scene background estimation. Since the video is processed in real-time, the video may have varying frame rate. Therefore stable scene background is calculated from fame images by taking into account time between frames.

$$\begin{cases} fr_{bg} = p * fr_{current} + (1-p) * fr_{bg} \\ p = \exp\left(-k\left(t_{current} - t_{previous}\right)\right) \end{cases},$$

where fr_{bg} – background frame; $fr_{current}$ –current frame; $t_{current}$ – time of the current frame $t_{previous}$ – time of the previous frame.

By calculating an absolute difference between current frame and background frame we get image areas with moving objects. Using pattern matching approach, car number place is located. Optical character recognition (OCR) approaches are used to recognize the car license number. For the improved performance two cameras are used: low-resolution camera for moving objects detection, and high-resolution camera for car plate number recognition.

All tracking history is stored in the central server (see Fig. 2), which includes a data analysis module. The distributed system should be a scalable tree architecture, which is capable of interconnecting districts to one city, cities to country regions, and to the whole country level. Data retrieval and automatic processing features of the central server should include following:

- Passed vehicles count statistics per hours, days, weeks, months, seasons, and years, as well as by location: road, district, city, and country;
- Speed monitoring statistics for individual vehicles and for overall traffic capacity on different levels, such as road, district, etc.;
- Speed regulation for road laws enforcement including automatic generation and sending fine sheets to recognized vehicles owners.

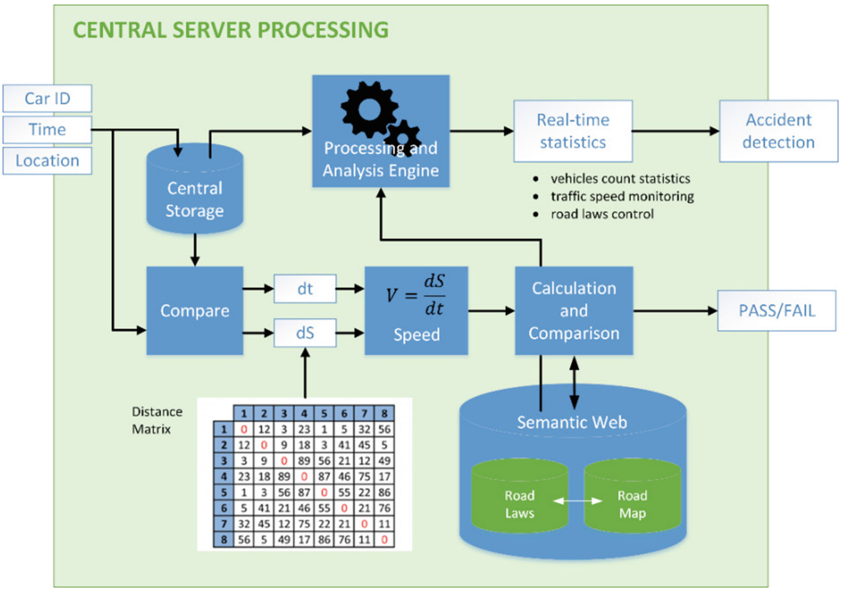


Fig. 2. Central server architecture.

Speed calculation in the system is performed through the following main items: vehicle plate number detection at monitoring points, central storage of data (time, point, plate number), and distance table between monitoring points. When the system detects a new vehicle it perform following steps: (1) Recognize vehicle license number and sent to central server to process; (2) Find previous point and time the car was last seen; (3) Get distance between current and previous point; (4) Calculate speed from known distance and time difference.

To minimize data transfer size between tracking devices and central server plate number is determined on place by recognition from video stream. To keep the system accurate on speed calculation it is crucially important to have the same clock time on every device in the network. Therefore, an important procedure in system operation is clock synchronization.

3 Experiments

Current research is important for Latvia because road safety situation in the country is low. Latvia has the highest road fatality rate among countries of EU according to report for year 2014 from Mobility and Transport department of European Commission. The average road fatality rate in European Union is equal to 51 dead per million inhabitants whereas road fatality rate in Latvia is 105 dead per million people [1].

3.1 Current Situation

Currently 48 stationary speed cameras are installed in Latvia plus there are 105 additional dedicated places for speed cameras in total, according to road transportation safety department CSDD [2], and country police data [3]. The Fig. 3 shows the locations of these speed cameras in Riga city and nearby. Official statistics about speed violations detected by speed cameras is described in [6].

Fig. 3. Locations of speed cameras in Riga city and nearby.

According to [6] the radar price varies between 42 000 and 87 500 Euro each, which is very high price. For a small country such is Latvia, it is very expensive to install new radars. Therefore, much less expensive devices are needed to cover large areas, which is crucially important for influence vehicles speeds on the roads.

In the paper experiments for demonstrating the current situation on the roads and the local effect of spot speed cameras were presented. During the experiments data was collected and processed manually. An individual person was recording a video stream at every measurement location during the whole experiment. Then all video files were processed together to extract information about passed cars. It was required to synchronize time on all recording devices to ensure precise calculation of travel times between points. Travel time for every car was calculated, based on known creation time of video file and the time offset within the video stream when the car was detected.

3.2 Experiment 1. Karla Ulmana Street

In this experiment we covered a road from Riga city towards Jurmala city, which is the main connection between two cities, which has three lanes on the road for both directions. In the middle of the road there is a static speed camera installed on a regular basis. The speed limit on this road is 90 km/h.

The experiment on this road includes 4 measurement points as shown on the Fig. 4 with the red marks. The location of static speed camera is marked with a violet circle icon. Every measurement point was capturing a video of passing cars and license numbers were registered with timestamp information. The experiment was performed in July on Saturday from 3 pm till 3:30 pm. Measurement was performed for one direction: from Riga to Jurmala. During 30 min of the experiment 920 cars were registered in the study area segment A, and 795 cars at the segment C. It can be explained by the road junction in the middle, which was letting some cars to enter and exit the study area.

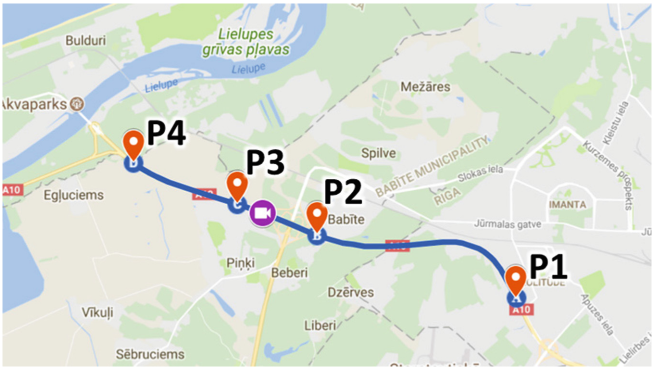


Fig. 4. Experiment 1 study area and measurement points.

Table 1 shows the aggregated information from this experiment on average travel time between check points, calculated average speed for every measured segment, and the amount of speed limit violation according to the regulation of the road.

From Table 1 it can be clearly seen that drivers do not violate the speed limit in the segment with a speed camera. However, there is a quite big difference on other segments in the study area. This fact shows that spot speed cameras have very limited local impact on driving speed – drivers are slowing down exactly near the speed camera and accelerate straight after passing it. Although, it is very important to mention that drivers afraid of fines and slow down when they are checked on speed. This means that the longer is the speed check distance, the longer drivers would follow the speed limit. This allows

Table 1. Collected data summary for experiment 1.

	Road segment		
	A	B	C
From–to	Point 1–Point 2	Point 2–Point 3	Point 3–Point 4
Cars passed	920	750	795
Avg travel time, s	149.91	65.56	70.75
Distance, m	4 360	1 630	2 150
Average speed, km/h	104.7	89.5	109.4
Speed limit, km/h	90	90	90
Violation, km/h	14.7	–0.5	19.4
Has speed camera	No	Yes	No

us to make a strong assumption that proposed system will have the desired effect of reducing the average speed of cars on the roads and contribute to road safety.

3.3 Experiment 2. Road A9: Riga–Liepaja

In this experiment we covered a part of the road between Riga city and Liepaja city. This road has only one lane for every direction with the speed limit 90 km/h.

The experiment on this road includes 2 measurement points as shown on the Fig. 5 with the red marks. Same as for previous experiment, a video of passing cars was captured at every measurement point, and license numbers were registered with time-stamp information. The experiment was performed in August on Saturday from 11 am till 11:30 am. Both directions were measured. The road in the study area does not have crossings and significant junctions so the entire transport flow is going straight from one end to another.

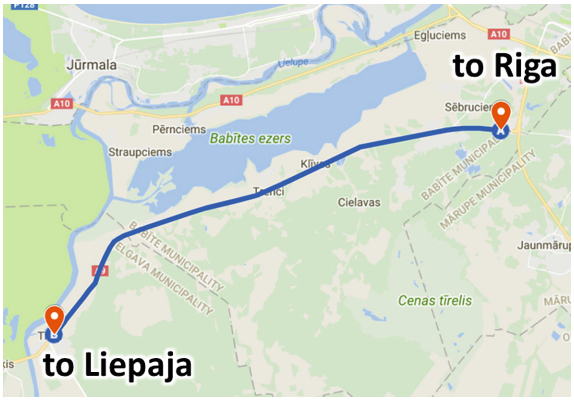


Fig. 5. Experiment 2 study area and measurement points.

Table 2 shows the aggregated information from this experiment on average travel time between check points, calculated average speed for every measured segment, and the amount of speed limit violation according to the regulation on this road.

Table 2. Collected data summary for experiment 2.

	Direction	
	Riga–Liepaja	Liepaja–Riga
Cars passed	137	145
Avg travel time, s	750.11	727.27
Distance, m	22 000	22 000
Average speed, km/h	105.6	108.9
Speed limit, km/h	90	90
Violation, km/h	15.6	18.9
Has speed camera	No	No

This experiment shows that on urban areas with no speed cameras drivers tend to violate the speed limits very much. As shown on speeds histogram on Fig. 6, very few drivers were following the road rule of 90 km/h limit. Since the road has only one lane for each direction, some drivers were overtaking other cars, which also contributed the higher average speed. Introduction of the proposed system on this study area could significantly reduce the driving speed.

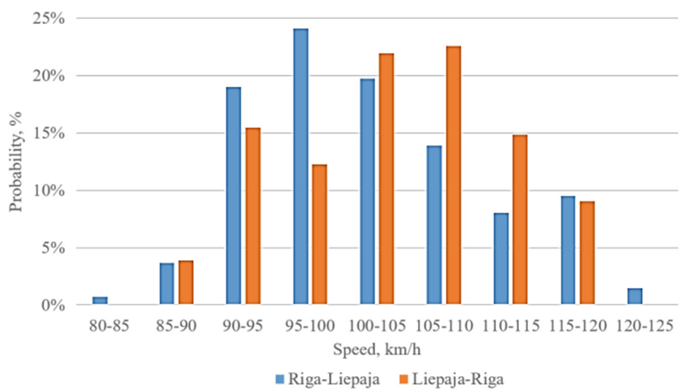


Fig. 6. Speed distribution histogram for both directions of study area in experiment 2.

4 Results Discussion

Urban and rural areas have fundamentally different characteristics with regard to density of road networks, land use, and travel patterns. Consequently, the characteristics of fatal motor vehicle crashes differ between rural and urban areas. For example, pedestrian and bicyclist deaths and deaths at intersections are more prevalent in urban areas, whereas a larger proportion of passenger vehicle and large truck occupant deaths and deaths on high-speed roads occur in rural areas. Speeding has been a factor in about 30% of crash deaths in both rural and urban areas since 2006 [7].

According [10] an increase of 1 km/h in average vehicle speed results in an increase of 3% in the incidence of crashes resulting in injury and an increase of 4–5% in the incidence of fatal crashes.

Rural areas. One part of the proposed system introduction supposes the coverage of highways in rural areas between cities. There are 9 cities in Latvia in total which includes Riga, the capital, and 8 other cities: Daugavpils, Liepāja, Jelgava, Jūrmala, Ventspils, Rēzekne, Valmiera, Jēkabpils. Their total population is equal to 1.112 mln inhabitants. Compared to total population of Latvia 1,953,200 all these main cities cover 52% of country population [8].

To cover all main roads between the 9 cities of Latvia it is required to install road sensor devices on 9 roads which makes up only 18 locations. Figure 7 shows these locations on the map. Every road between two cities has two corresponding locations of road sensors – at the beginning of the road and at the end of the road. Setting one device per direction would require 2 devices for every location, which in total would give 36 road sensor devices.

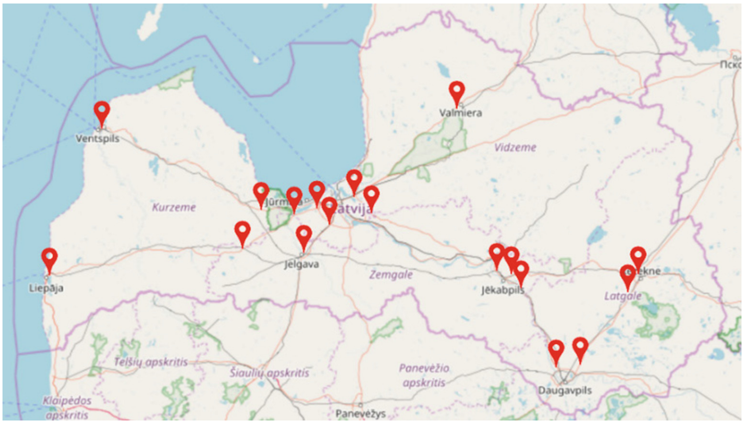


Fig. 7. Road sensor devices locations for covering rural areas between cities.

By introduction of the proposed system on the roads of Latvia, we estimate that average speed on main rural roads drops at least by 10–15 km/h. According to [9, 10] such change should decrease the road accidents rate by 55% and road fatalities rate by 40%. With these changes we expect that Latvia significantly improve safety on city roads and Latvia will no longer be a country with most dangerous roads in Europe.

Urban areas. Another part of proposed system introduction supposes the coverage of main roads in urban areas within the main cities of Latvia. Riga city is the biggest and the most important city that has to be covered.

To cover all main roads of Riga it is required to install road sensor devices in 18 locations. Figure 8 shows these locations on the map of Riga. Every main road with straight segments between crossings has road sensors – at the beginning of each segment and at the end of the segment.

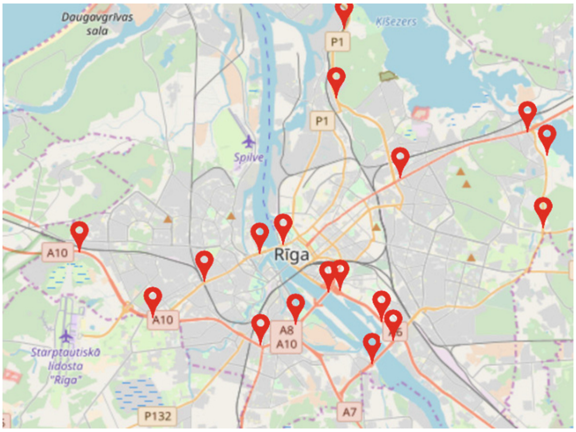


Fig. 8. Road sensor devices locations for covering main roads in Riga city.

By this introduction of the proposed system on roads of Riga we estimate the average speed decrease at least by 5–8 km/h. According to [9, 10] such change should decrease the road accidents rate at least by 35–40% and road fatalities rate by 25%.

With all these changes in rural and urban areas we expect that Latvia significantly improve safety on roads and move from the 1st place to 8th place in the list of countries with most dangerous roads in Europe with estimated no more than 12.0 fatalities per 100 000 motor vehicles.

5 Conclusions

In this paper a distributed road tracking system is proposed. This system has multiple devices installed on city roads equipped with cameras and automatic car number recognition module. All these devices are connected to a central server where they send information about the detected cars by providing information about recognized car license number, precise time, and location. The central server stores all this information and determines the amounts of cars various districts of the city in real-time, determines traffic congestions, collects statistics by time, days and seasons, and calculates average vehicles speed for travelling between points of detection.

In the two experiments we show that spot speed cameras have very limited local impact on driving speed – drivers are slowing down exactly near the speed camera and accelerate straight after passing it. What is very important, drivers afraid of fines and slow down when they are checked on speed. This means that the longer is the speed check distance, the longer drivers would follow the speed limit. This allows us to make a strong assumption that proposed system introduction will have the desired effect of reducing the average speed of cars on the roads and contribute to road safety.

References

1. Road safety in the European Union, Trends, statistics and main challenges, March 2015, ISBN 978-92-79-45654-1 (2015)
2. CSDD Homepage. https://www.csdd.lv/stacionarie-fotoradari/vispariga-informacija. Accessed 21 Sep 2017
3. Valsts Policija Homepage. http://www.vp.gov.lv/?id=803. Accessed 18 Aug 2017, (in Latvian)
4. Altintasia, O., Tuydes-Yamanb, H., Tuncayc, K.: Detection of urban traffic patterns from Floating Car Data (FCD). Transp. Res. Procedia **22**, 382–391 (2017)
5. Demirel, H., Shoman, W.: Spatio-temporal analyses for dynamic urban road network management. Transp. Res. Procedia **22**, 519–528 (2017)
6. Klūga, M.: Šogad uz Latvijas ceļiem plānots uzstādīt 20 jaunus fotoradarus. http://www.lsm.lv/raksts/zinas/latvija/sogad-uz-latvijas-celiem-planots-uzstadit-20-jaunus-fotoradarus.a217447/, (in Latvian)
7. Insurance Institute for Highway Safety. http://www.iihs.org/iihs/topics/t/roadway-and-environment/fatalityfacts/roadway-and-environment
8. Latvijas iedzīvotāju skaits pašvaldībās. http://www.pmlp.gov.lv/lv/assets/documents/Iedzivotaju%20re%C4%A3istrs/07022017/ISPV_Pasvaldibas_iedzivotaju_skaits.pdf
9. Richards, D.C.: Road Safety Web Publication No. 16, Relationship between Speed and Risk of Fatal Injury: Pedestrians and Car Occupants. Transport Research Laboratory, Department for Transport, London (2010)
10. World Health Organization, Road traffic injuries fact sheet. http://www.who.int/mediacentre/factsheets/fs358/en/
11. Tefft, B.: Impact Speed and a Pedestrian's Risk of Severe Injury or Death. AAA Foundation for Traffic Safety, Washington, DC (2011)
12. Eurostat, Transport Statistics at Regional level. http://ec.europa.eu/eurostat/statistics-explained/index.php/Transport_statistics_at_regional_level

Introducing Fixed-Wing Aircraft into Cooperative UAV Collision Avoidance System

Dmitrijs Lancovs(✉)

Transport and Telecommunication Institute, Lomonosova 1, Riga 1019, Latvia
dmitrijs.lancovs@outlook.com

Abstract. Unmanned aerial vehicles (UAV) have become popular means of carrying light payloads for survey, mapping, delivery and various other purposes. Collision avoidance mechanisms are being actively researched, but non-cooperative solutions are still unreliable. A cooperative solution is required, based on commercial off-the-shelf hardware to provide this capability.

This system would be classified as "Class A" in manned aviation based on potential for human injury or death in case of failure, requiring a specific level of reliability.

A stepwise approach to designing such system was proposed, starting with initial operating range and then integrating various internal and external factors at each stage of modelling. UgCS mission planning and flight control software was used, and model was verified and validated.

First results of modelling for multicopter craft were produced, establishing operating ranges of ideal transmitter and receiver in a deterministic, noiseless environment, with a completely reliable channel.

This article introduces fixed-wing airframes into the simulation and updates initial operating range requirements, performing simulation across all the previous scenarios with fixed-wing UAV encounters against other fixed-wing UAVs as well as multicopters.

Initial operating range requirements were updated. No significant changes in operating range were observed, and it still is significantly shorter than that of automatic dependent surveillance-broadcast (ADS-B). Transmitters and receivers operating within this range exist and are used on commercial UAVs, albeit for different purposes, indicating that such system is feasible with current technology.

Keywords: Unmanned aerial vehicles
Automatic dependent surveillance–broadcast · Unregulated airspace
Collision avoidance system · Simulation

1 Introduction

Advances in electronics, automation and electric power storage have provided a combination of capabilities necessary for commercial unmanned aircraft. Such craft are already being used for various purposes, such as photogrammetry, inspection, security,

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_38

while more research and development is done in the field of unmanned cargo delivery systems.

These solutions, however have a weakness: unless ICAO-certified and registered, such craft are restricted to operate in unregulated airspace, or require special airspace reservations. In case of the former, this is usually a thin layer of airspace between ground level and up to a couple hundred meters above it.

Such operating environment is shown on Fig. 1. Note that in an urban setting, UAVs are operating above people and property, and therefore present a hazard in case of failure. Also, as more craft are entering airspace, it may become densely populated, making collisions a significant concern, especially in a completely autonomous operating mode.

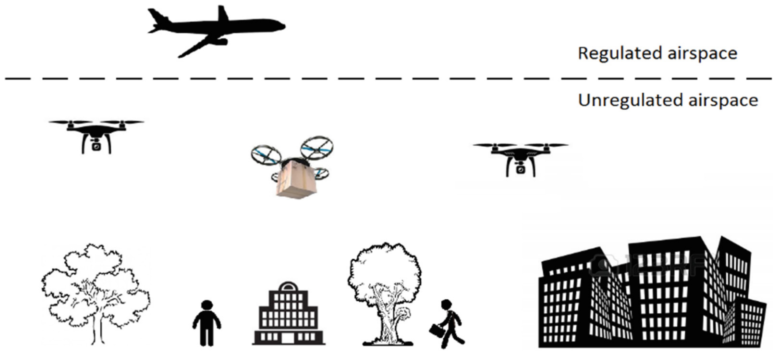


Fig. 1. Operating environment for commercial UAVs [1].

However, while failure can be prevented by redundancy of craft systems, there are no reliable solutions for collision avoidance with onboard means [1], and currently UAVs are not legally permitted to fly completely autonomously in EU, being restricted to within line of sight operations.

Indeed, both radar and lidar tech is feasible for collision avoidance [2, 3], but ranges, weight and power requirements preclude use of said devices on smaller UAVs, while their collision detection capability is limited to larger objects.

Optical solutions are much more robust at obstacle detection, but do not provide high accuracy in range assessment, also preventing their use in this scenario.

A cooperative system for collision avoidance was proposed by the author, and, since it deals with potentially lethal falling hazard to humans, it should be classified as "Class A" system [4], requiring a specific level of reliability – with the probability of failure at 10^{-9}.

Collision avoidance was proposed to be done cooperatively, in a similar fashion to ADS-B. However, there are many notable differences. ADS-B is a long-range system with limited precision (not needed for ranges and speeds at which manned aviation operates) and infrequent messages at 1–2 per second.

Craft broadcast messages about their type, heading, bearing, speed, etc. Other craft receive this information and make decisions whether there is potential for collision or not. In the event of a potential collision a set of rules defines how to resolve it. To aid in decision-making, some additional information is transmitted, such as more detailed craft capability report.

Also, having updates once or twice per second (as in typical ADS-B implementations) is not enough to guarantee reception of message in a real-life situation with other broadcasters and noise, which would restrict UAVs at maintaining unreasonably large distances between each other (in case of manned craft such distances are reasonable, since there is a limited number of aircraft).

Proposed system should work with significantly more frequent updates. The latter requirement means that in a saturated environment transmit range should be kept as low as possible, since all UAVs share the same channel. ADS-B transmits up to several hundred kilometers, and this provides a notable power draw and potential interference with onboard electronics for smaller UAVs.

However, range should still be sufficient to prevent collisions between all kinds of UAVs that may rely on the system, while providing "Class A" reliability.

Several stages were proposed to design such system, among them initial state required to establish its operating range. As stated above, this should be kept minimal, yet sufficient for various UAV types involved in future traffic.

Once that range is established, subsequent stages introduce uncertainties coming from transmit channel and equipment, other broadcasters, radio frequency noise, etc. At the end of the design phase a set of requirements is produced for physical implementation of the system.

Earlier experimentation [5] has provided data on multicopter ranges. However, fixed-wing aircraft have not been included in those experiments, and must be simulated now in order to obtain initial operating range requirements for use case involving such craft.

2 Collision Avoidance Scenario Simulation

2.1 Simulation Environment

Simulation was performed in UgCS software [6]. The purpose was to identify minimal range required to successfully perform collision avoidance maneuver in a potential collision scenario among various craft types. Environment is deterministic, decision to perform collision avoidance is completely reliable, transmitter, receiver and channel are not factored in at this stage.

Simulation consisted of a flight path that imitates actions taken during collision avoidance, and an emulator – a physical simulation of a specific craft type. Several scenarios (see Table 1) were used with craft positioned in different courses and reacting in a different way. At the end of each scenario a specific minimal safe range was obtained.

Scenarios are divided into three groups, each representing various states of the system. First group deals with craft cooperatively avoiding each other. This is the most likely scenario, with both reports and avoidance decisions produced by every craft in the airspace.

Second group involves a situation when one of the craft can only report its course, but not alter it.

Table 1. Simulation scenarios [5].

Aircraft state	Course	Description
Both aircraft moving, both responding	Head-on	Incoming straight at each other
	Intersect	Incoming at a right angle
	Chase	Faster craft chasing slower one
	Climb/descend	Head on with vertical resolution
Both aircraft moving, only one responding	Head-on	Incoming straight at each other
	Intersect	Incoming at a right angle
	Chase	Faster craft chasing slower one
	Climb/descend	Head on with vertical resolution
One aircraft stationary and not responding	Head-on	Incoming at a stationary craft
	Climb	Head on with vertical resolution
	Descend	Head on with vertical resolution

Third group represents a case with stationary UAVs, such as tethered systems, that are fed power through a wire to the ground, and are unable to move away from their position. This data may also be used for stationary objects with beacons, such as radio towers. Note that in case of tethered systems, vertical collision resolution is only possible by climbing, as a descending course may result in hitting the wire.

Still, it may be possible to have a UAV that is not tethered, yet cannot move away from its position for whatever reason, and a situation like this may involve diving under it, therefore such scenario is also present.

2.2 Previous Experiment Results – Multicopters

Data on multicopters was already obtained in previous experimentation [5]. It simulated three common UAV types in every scenario from Table 1. Results for each scenario were obtained and are presented in Table 2.

Table 2. Simulation results for multicopters [5].

Aircraft state	Course	d_{min}, m	d_{max}, m
Both aircraft moving, both responding	Head-on	5.50	11.41
	Intersect	9.02	12.70
	Chase	2.80	54.74
	Climb/descend	6.33	12.59
Both aircraft moving, only one responding	Head-on	10.57	98.35
	Intersect	7.48	90.39
	Chase	4.61	4.61
	Climb	21.13	130.86
	Descend	22.68	132.37
One aircraft stationary and not responding	Head-on	5.57	12.37
	Climb	11.13	18.24
	Descend	8.97	15.46
Resulting values		2.80	132.37

Minimal sufficient range per scenario is shown in d_{min} column, while maximum is in d_{max} column. Result is the smallest and largest ranges observed. As can be seen here, ranges are significantly shorter than what is provided by ADS-B.

2.3 Experiment Setup

No significant changes were made to the simulation other than replacing the aircraft with a different type. Route approach distance has been lengthened to make sure fixed-wing simulator is travelling straight when reaching the collision avoidance waypoint.

Vertical resolution was set to a 45° desired slope path, but the simulator was instructed to fly at a maximum angle it can support. As can be seen in Fig. 2, craft could not support such a steep angle.

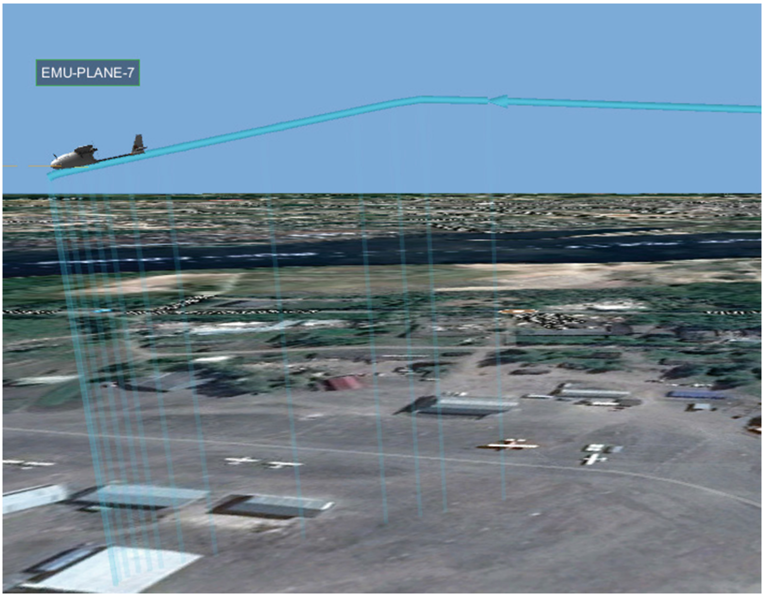


Fig. 2. Path performed by a fixed-wing aircraft.

2.4 Current Experiment Results – Fixed-Wing Aircraft

The purpose of this article was to add fixed-wing aircraft to the simulation and obtain range requirments for them. This implies not only simulating encounters among fixed-wing aircraft, but also their interaction with multicopters.

While fixed-wing aircraft can have a wide range of capabilities, simulation was done using typical "flying camera" setup, with the craft of 2-m wing span, controlled by PixHawk autopilot [7], and maximum flight speed of 25 m/s. It was expected that ranges would significantly differ from those for multicopters thanks to completely different aerodynamic setup. The opposite was observed (see Table 3).

Table 3. Simulation results for fixed-wing craft and fixed-wing craft vs multicopters.

Aircraft state	Course	d_{min}, m	d_{max}, m
Both aircraft moving, both responding	Head-on	44.60	99.13
	Intersect	33.63	68.88
	Chase	2.26	40.21
	Climb/descend	16.09	51.05
Both aircraft moving, only one responding	Head-on	30.57	138.56
	Intersect	25.61	102.26
	Chase	5.56	33.56
	Climb	48.40	168.55
	Descend	35.05	112.37
One aircraft stationary and not responding	Head-on	48.56	48.56
	Climb	26.60	33.40
	Descend	25.05	25.05
Resulting values		2.26	168.55

While there indeed are differences in various specific scenarios, most notably head-on and intersect situations, in others we see similar values. Chase is still the best situation to avoid a collision, in both cooperative and non-cooperative situation, while vertical resolution is still the costliest.

3 Analysis of Simulation Results

3.1 Overview

Results obtained in the second experiment provide a slightly larger operating range [2.26, 168.55] m, which completely overlaps that from the first experiment [2.80, 132.37] m. It is still very similar to the first range, only 32 m longer.

Furthermore, as noted above, biggest difference is in head-on scenarios, since the plane has to bank to make a turn, while multicopters can change heading without having to turn at all.

Also of note is a limit imposed on a fixed-wing aircraft in sustainable climb and descent angles. Those can differ significantly between different aircraft and it could potentially be advantageous to perform vertical collision resolutions for such craft. The simulated aircraft, however, is very limited in such performance, so expecting every UAV capable of such resolution is unreasonable.

Multicopters have already shown that horizontal collision resolution is preferable whenever possible. Still, as a potential solution when horizontal resolution is unavailable due to environments (which may include buildings, tall trees, etc.), it can be used.

However, in this case having a set of rules on collision avoidance is not enough; a negotiation between craft may be required to decide which resolution to use. This requires further study.

3.2 Minimum Operating Range

Both experiments have shown that it is technically possible, at times, to avoid a collision within very narrow ranges. This is expected, especially in a chase scenario with faster craft following a slower one, if their speeds are close.

Considering that GPS and GLONASS precision is limited, it makes little sense to rely on collision resolution at such short ranges. Instead, setting a minimal allowed distance between UAVs is preferable. Entering such distance should be prohibited, even if craft speeds and agility could theoretically allow them to operate closer to each other.

The most logical solution would be to choose this minimal permitted distance based on performance of present day GPS/GLONASS sensors, as well as data that would be obtained from next stages of current research, when transmitter, receiver and channel are introduced into the model.

3.3 Maximum Operating Range

Second experiment has shown that even with a slower fixed-wing aircraft (25 m/s) than a multicopter (as in the first experiment, 30 m/s), maximum operating range has extended by 32 m. With faster yet craft it is expected to grow further.

At the same time, this is a minor difference, and mostly due to how fixed-wing aircraft turn, requiring more time before they shift from the course by a necessary amount.

This maximum operating range is still within what is possible in commercial radio equipment used for various purposes. It is also by several orders of magnitude lower than what ADS-B provides, showing, again, that ADS-B transmit power is not necessary for small commercial UAV systems in unregulated airspace.

One concern is channel saturation by craft broadcasting updates. While frequent updates are desired, a balance between range and update frequency should be observed.

Further limitations may be imposed based on local radio broadcast laws and regulations. However, the range established here is the absolute minimum necessary for the system to work reliably, in case of ideal transmitter, receiver and noiseless channel. It is expected to grow when transmitter, receiver and channel models are introduced into the system. Limitations will also arise from power requirements for said hardware and radio frequency noise.

4 Conclusion

Both minimum and maximum ranges for collision avoidance system were established, now with fixed-wing aircraft accounted for. There were no significant changes in required range, and it remains technically feasible with modern commercial off-the-shelf hardware.

One interesting observation is that in certain scenarios vertical collision resolution was preferable. It is not surprising, considering manned aviation uses such resolution, as it appears preferable for fixed-wing aircraft. Multicopters, however, display preference for horizontal collision resolution.

This observation presents a challenge, especially in situations when either horizontal or vertical resolution may be unavailable due to stationary obstacles. Rules cannot be established to cover for both possibilities at the same time, unless some negotiation mechanism is implemented, taking this system one step further than ADS-B, which passively informs other craft of its course and relies on collision resolution rules, but doesn't negotiate course changes with partners, leaving that task to the dispatchers.

It is still technically possible to design such system, and it could still work as an infrastructure-independent system, but would then be reclassified from "cooperative" to "interactive", since decision making would be performed by both craft.

Since non-cooperative scenarios were also simulated, safety would be assured in case one of the craft fails to negotiate as well.

At the same time, question of security remains. ADS-B is prone to spoofing [8], something that a similarly designed system would be susceptible to as well. Encryption methods could be used, similar in digital signature area, to fight this capability.

First stage of designing a cooperative, infrastructure-independent collision avoidance system for UAVs in unregulated airspace is complete, and operating range has been established.

Future research will incorporate radio equipment used by craft to communicate. Signal encoding, message frequency and content should be chosen, based on channel bandwidth and reliability. This in turn requires identifying said parameters.

Additional phases such as encryption and possible negotiation between craft are expected to introduce additional delays between potential collision detection and resolution, which, at UAV speeds, translates into range increase. Indeed, d_{min} and d_{max} are expected to grow at further stages of research, and the current decision to provide d_{min} cutoff at distances based on navigation precision of GPS/GLONASS may get reversed.

Such changes are expected, and are the very reason why staged approach is used in collision avoidance system design.

References

1. Lancovs, D.: On sharing of uncontrolled airspace for low flying unmanned aerial vehicle systems. In: Kabashkin, I., Yatskiv, I. (eds.) Proceedings of the 15th International Conference Reliability and Statistics in Transportation and Communication, pp. 214–222. Riga, Latvia (2015)
2. Moses, A., Rutherford, M.J., Kontitsis, M., Valavanis, K.P.: UAV-borne X-band radar for collision avoidance. Robotica **32**(01), 97–114 (2014)
3. Sabatini, R., Gardi, A., Richardson, M.A.: LIDAR obstacle warning and avoidance system for unmanned aircraft. Int. J. Mech. Aerosp. Ind. Mechatron. Manuf. Eng. **8**(4), 718–729 (2014)
4. Youn, W.K., Hong, S.B., Oh, K.R., Ahn, O.S.: Software certification of safety-critical avionic systems: DO-178C and its impacts. IEEE Aerosp. Electron. Syst. Mag. **30**(4), 4–13 (2015)
5. Lancovs, D.: Building, verifying and validating a collision avoidance model for unmanned aerial vehicles. Procedia Eng. **178**, 155–161 (2017)
6. SPH Engineering, Universal Ground Control Software. https://www.ugcs.com/. Accessed 25 Aug 2017
7. Pixhawk.org, Pixhawk autopilot. https://pixhawk.org/. Accessed 25 Aug 2017
8. Costin, A., Francillon, A.: Ghost in the air (traffic): on insecurity of ADS-B protocol and practical attacks on ADS-B devices. In: Proceedings of Black Hat USA 2012 (2012)

Telematics

The Importance of Mapping Regional Disparities and Regional Development: Case Study of the ICT Sector in the Slovak Republic

Emilia Madudova(✉)

Žilinská univerzita v Žiline, Univerzitná 8215/1, 010 26 Žilina, Slovak Republic
Emilia.Madudova@fpedas.uniza.sk

Abstract. The paper deeply describes regional disparities of the ICT sector in the Slovak Republic and the importance of measuring these disparities due to the unequal localization of ICT firms in the individual regions of the Slovak Republic. The ICT industry in the Slovak Republic (according to the Statistical Office of the Slovak Republic) has been one of few industries in which the employment has continuously grown. This sector has the lowest unemployment rate of all sectors of the economy in the SR as well. As concerns a number of exports, the ICT sector follows the automotive industry by a narrow margin, with the relative share of imports as well as exports of the Slovak ICT sector being one of the highest in the world. The existence of disparities leads to lower production and scarce use of available resources as well but also differences in living standard. The results show that that location of ICT firms in individual regions of SR contributes to higher employment, rising living standard and other positive economic influences conducive to economic growth in the regions and contributes to regional development. Many authors have already been interested in concepts that include a variety of phenomena that can be evaluated through quantitative and efficacious indicators. As findings shows, the ICT sector is the driving force of the Slovak economy and boosts economic development. All these factors mentioned above are the reasons why to pay attention to regional disparities in ICT sector in the Slovak Republic is so important.

Keywords: Regional disparities · Regional development · ICT sector

1 Introduction

The development of a region is affected by concepts linked to the ability to lagging economic conditions [1, 2]. Local economic development may be encouraged by the realization of a regional competitive advantage based on location–specific and specialized capabilities and competencies nurtured by socio-institutional and cultural structures. Since such conditions are context-specific, they are extremely difficult to replicate in different settings and each location has to shape its own competitive advantage on the basis of functional and effective interactions between local economic agents and socio-institutional forces [3].

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_39

However, not all locations are characterized by such favourable conditions and [4, 5] suggested that national growth is often led by few fast-growing and innovative places within a country, mostly coinciding with large metropolitan areas. As a consequence, when looking at the performance of sub-national entities it is not surprising that welfare concentrates in these same few places giving potentially rise to spatial inequalities at the national level.

Even if economic development processes are highly localized not all 'locations' are equally able to succeed in the global competitive environment. In fact, knowledge and innovation activities require a favourable environment to make positive feedback and interactions possible. Therefore, differences in local social, political and institutional settings determine different interactions between local economic agents, knowledge and innovation activities [6].

Cumulative and path-dependent processes of accumulation of knowledge shape the distribution of welfare across space, suggesting the existence of a more complex economic geography than that of a flat world. In other words, economic development is ultimately spurred at the local level where knowledge externalities are generated. As a matter of fact, while codified knowledge becomes largely available and accessible as a result of improvements in communication technologies, tacit knowledge remains spatially bounded and its economic value has even increased as a consequence of its relative scarcity in respect to codified knowledge [7]. On the other hand, some economists have long recognized that technological change is one of the most important forces driving economic growth, together with human capital and knowledge accumulation [8, 9].

1.1 The Importance of ICT

The perspective of Information and Communication Technology (ICT) as communication measures is related to the demand side of the economy [10]. The scope of the economic effect of the divide in East Asia, despite the region having captured a high share of the global production of ICT goods was estimated [11, 12]. Authors found that a significant digital divide exists between the five leading countries in East Asia and the other seven developing countries [13, 14].

From a regional perspective, it is important to know whether changes in an ICT sector have any impact in terms of economic development and innovation dynamics differentials across the Slovak Republic. The diffusion of ICT across the economy appears to be a major level for improving both productivity levels and competitiveness.

The ICT industry in the Slovak Republic (according to the Statistical Office of the Slovak Republic) has been one of few industries in which the employment has continuously grown. This sector has the lowest unemployment rate of all sectors of the economy in the SR as well. As concerns a number of exports, the ICT sector follows the automotive industry by a narrow margin, with the relative share of imports as well as exports of the Slovak ICT sector being one of the highest in the world.

The Slovak ICT sector has maintained growth trajectory over the past period of 2016. A performance indicator of value added has annually increased by 11.4% and sales increased by 18.3% to 2.96 billion EUR. The cumulative profit is by almost the one-fifth higher as well. The key features of the year 2016 are finishing of major projects, more pronounced start of the international activities, acquisitions and introducing a new services and goods on the market [15, 16].

2 Methods

The spatial effects have been calculated by the exploratory data analysis by means of descriptive statistics that have been extended to the spatial domain with the Location Index, Location Coefficient, Coefficient of Specialization, Coefficient of Concentration, Theil Index, Lorenz Curve and Gini coefficient. The paper has approached the issue of both sectorial ICT specialization of regions and geographic concentration of the ICT sector. Regional specialization is usually analysed in connection with industrial concentration and its distribution in the regional dimension.

Location index (LI) measures a region's industrial specialization relative to a larger geographic unit. An *LI* is computed as an industry's share of a regional total for some economic statistic divided by the industry's share of the national total, where, X_{ij} means number of employees of the i-th sector in the selected region, Y_i – number of employees of the i-th sector in the relevant country, S_i population in the selected region and S – population in the relevant country.

$$IL = \frac{\frac{Xij}{Yi}}{\frac{Sj}{S}}. \tag{1}$$

Location Quotient (LQ), the employment share of sector i in region j in all sectors employment of region and the employment share of industry i in all industries of the entire country, where x_{ij} means regional industry employment, x_{kj} – total regional employment, x_i state industry employment and x_k total state employment.

$$LQ = \frac{\frac{X_{ij}}{\sum_{k=1}^{n} X_{kj}}}{\frac{X_i}{\sum_{k=1}^{n} X_k}}. \tag{2}$$

Hoover Coefficient of Specialization measures specialization through measuring industrial structure difference of inter-region.

$$CS = \frac{X_{ij}}{\sum_{k=1}^{n} X_{kj}} - \frac{X_i}{\sum_{k=1}^{n} X_k}. \tag{3}$$

Hoover Coefficient of Concentration compares the intensity of one phenomenon in the region with the intensity of the other phenomenon in the region.

$$CC = \frac{X_{ij}}{X_i} - \frac{\sum_{k=1}^{n} X_{kj}}{\sum_{k=1}^{n} X_k}. \tag{4}$$

Theil index improves upon Gini Coefficient by identifying the share of inequality attributable to the between-region component. It is a better tool for the analysis of regional inequality as it suggests the relative importance of spatial dimension of inequality, where x_j means employment in the region, x – employment in a state, y_j – average regional wage, $\bar{y}$ – average wage in the state, m – number of regions.

$$TE = \sum_{j=1}^{m} \frac{x_j}{x} \cdot \frac{y_j}{\bar{y}} \cdot ln\frac{y_j}{\bar{y}}, \tag{5}$$

where

$$\bar{y} = \sum_{j=1}^{m} \frac{x_j \cdot y_j}{X}. \tag{6}$$

In order to consider the role and signification of ICT Industry, author based on the definition of the ICT sector according to the Statistic Office of the Slovak Republic. The statistical definition of ICT Sector through SK NACE rev.2 categorization only includes companies which allocate activities related to ICT as the main activity. The ICT sector according to the Statistical Office of the SR (SK NACE rev.2) includes Publishing Activities, Film, Video, TV Production, Broadcasting, Telecommunications, Computer Programming, and Information Services. The research focuses on ICT sector mainly on the most significant ICT Division 62 Computer Programming.

3 Results

The following measures of ICT industry, especially Division 62 regional specialization and concentration from 2010 to 2015 describes changing the trend of spatial and economics disparities in individual regions of the Slovak Republic.

Ranking of the fastest growing technological companies in the Middle Europe for this year includes five Slovak companies of division 62. Even when it is very complicated to state the proportion of ICT sector on GDP, several studies present the contingency between the capital investments of ICT and GDP increase. From this reason, the government should pursuit better conditions for the ICT sector, in order to appease the disparity of the ICT sector Slovak regions. Furthermore, the government should support the ICT sector with the view of potential economic growth.

In order to maintain a growing economy, regions should provide such conditions that attract business and firms to the region and persuade these firms of the advantages of location in the given area as well. The individual regions should also try to promote the attractiveness of the region and strengthen the competitiveness of the regional economies.

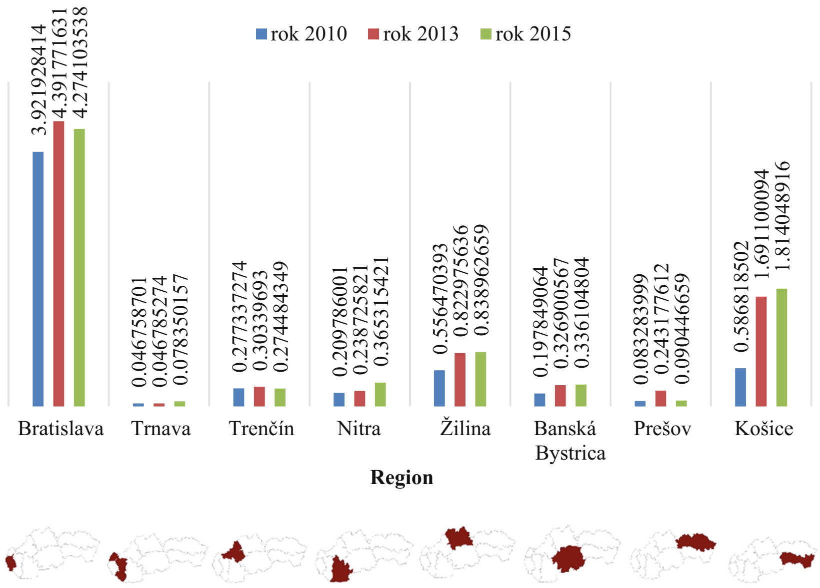


Fig. 1. Location index of ICT division 62: computer programming 2010–2015.

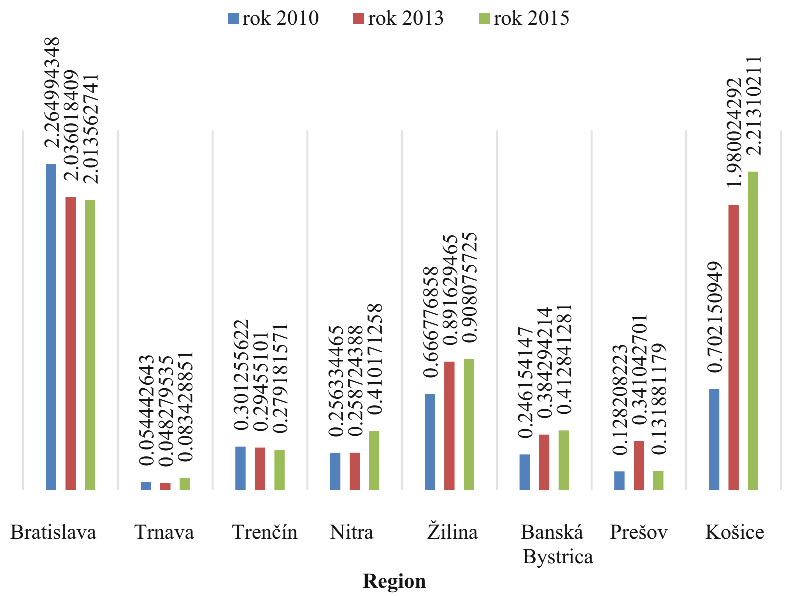


Fig. 2. Location quotient of ICT division 62: computer programming 2010–2015.

Figure 1 presents the development of Location Index in Division 62 for the years 2010 to 2015. The largest increase of this index in this time period was recorded by the Košice Region, with the value increased from 0.59 to 1.81. This change means, that in relation to the population, the location of ICT industry in Košice Region records growing trend. There is an ICT cluster located in this region. This cluster Košice IT Valley plays an important not only regional but also a national role as well. In 2015, this cluster was certified as the first in the central Europe and is one of three certified clusters in the area of information and communication technologies. In the central part of the Slovak Republic plays the important role the Žilina Region, even when the Location Index score was proportional to the population, in the course of 5 years, this index has increased, indicating an improvement in the sector.

The location has also been evaluated in relation to employment (Fig. 2). It is clear, that the employment concentration has increased the most in the Košice Region. Surprisingly decreasing trend this quotient occurred in Bratislava Region, where the capital city Bratislava is located. On the other hand, this can mean, that marked disparities in the past moderate at present. This fact can reflect the job creation and increasing working places in Division 62 in other regions, mainly in Košice, Žilina, Nitra and Banská Bystrica Region.

The specialization (Fig. 3) of Košice Region confirms also the Coefficient of Specialization. There can be seen the increasing trend in Žilina Region. Winding down disparities can be seen on declining coefficient of Bratislava region.

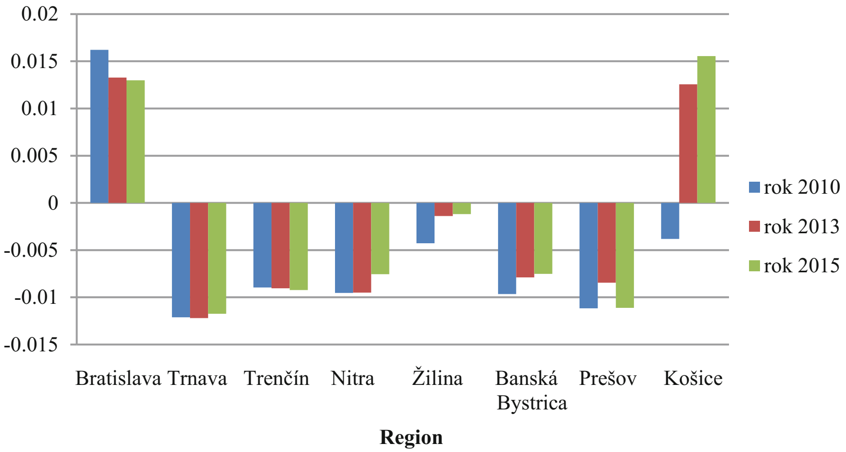


Fig. 3. Hoover coefficient of specialization of ICT division 62: computer programming 2010–2015.

The development of spatial distribution taking account employment in division 62 shows Fig. 4. Hoover coefficient of concentration confirmed trend of decreasing the regional disparities in individual regions of the Slovak Republic.

ICT sector has become an essential part of the world economy. The good sign is, that the ICT sector has established in the Slovak Republic. As research shown, the regional disparities in the Slovak Republic decrease and there can be found influential and significant companies in most of the regions of the Slovak Republic not only in Bratislava region as it was in the past.

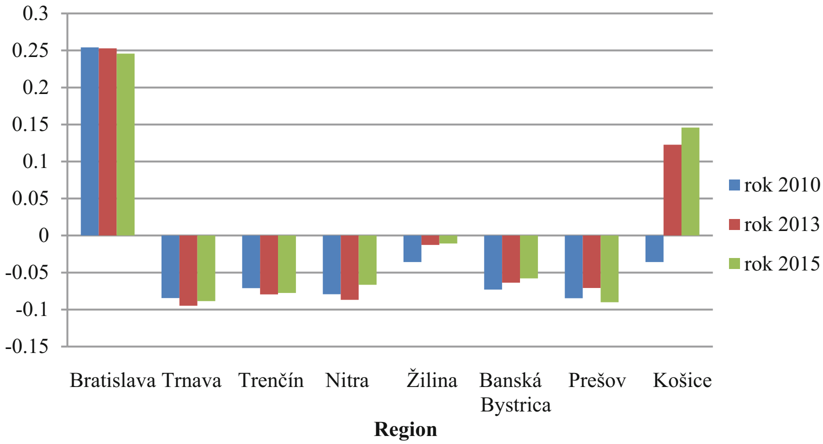


Fig. 4. Hoover coefficient of concentration of ICT division 62: computer programming 2010–2015.

All this reflects the ability of Slovak universities educate ICT specialists, a government regulation focused on promoting ICT sector and knowledge-based economy in the Slovak Republic. ICT sector has a great potential for economic growth and post-crisis economy recovery.

4 Conclusion and Discussion

Based on the research, we can conclude, that an ICT sector in the Slovak Republic faces an economic growth what can result in regional development. The final value of the Gini coefficient in 2015 represents the value of 0.13173 (when comparing average monthly wage and number of employees) in Division 62 and value of 0,0914 when comparing average monthly wage and number of employees of all the sectors and value of 0.09338 in ICT sector.

The Lorenz curves at Fig. 5 presents the 2015 distribution of income. Measuring changes in inequality helps determine the effectiveness of policies aimed at affecting inequality and generates the data necessary to use inequality as an explanatory variable in policy analysis. In Division 62, the average income varies from 1 041 EUR to 2 501 EUR with employment of (151–9 135) employees in individual regions, in ICT: from 987 EUR to 2 165 EUR and employment (623–22 075), and all sectors: from (800 EUR to 1 322 EUR and with employment of (151 515–354 113) regional employment). This implies a fact, that the distribution of incomes within the division 62 in Slovakia is far unequal as the distribution of incomes in all sectors.

Following table includes information used for calculation of Theil's index within the division 62 in Slovakia. The value of 0.04 presents the relatively equal distribution of monthly income within the division 62 in individual regions in regard to a number of regions concerning a number of employees within this sector (Table 1).

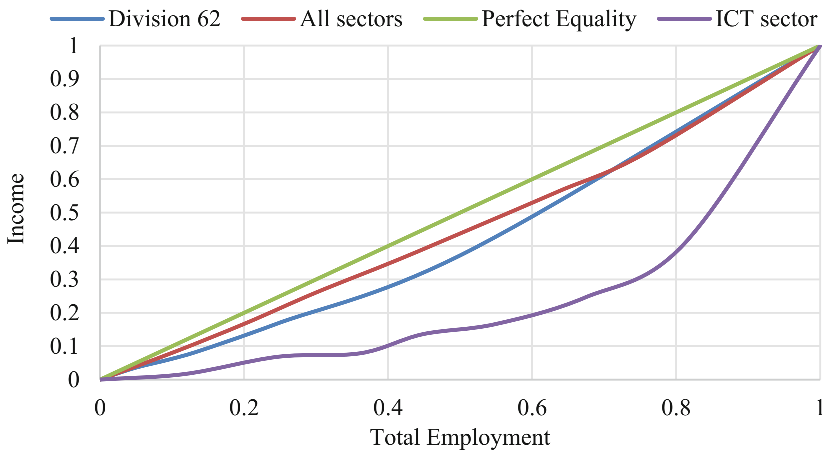


Fig. 5. Lorenz curves of 2015 values.

For comparison, Theil's index has also been calculated for all the sectors (0.016). This implies a fact, that the distribution of incomes within the division 62 in Slovakia, is far unequal as the distribution of incomes in all departments.

Table 1. Theil index of division 62 and all sectors.

Region	Average monthly wage division 62 (EUR)	Employment division 52	Theil index division 62	Average monthly wage all sectors	Employment all sectors	Theil index all sectors
Bratislava	2 501	9 135	0.142267302	1 322	354 113	0.090803883
Trnava	1 352	151	–0.002109141	938	141 273	–0.005521929
Trenčín	1 349	562	–0.007877832	909	157 126	–0.009031836
Nitra	1 041	867	–0.01567548	860	164 988	–0.014371637
Žilina	1 500	2 002	–0.022627109	920	172 084	–0.008703657
Banská Bystrica	1 236	763	–0.012020366	866	144 258	–0.012057141
Prešov	1 613	256	–0.002303998	800	151 515	–0.018296877
Košice	1 621	4 983	–0.043993608	947	175 747	–0.005843973
Slovak Republic		18 719	0.035659768		1 461 104	0.016976833

Gini coefficient, Lorenz curve, and Theil index has been chosen for better evaluation of disparities in individual regions. Two regions may have same Gini coefficient, but Lorenz curves can have for same area different shapes. Theil's index is more appropriate when comparing differently defined population subgroups. Thus Theil's Index improves upon Gini Coefficient by identifying the share of inequality attributable to the between-region component. It is a better tool for the analysis of regional inequality as it suggests the relative importance of spatial dimension of inequality.

Regional development should contribute to increasing the competitiveness of the region and reducing economic and social disparities. It must lead to economic, territorial and social development. Regional development cannot be tied to economic growth, but economic growth is a fundamental element of regional development. If the total volume of real gross domestic product in a given period is greater than in the previous period, economic growth is occurring. Even when factors of regional development vary in time, Slovak Republic should promote regional development policy on the way to knowledge-oriented economy.

Information and Communication technology sector together with automotive industry become a major source of entrepreneurship and job generation in the Slovak Republic. Many authors focus on the emergence of high-tech. knowledge-based entrepreneurship, but less attention is paid to the effect of the regional environment new firm formation and regional conditions region can offer.

Acknowledgement. This paper is an output of the projects No. VEGA 1/0515/15, VEGA 1/0693/16 and No. 5/KS/2017.

References

1. Romer, P.M.: Increasing returns and long-run growth. J. Polit. Econ. **94**(5), 1002–1037 (1986)
2. Romer, P.M.: Endogenous technological change. J. Polit. Econ. **98**(5), 2–15 (1990)
3. Soltes,. V., Stofkova Repkova, K.: The impact of business environment on regional disparities. In: Proceedings of the International Conference on Innovations in Science and Education, pp. 187–191. CBUIC, Prague (2016)
4. Rodríguez-Pose, A., Storper, M.: Better rules or stronger communities? on the social foundations of institutional change and its economic effects. Econ. Geogr. **82**, 1–25 (2006)
5. Carlino, G., Chatterjee, S., Hunt, R.: Knowledge spillovers and the new economy of cities. Federal Reserve Bank of Philadelphia Working Paper No 01-14 (2001)
6. Crescenzi, R., Rodríguez–Pose, A.: Reconciling top-down and bottom -up development policies. Environment and Planning **43**(1), 773–780 (2011)
7. Rodríguez–Pose, A.: Do institutions matter for regional development? In: Working Paper in Economics and Social Sciences, vol. 2, pp. 1–5. IMDEA Social Sciences, Madrid (2010)
8. Sonn, J., Storper, M.: The increasing importance of geographical proximity in knowledge production: An analysis of US patent citations, 1975–1997. Environ. Plann. **A40**, 1020–1039 (2008)
9. Buček, M., Rehák, Š., Tvrdoň, J.: Regionálna ekonómia a politika. Regional economics and policy. Iura Edition, Bratislava 2010
10. Buted, M., Stofkova, Z.: Relocalisation Trends in the Context of Globalization. In: Notes in Management Science (ISSGBM – IB), pp. 33–38. Hong Kong (2016)
11. Horrigan, J.: Online communities: Networks that nurture long-distance relationships and local ties. Pew Internet and American Life Project (2006). http://207.21.232.103/pdfs/PIP_Communities_Re-port.pdf
12. Wong, P.K.: ICT production and diffusion in Asia: Digital dividends or digital divide? Inf. Econ. Policy **14**, 167–187 (2002)
13. Billon, M., Lera-Lopez, F., Marco, R.: Patterns of combined ICT use and innovation in the European regions. J. Glob. Inf. Technol. Manage. **20**(1), 28–42 (2017)

14. Buyst, E.: Continuity and change in regional disparities in Belgium during the twentieth century. J. Hist. Geogr. **37**(3), 329–337 (2011)
15. Corejova, T., Al Kassiri, M.: Knowledge-intensive business services as important services for innovation and economic growth in Slovakia. In: CBU International Conference Proceedings 2016: Innovations in Science and Education, pp. 42–47. Central Bohemia University, Prague (2016)
16. Madudova, E., Kolarovszki, P.: Results of a cluster analysis of the ICT sector. In: Proceedings of the 2016 International Conference on Engineering Science and Management, pp. 1–2. Atlantis Press, Zhengzou (2016)

Quantitative Analysis of the Competitive Environment in the Electronic Communications Sector

Lucia Madleňáková[1(✉)], Mária Matúšková[1], Radovan Madleňák[1], and Paweł Droździel[2]

[1] University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia
lucia.madlenakova@fpedas.uniza.sk
[2] Lublin University of Technology, ul. Nadbystrzycka 38 D, 20 – 618 Lublin, Poland

Abstract. The contribution is devoted to the problem of determining the degree of concentration in relation to indicators typical of the electronic communications sector. At present, concentration monitoring is very important. The reason is the significant impact of the incumbent on some relevant markets, which is not particularly desirable from the point of view of end-user protection or existing barriers to entry. So, this situation is primarily in the countries of Central and Eastern Europe, even though the market has been fully liberalized for several years. The rapid growth of technologies in this sector is a significant factor in the growth of concentration too.

An important part for the determination of "concentration ratio" is represented by models, methods and procedures for identifying absolute or relative concentration. They differ substantially in the way they take into account the number of subjects who are the carriers of the indicator under observation. Sectoral concentration analysis is conducted through indexes. They monitor the competitive behaviour of the entities operating on the relevant market. This analysis provides information on the structure of the market to regulators and makes recommendations for market regulation. The concentration ratio of the strongest subjects and the Herfindahl-Hirschman Index of concentration is most commonly used in electronic communications. At this time, it is also important to look at the relative concentration in relation to the unequal distribution values of key market indicator in the relevant market. For this case, we present a variation coefficient in the contribution.

Keywords: Electronic communication · Protection of competition · Regulation Concentration ratio · Herfindahl-Hirschman index

1 Introduction

Implementing rules to protect economic competition in electronic communications is linked to monitoring the competitive environment. The competitive structure of the sector can have a different character depending on the number and size of the entities operating in given sector. Defining the influence of individual entities can be affected also by benefits resulting from competition in the whole group of other sectors, on the

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_40

horizontal or vertical plane. Horizontal concentration occurs by merging subjects operating in the same industry that produce the same or similar products and sell them in the same geographical conditions. The consequence of horizontal concentration is the increase of the size of the entities and decrease of the number of entities in the sector. Vertical concentration goes beyond the industry in that it expands business to supply industries (upstream integration), or penetrates to industries that consume its products (downstream integration).

The more the industry is concentrated in the hands of a small number of subjects, the smaller the competition in given market. Vice versa, decreasing concentration leads to rise of competition. The level of concentration thus serves for indirect measuring of the level of competition in the sector [10].

2 Theoretical Background of the Current State

Many researchers as Kwoka [15], Golan et al. [12], Bikker [5], and others deals with asserting the level of concentration in individual sectors of the economy. In terms of the Slovak economy, this issue is address especially in the works of Brezina [6, 7], Mihalčová [20], Kočišová [14], Čorejová et al. [8, 10, 11].

Measuring the level of concentration in the market plays a major role especially in regulated sectors. Network industries are a typical example, where the development of services and competition is not growing evenly in all relevant markets. Key attribute of this behavior can be the existing position of the incumbent. The sector of electronic communications is no exception, where the development and market potential grow rapidly. This is linked mainly to the development of ICT, networks and new services. The issues of market structure, regulation, and concentration in the field of electronic communications in Slovakia is discussed in the works [8, 10, 11, 17, 19] and others.

The state pays a great deal of attention to monitoring the level of concentration and its potential changes happening today due to acquisitions and fusions. The legal regulation of concentration deals with the external implementation of concentration, which in terms of economic competition is much more dangerous than natural development. Excess concentration can become a source of potential breach of rules for the protection of economic competition. This phenomenon is considered long-term as decisively negative despite the fact that it in short-term it is positive for individual actors. That's why it is the task of the state to play the role of a sponsor and ensure the protection of principles of economic competition through legislature [17, 18].

In the conditions of the Slovak Republic the Act No. 136/2001 on Protection of Competition [1] deals with the issue of concentration, which is in accordance with the legal regulation within the EU. According to said law, any procedures and agreements on part of entrepreneurs that would limit competition in any for and in any way are prohibited, whether it is in the field of prices or market division. Given the fact that companies are trying to increase their economic strength and competitiveness, which can be against the market, it is very important to regulate the concentration in individual sectors. There are many measures that control and approve concentration. Based on these measures, competent authorities can approve, limit, or prohibit concentration, if this is in breach to the conditions of fair economic competition that is imposed by the law. The Act No. 351/2011 on Electronic Communications [2] ensures the development of effective economic competition in the market of electronic communications and the regulatory authority carries out supervision.

2.1 Development of the Value of the Telecommunications Market in Slovakia

Strong competitive environment causing price erosion as well as ongoing regulation is factors faced by the Slovak telecommunication market for several years. In 2016 these cause another, albeit a slight decrease of value of the telecommunication market. The total value of the telecommunication market in 2016 compared to the previous year decreased by 0.5% and reached 1,815 billion Euros. Market decline occurred despite increase of the number of customers of telecommunication providers. However, the number of customers in the telecommunication market increased in all segments. Compared to the previous year it grew by 2.4% to almost 11 million active customers (Fig. 1).

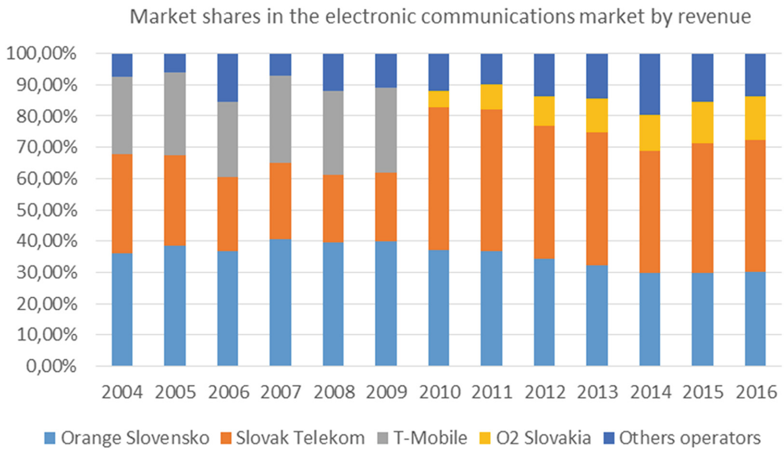


Fig. 1. Market share in the electronic communications market in Slovakia by revenues [13].

2.2 Development of the Value of the Telecommunications Market in Slovakia by Services

The dominant share in the value of the telecommunications market is still kept by the segment of mobile services, amounting to EUR 935 million, representing 51.6% of the total market value. The value of mobile voice and SMS has long been in decline, which is only partially offset by an increase in revenues from mobile data. Revenues of mobile telecommunication services in 2016 decreased by 0.7%, on the other hand, the revenues from fixed services and pay-TV grew by 0.7%. The fixed Internet segment recorded a

5.1% year-on-year increase in the number of connections, thus achieving a penetration of fixed Internet of 70% in Slovakia (Fig. 2).

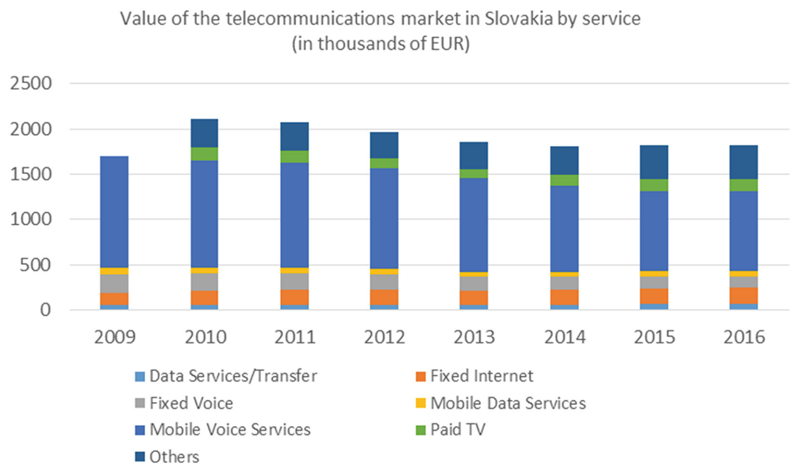


Fig. 2. Value of the telecommunications market in Slovakia by services [13].

3 Methodology of Concentration Measurement

There are a number of methods and ways how to measure and monitor the concentration. The basis for majority of the methods is to define the concentration indicators that characterize the status of the competitive environment in given sector. The important of concentration indicators results from their ability to capture structural features of the market. For this reason concentration indicator are oftentimes used in structural models that describe the relationship between the level of concentration in given sector and the market structure. Concentration indicators are also capable to capture changes of concentration due to changes in the number of entities in the market, e.g. when a new provider joins the market, a provider leaves the market, or when network or service providers merge [9, 16, 21].

As stated in the works of other authors, measuring the concentration indirectly ensures assessment of the structure and competition in the market. Most commonly used are the absolute concentration indicator "Herfindahl-Hirschman Index of concentration" and "the Concentration Ratio" in combination with other tools [6, 7, 10, 17]. To conduct an analysis of the telecommunication market and determining concentration indicators, the authors used primary data on revenue and performances acquired from relevant sources provided by the regulation authority of the Slovak Republic [22], dominant telecommunication providers (Slovak Telecom, a.s., Orange Slovensko, a.s., a O2 Slovakia, a.s.), and data published in IT yearbooks in the Slovak Republic [13]. However, due to unavailability of some data, it is not possible to publish comprehensive time developments as well, therefore we have selected data for the analysis that contained sufficiently long time period to get representative results.

4 Results and Discussions

Figure 1 depicts market shares and individual providers. As can be seen, during the monitored period there were no fluctuations in the telecommunication market. In connection to the analysed data only a smaller decline of sales of the dominant provider Slovak Telekom can be seen. However, this relates primarily to the liberalization of the market. The company Orange Slovensko is an obvious leader especially in the segment of mobile voice services, the market share of which grew 1% annually until the entrance of a third mobile provider. The breakthrough period (2007) in the development of concentration in the electronic communications market relates to the entrance of a third provider, O2 Slovakia, which was significantly reflected in revenues in 2008, and the share of this company since this moment grew c. 3% annually.

Another fluctuation in the market in 2010 was cause by the merger of the company Slovak Telekom and ST-Mobile, which significantly contributed to the decline of revenue of the company Orange Slovensko. After this merge, the Slovak Telkom and the Orange Slovensko are companies with a major influence in almost all monitored relevant markets. This is also reflected in activities of the regulatory authority, which is forced to intervene in these markets to maintain effective economic competition by determining the method, form, and conditions of behaviour in the relevant markets. In addition to regulation, the development of prices also reflects the effect of the economic crisis (decline in the purchasing power of companies and the population) and increasing price competition. Revenues of the largest providers have been declining also due to regulation of prices for accessing and interconnecting roaming.

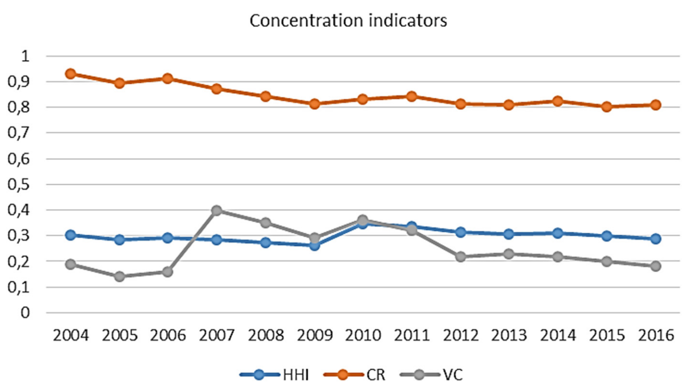


Fig. 3. Concentration indicators of telecommunication market in Slovak Republic: HHI, CR and VC.

In 2012 the Slovak Telekom held its first place in the overall telecommunication market with c. 42%, Orange with 34% and the O2 had a 9% share. In 2016 the market shares of these companies are around 41% for Slovak Telekom, 31% for Orange, 13% for O2, and a new mobile provider Swan Mobile achieved a 2% share in the mobile market [13, 22].

The resulting values of the concentration indicators are shown in Fig. 3. The market concentration is relative. Its value depends on the criteria used in the analysis.

4.1 Concentration Measurement by Herfindahl-Hirschman Index

For the purpose of determining concentration in the telecommunication market, the indicator HHI has been used, as well as market share values. We define according [8, 10] this rate is the sum of squares of market shares $s_i, (i = 1, 2, \ldots, n)$ of all companies in given field.

$$HHI = \sum_{i=1}^{n} \left(s_i\right)^2. \tag{1}$$

This index can have a maximum value of 1. This occurs if 1 company concentrates the offer in the sector and the minimum value 1/n is achieved if all companies have equal market share. However, due to its construction, the HHI does not reflect the impact of small companies on concentration in the sector, the share on industry production of which is less than 1%. The market share of a company or of a merged company can be used when determining concentration and its subsequent approval or rejection. The flow of values (see Fig. 3) is linked to the market analysis situation. Thus, we record a significant leap in 2010, when the concentration level from 0.2638 to 0.3477. In the next period the concentration level declines slightly. The entrance of a fourth mobile provider did not cause any significant change. Since the calculated concentration values are higher than the 0.18 limit, we can state that the market of electronic communications assessed based on HHI is highly concentrated during the entire monitored.

4.2 Concentration Measurement by Concentration Ratio

The concentration monitored based on the indicator concentration ratio (CR_m) [8, 10] allows to assess concentration for the strongest subjects in the market.

$$CR_m = \sum_{i=1}^{m} s_i. \tag{2}$$

By Čorejová et al. [10] this means that the CR_m index gives the same weight m to the largest providers ($w_i = 1;\ \forall I \leq m$) and zero significance ($w_i = 0;\ \forall i > m$) to the remaining small providers in the telecommunication market. This indicator can have values of $0 \leq CR_m \leq 1$. Values approaching one mean that m telecommunication providers included in the calculation of the indicator comprise the entire industry production, so the level of concentration is high. If the indicator value approaches 0, it means that there is an infinite number of small providers in the telecommunication sector of equal size. However, as opposed to HHI, the indicator concentration ratio reveals more fluctuations during the monitored period (see Fig. 3). These are slight fluctuations that have their reasons. Since until 2009 there were three dominant providers in the telecommunication market, for which the concentration ratio was indicated, its value declined slightly, this indicates strengthening positions of alternative providers. Another breakthrough was in 2011, when due to mentioned acquisition and position of the three major subjects (Slovak Telekom, T-Mobile and Orange), another subject, O2 Slovakia entered the market. Thus, in the following year we can observe a slight decline in the concentration level, which is again caused by increase of market share of alternative providers. The concentration ratio confirms strong concentration during the monitored period, as suggested by the development of values of the monitored indicator. Since 2010 we can talk about a tri-pol market structure.

4.3 Concentration Measurement by Variation Coefficient

The variation coefficient (VC) expresses to what extent is the impact distributed among market subjects, i.e. it characterizes the relative variability rate. It represents the equivalent of a relative concentration indicator to HHI, when determining the level of concentration it considers the specific number of subjects in the market during the observed period. As long as the value of the variation coefficient equals 0, we can assume that the market share in the industry is absolutely evenly distributed among the subjects. In practice, two relationships are used to calculate the variation coefficient [10].

$$VC = N \sum\nolimits_{j=1}^{N} \left(s_j - \frac{1}{N} \right)^2, \tag{3}$$

$$VC = N \times HHI - 1. \tag{4}$$

As Fig. 3 clearly shows, the values of the variation coefficient suggest that in 2007 the distribution of market shares was least even (value VC = 0.4). In the following years a more moderate divergence of VC values was identified, whereby 2010 was a breakthrough year also for this indicator. Both key players (Slovak Telekom, Orange) are recording annual decline of revenue, which is also reflected in the decline of values of the variation coefficient, but we still cannot talk about an effective economic competition model.

5 Conclusion

In connection to the value of the telecommunication market assessed through revenue, the goal of the paper was to assess the level of concentration in the Slovak telecommunication market during the monitored period through selected concentration indicators. Based on monitored data we can state that in 2015 the telecommunication market finally improved. The revenues of the largest companies grew slightly after years of continuous decline by one and a half per cent to 1.88 billion Euros. The decline of revenues is primarily the consequence of price competition that also causes lower dynamic of Internet growth. The use of absolute and relative concentration indicators seems as suitable and the presented results confirm changes in the concentration due to the entry of a new telecommunication provider or due to acquisitions.

In relation to the obtained values of listed indicators we can state a strong concentration in the telecommunication market. It was not significantly affected even due to the liberalization of the market between 2001 and 2003, or legislative and regulatory changes. However, we can expect fluctuations of concentration in the upcoming period. These can be linked to further expected acquisitions or possible future dynamic growth of the fourth mobile provider.

Acknowledgement. VEGA 1/0721/15 Research on the impact of postal services and telecommunication convergence on regulatory approaches in the postal sector.

VEGA 1/0515/15 Endogenous factors of the IPR intensive Industries in the regional enterprise environment in Slovak Republic.

References

1. Act No. 136/2001 on Protection of Competition. http://www.antimon.gov.sk/data/files/163_act-no-136-2001-on-protection-of-competition-amended-by-387-2011.pdf. Accessed 30 May 2017
2. Act No. 351/2011 on Electronic Communications. http://www.teleoff.gov.sk/data/files/22211.pdf. Accessed 30 May 2017
3. Anbarci, N., Katzman, B.: A new industry concentration index. Econ. Pap. **34**(4), 222–228 (2015)
4. Andrisková, I., Čorejová, T.: Analysis of market environment and factors influencing the demand for frequency spectrum. In: Proceeding of 7th International Scientific Conference Company Diagnostics, Controlling and Logistics, pp. 6–11. EDIS – University of Žilina, Žilina (2014)
5. Bikker, J.A., van Leuvensteijn, M.A.: New Measure of Competition in the Financial Industry: The Performance-Conduct-Structure Indicator, 1st edn. Routledge, New York (2014)
6. Brezina, I., Oršulová, A., Pekár, J.: Analyses of absolute concentration of the selected branch by Herfindahl-Hirschman index. J. Econ. **1**(57), 77–94 (2009)
7. Brezina, I., Pekar, J., Cickova, Z., Reiff, M.: Herfindahl-Hirschman index level of concentration values modification and analysis of their change. CEJOR **1**(24), 49–72 (2016)
8. Čorejová, T., Strenitzerová, M., Madleňáková, L., Majerčáková, M., Jankalová, M.: Network Economy. 2nd edn. EDIS – University of Žilina, Žilina (2010)
9. Čorejová, T., Imrišková, E.: Convergence at the postal market. Maintenance Reliab. **3**(39), 74–76 (2008)
10. Čorejová, T., Sedláková, I., Sedláková, E.: Analysis of electronic communication market in Slovak Republic by revenues in 2004–2012 years. Autobusy: Technika, Eksploatacja, Systemy Transportowe **6**(17), 1653–1656 (2016)
11. Čorejová, T.: Regulatory impact indicators in selected sectors and their comparison within V4 countries. In: Proceedings of International Scientific Conference POSTPOINT 2013, pp. 36–42. EDIS – University of Žilina, Rajecké Teplice (2013)
12. Golan, A., Judge, G., Perloff, J.M.: Estimating the size distribution of firms using government summary statistics. J. Ind. Econ. **1**(XLIV), 69–80 (1996)
13. IT Yearbook (2004–2016). https://www.pcrevue.sk/it-rocenka. Accessed 18 June 2017
14. Kočišová, K.: Concentration of Slovak banking sector and bank performance. Finan. Markets. Sci. J. **3**(11), 1–12 (2014)
15. Kwoka Jr., J.E.: Large firm dominance and price-cost margins in industries. South. Econ. J. **1**(44), 183–189 (1977)

16. Madleňáková, L., Madleňák, R.: The convergence of telecommunication and postal services and their impact on regulatory approaches. In: Proceedings of 15th International Conference RelStat 2015, pp. 158–165. Transport and Telecommunication Institute, Riga (2015)
17. Madleňáková, L., Majerčáková, M.: Legal Framework in Electronic Communications and Postal Services, 1st edn. DOLIS, Bratislava (2015)
18. Madudová, E., Kolarovszki, P.: Results of a cluster analysis of the ICT sector case study from the Slovak republic. In: Proceedings of the International Conference on Engineering Science and Management (ESM), pp. 156–159. Atlantis Press, Zhengzhou (2016)
19. Majerčáková, M.: Consumer Protection in the Electronic Communications Market. 1st edn. EDIS – University of Žilina. Žilina (2016)
20. Mihalčová, B., Hintošová, A.: Evaluation the concentration of the sector: theoretical approaches and legal tools. J. Econ. **8**(52), 984–995 (2004)
21. Pulaj, E., Kume, V.: Measuring market concentration of construction industry. Vlora region evidence. Eur. Sci. J. **9**(32), 121–136 (2013)
22. Regulatory Authority for Electronic Communications and Postal Services: Statistic Data from Electronic Communications. http://www.telecom.gov.sk/files/statistika_vus/telekom_ang/index.htm. Accessed 15 May 2017

The Analysis of the System that Includes Two Ferromagnetic Spheres in Outer Magnetic Field

Yury A. Krasnitsky(✉)

Transport and Telecommunication Institute,
1 Lomonosova Street, Riga 1019, Latvia
krasn@tsi.lv

Abstract. The procedure for solving the Laplace equation in bispherical coordinates for a system of bodies consisting of two spheres located in an unbounded medium parallel to an external magnetic field is considered. The boundary conditions are used consisting of the continuity of the potential and the normal component of the induction flux density on the surfaces of the spheres. Particular cases are investigated, when the spheres are the same, and $\mu >> \mu_0$.

Keywords: Bispherical coordinates · Laplace's equation
Magnetic field among the spheres

1 Introduction

The problem of the influence of a body of a certain shape on the external static field acting on it can be solved by analytical methods, provided that the coordinate system in which the variables in the Laplace equation can be separated is adapted to the shape of the body [1]. In other words, the coordinate surfaces of this system must coincide with the surface of the body. In other cases, one should turn to numerical calculations and modeling the standards for which the exact solutions of the boundary value problems always serve.

In electrostatics, the problem of two spherical conductors has been studied in some detail using bispherical coordinates and has numerous applications. Studies on the analysis of electric fields in the presence of spheres are related in particular to the theoretical verification of the Coulomb law [2], the determination of forces on particles and the amplification of the field in colloidal suspensions, the study of the behavior of water droplets and aerosol particles in the electric field [3, 4], the calculation of fields in material structures with built-in arrays of bodies of small electrical sizes [5] and the reaction of nanostructures to electromagnetic fields [6]. There are many examples [7] of the practical use of these results, including in medicine (deposition of inhalation aerosols on the surface of the lungs), food industry (development of electrostatic powder coatings for food), gas cleaning, etc. Appropriate solutions are useful in interpreting the results of geophysical prospecting [8].

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_41

In most of the literature, various particles and bodies of small electrical sizes are modeled as metal spheres located in an external electric field. At the same time, as far as is known [5], there is no exact analytical solution for the magnetic field in the presence of two magnetic spheres in the literature. It would be useful when using the concept of magnetic charges in electrodynamics and antenna theory.

In this paper, we are going to obtain the exact solution of the Laplace equation for a system of two spheres with $\mu \gg \mu_0$. With its help, it is possible to model the influence of the gap on the performance of magnetic circuits, in particular receiving ferrite rod antennas. However, the solution of this issue is beyond the scope of this article. Here we are going to consider the boundary-value problem of two ferromagnetic spheres embedded in a homogeneous and infinite medium in which, in the absence of spheres, there is a homogeneous magnetic field. The solution of the Laplace equation in a bispherical coordinate system makes it possible to find the distribution of the potential and field in the entire space, including the region between the spheres, i.e., in the gap. We will use existing solutions for conducting and dielectric spheres, relying on the dual-permutability property of Maxwell's equations. A particular case is considered, when the spheres are identical and $\mu \gg \mu_0$. The concept of effective magnetic permeability for a system of spheres is introduced.

2 Basic Information About Bispherical Coordinates

Consider two identical spheres of radius b made of a ferromagnetic material with permeability μ_i. The permeability of the environment is μ_e. The centers of the spheres are located symmetrically on the z axis with respect to the origin at the points $z = +a$ and $z = -a$ (Fig. 1). The length of the tangent drawn from the origin to any of the spheres is $c = (a^2 - b^2)^{1/2}$. The bispherical coordinates of the observation point are determined by the equalities (1), shown in Fig. 1.

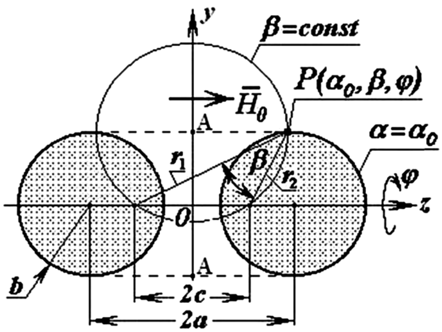


Fig. 1. Bispherical coordinates.

$$\left.\begin{aligned} x &= c\tfrac{\sin\beta\,\cos\,\phi}{\operatorname{ch}\alpha - \cos\beta} \\ y &= c\tfrac{\sin\beta\,\sin\,\phi}{\operatorname{ch}\alpha - \cos\beta} \\ z &= c\tfrac{\operatorname{ch}\alpha}{\operatorname{ch}\alpha - \cos\beta} \end{aligned}\right\}, \tag{1}$$

$$\alpha = \ln\frac{r_1}{r_2},$$

$$-\pi \le \beta \le \pi,\ \ \infty < \alpha < \infty,\ 0 \le \phi \le \pi.$$

It is easy to find that:

$$x^2 + y^2 + (z - c\,\operatorname{cth}\alpha)^2 = c^2/\operatorname{sh}^2\alpha \tag{2}$$

and, therefore, the surface α = const describes the sphere. For the left sphere $\alpha = -\alpha_0$; for the right sphere $\alpha = +\alpha_0$.

Geometric parameters of spheres are related to bispherical coordinates by relations

$$a = c\,\operatorname{cth}\alpha_0,\ \ b = c/\operatorname{sh}\alpha_0 \tag{3}$$

The values $\alpha = \pm\infty$ correspond to the points $x = y = 0$, $z = \pm c$. It also follows from (3) that when $\alpha_0 > 0$, then $a - b < c < a + b$, and these points lie inside the left and right spheres respectively.

The Laplace equation in bispherical coordinates has the form:

$$\begin{aligned} \nabla^2 V = \frac{(\operatorname{ch}\alpha - \cos\beta)^3}{c^2\sin\beta}&\left[\sin\beta\frac{\partial}{\partial\alpha}\left(\frac{1}{\operatorname{ch}\alpha - \cos\beta}\frac{\partial V}{\partial\alpha}\right) + \frac{\partial}{\partial\beta}\left(\frac{\sin\beta}{\operatorname{ch}\alpha - \cos\beta}\frac{\partial V}{\partial\beta}\right)\right. \\ &\left. + \frac{1}{\sin\beta\,(\operatorname{ch}\alpha - \cos\beta)}\frac{\partial^2 V}{\partial^2\varphi}\right] = 0, \end{aligned} \tag{4}$$

and its most general solution can be represented as a potential:

$$\begin{aligned} V = (\operatorname{ch}\alpha - \cos\beta)^{1/2}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}&\left(A_n e^{(n+1/2)\alpha} + B_n e^{-(n+1/2)\alpha}\right) \\ &\times\left[C_{nm}\,\mathrm{P}_n^m(\cos\beta) + D_{nm}\,\mathrm{Q}_n^m(\cos\beta)\right](F_{nm}\sin m\varphi + G_{nm}\cos m\varphi). \end{aligned} \tag{5}$$

Here $\mathrm{P}_n^m(\cos\beta)$ and $\mathrm{Q}_n^m(\cos\beta)$ are the Legendre functions of the first and second kind, respectively. The solution must be regular in the xz plane for $\alpha = \pi/2$, which holds if $D_{nm} = 0$.

Let the external magnetic field H_0 be oriented along the z axis (Fig. 1). Then the magnetic scalar potential V will have a cylindrical symmetry with respect to this axis, there is no dependence on the angle φ, and in (5) only the term with $m = 0$ should be retained, from which

$$V = (\operatorname{ch}\alpha - \cos\beta)^{1/2} \sum_{n=0}^{\infty} \left(A_n e^{(n+1/2)\alpha} + B_n e^{-(n+1/2)\alpha} \right) \mathrm{P}_n(\cos\beta). \tag{6}$$

3 Calculation of Scalar Magnetic Potentials

In general, the radii of spheres can differ from each other. We denote by α_1 and α_2 the coordinates of the surfaces of the right and left spheres, and by V_1, V_2 and V_3 – the potentials inside them and in the external medium, respectively. The potential due to the external field, in the absence of spheres, is equal to $V_0 = -H_0 z$, if the vector $\mathbf{H}_0$ is oriented in the direction of increasing z. The potential V_0 is antisymmetric with respect to the xy plane (odd in z, that is, for $z < 0$ there is a transition from α to $-\alpha$).

The potentials V_3 and V_1, V_2 also have this property:

$$\left.\begin{aligned} V_3(-\alpha,\ \beta) &= -V_3(\alpha,\ \beta), \\ V_1(-\alpha,\ \beta) &= -V_2(\alpha,\ \beta). \end{aligned}\right\}. \tag{7}$$

To find the potential V_3 in the outer region, we use the symmetry condition (7) and the circumstance that $V_3 \to V_0$ as $z \to \infty$. Then:

$$\begin{gathered} V_3(\alpha,\ \beta) = -H_0\{z + c(\operatorname{ch}\alpha - \cos\beta)^{1/2} \sum_{n=0}^{\infty} [A_n e^{-(n+1/2)\alpha} + B_n e^{+(n+1/2)\alpha}]\, \mathrm{P}_n(\cos\beta)\}, \\ \alpha_2 \le \alpha \le \alpha_1\ , \end{gathered} \tag{8}$$

where instead of z one can substitute its value from (1). Using (7), taking into account the fact that V_1 and V_2 must be finite at the points $x = y = 0$, $z = \pm c$, in which the $\alpha \to \pm\infty$, we get

$$V_1(\alpha,\ \beta) = -H_0[z + c(\operatorname{ch}\alpha - \cos\ \beta)^{1/2} \sum_{n=0}^{\infty} C_n e^{-(n+1/2)\alpha}\, \mathrm{P}_n(\cos\beta)]\ , \quad \alpha \ge \alpha_1, \tag{9}$$

$$V_2(\alpha,\ \beta) = -H_0[z + c(\operatorname{ch}\alpha - \cos\beta)^{1/2} \sum_{n=0}^{\infty} D_n e^{(n+1/2)\alpha}\, \mathrm{P}_n(\cos\beta)]\ , \quad \alpha \le \alpha_2\ . \tag{10}$$

Now we impose two boundary conditions.

1. The first consists in the continuity of the potential on the surfaces of spheres, that is, for $\alpha = \alpha_1$ and $\alpha = \alpha_2$ the following equalities must hold: $V_1(\alpha_1,\ \beta) = V_3(\alpha_1,\ \beta)$ and $V_2(\alpha_2,\ \beta) = V_3(\alpha_2,\ \beta)$. Then:

$$C_n = A_n + B_n e^{(2n+1)\alpha_1}\ : \quad D_n = B_n + A_n e^{-(2n+1)\alpha_2}. \tag{11}$$

If the spheres are identical, i.e. $\alpha_1 = -\alpha_2 = \alpha_0$ (Fig. 1), then $A_n = -B_n$ and $C_n = -D_n$, since

$$C_n = -B_n(1 - e^{(2n+1)\alpha_0}): \quad D_n = B_n(1 - e^{(2n+1)\alpha_0}). \tag{12}$$

Then for points on the surfaces of spheres with allowance for (7), we can write:

$$V_1(\alpha_0, \ \beta) = -H_0[z + 2c(\mathrm{ch}\,\alpha_0 - \cos\ \beta)^{1/2}\sum_{n=0}^{\infty} B_n \mathrm{sh}(n+1/2)\alpha_0\ \mathrm{P}_n(\cos\ \beta)], \tag{13}$$

$$\begin{aligned} V_2(\alpha_0, \ \beta) &= -V_1(-\alpha_0, \ \beta) \\ &= H_0[z - 2c(\mathrm{ch}\,\alpha_0 - \cos\ \beta)^{1/2}\sum_{n=0}^{\infty} B_n \mathrm{sh}(n+1/2)\alpha_0\ P_n(\cos\ \beta)]. \end{aligned} \tag{14}$$

2. The second condition requires the continuity of the normal component of the magnetic induction, or, what is the same, the magnetic flux density, on the surfaces of the spheres:

$$\mu_i H_{n1}(\alpha_0, \ \beta) = \mu_e H_{n3}(\alpha_0, \ \beta), \text{ or } \mu \frac{\partial V_1}{\partial \alpha}|_{\alpha=\alpha_0} = \frac{\partial V_3}{\partial \alpha}|_{\alpha=\alpha_0}. \tag{15}$$

Using the dual-permutation property of the Maxwell equations and extending the results of the computations [11] to the case of identical spheres ($A_n = B_n$) for $\mu >> 1$, we obtain a three-term recursion relation

$$B_{n+1} f_{n+1}(\alpha_0, \ n+1, \ \mu) + B_n f_n(\alpha_0, \ n, \ \mu) + B_{n-1} f_{n-1}(\alpha_0, \ n-1, \ \mu) = g(\alpha_0, \ n, \ \mu) \tag{16}$$

for potential coefficients B_n, where $f_n(\alpha_0, n, \mu)$ and $g_{\mathrm{n}}(\alpha_0, n, \mu)$ are known functions determined by the parameters and arrangement of spheres.

In the case under study ($A_n = B_n$, $\mu \gg 1$), rather laborious calculations lead to the following expressions for the functions mentioned:

$$\left.\begin{aligned} & f_{n+1}(\alpha_0, n+1, \mu) = 2\mu(n+1)\mathrm{sh}[\alpha_0(n+3/2)], \\ & f_n(\alpha_0, n, \mu) = -4n\{\mu\{\mathrm{sh}[\alpha_0(n+1/2)]\mathrm{ch}\,\alpha_0 + \tfrac{e^{+\alpha_0(n-1/2)}}{4n}[1 - e^{-2\alpha_0(n-1)}]\} \\ & \ + \mathrm{ch}[\alpha_0(n+1/2)]\mathrm{ch}\,\alpha_0 + \tfrac{1}{2n}\mathrm{ch}[\alpha_0(n+3/2)]\}, \\ & f_{n-1}(\alpha_0, n-1, \mu) = 2n\,\mu\,\mathrm{sh}[\alpha_0(n-/2)], \\ & g_n(\alpha_0, n, \mu) = 2\sqrt{2}\mu[-ne^{-\alpha_0(n-1/2)} + (n+1)e^{-\alpha_0(n+3/2)}], \\ & n = 0, 2, 3\ldots \end{aligned}\right\} \tag{17}$$

Relation (15) is a system of non-homogeneous linear equations implicitly relating the coefficients B_n for three successive values of n. It is known that linear difference equations of the second order have analytical solutions only if their coefficients are polynomials or at least rational functions of the discrete variable n. Equation (15) is a second-order linear difference equation whose coefficients contain exponential functions of n and, consequently, its analytic solution is impossible. Various methods are realized to achieve the result, in particular, based on perturbation theory [9], and also the method of truncated recursion equations [10, 11], the Green's function method [12], the use of the asymptotics of the Legendre functions [13], the inverse decomposition method [14], and others.

The calculation process is very time-consuming and often requires additional calculations. Its results, i.e. the values of the potential coefficients are not always explicitly given, even when analyzing the system of ideally conducting spheres. Exceptions, in particular, are the works of [6.15]. According to the first, for the case of identical spheres we have:

$$B_n = 2^{1/2} c H_0 (2n+1)\{1+\exp[2(n+1/2)\alpha_0]\}/\{1-\exp[4(n+1/2)\alpha_0]\}. \quad (18)$$

The strength of the outer field into the system of the spheres is denoted by H_0 = const. From the property of the dual-permutability of Maxwell's equations, in particular, it follows that in the same system of spheres identical (up to a scale) distributions of the electric and magnetic potentials should be observed if sequentially we consider conducting spheres with $\sigma \to \infty$, dielectric spheres with $\varepsilon_i/\varepsilon_e \to \infty$ and spheres of magnetic material with $\mu = \mu_i/\mu_e \to \infty$.

It should be noted that as n grows, B_n decreases very rapidly (Fig. 2).

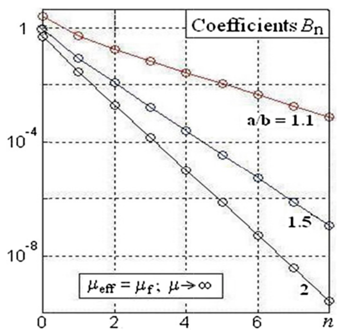


Fig. 2. The behavior of the coefficients B_n in (17).

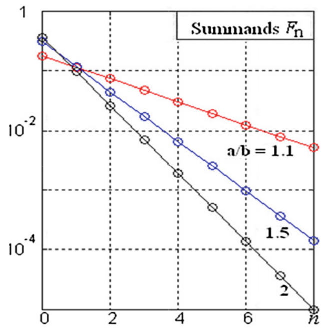


Fig. 3. Convergence of the series (18) for the potential.

We will assume that the expression (17) is applicable also in the analysis of a system of spheres with $\mu \gg 1$ and consider the potential distribution along the z axis in Fig. 1, using (8), where in the case of identical spheres it is necessary to set $A_n = -B_n$.

Then:

$$V_3(\alpha,\ \beta) = -H_0\{z + 2c(\mathrm{ch}\,\alpha - \cos\,\beta)^{1/2}\sum_{n=0}^{\infty} B_n\,\mathrm{sh}[(n+1/2)\alpha]\,\mathrm{P}_n(\cos\,\beta)\},\quad -\alpha_0 \le \alpha \le \alpha_0. \tag{19}$$

The Fig. 3 shows how the terms of the series (18) behave at the point (0, 0, a + b) lying on the intersection of the surface of the sphere with the z axis, where $\beta = 0$, for three values of α_0.

4 Effective Permeability of Two Spheres System

By effective permeability we will understand the ratio $\mu_{\mathrm{eff}} = \Phi/\Phi_0$, where

$$\Phi = \int_S \bar{B}_1 \mathrm{d}\bar{S},\quad \Phi_0 = \int_S \bar{B}_0 \mathrm{d}\bar{S}, \tag{20}$$

– fluxes of induction vectors found in the presence of spheres and without them, respectively. The ratio μ_{eff} characterizes the degree of concentration of the flux of the magnetic induction vector through a definite cross section S of the system.

To find the permeability μ_{eff}, it is necessary to calculate the values of the induction vectors $\bar{B}_1 = \mu_1 \bar{H}_1,\ \ \bar{B}_0 = \mu_0 \bar{H}_0$ in the central cross-section S (in Fig. 1 this is a circle with diameter $AA = 2b$ lying in the xy plane, where $\alpha = 0$) in the presence of spheres and without them, respectively.

The quantity H_1 is equal to the α-component of the field strength in the cross-section S. It can be easily determined in terms of the potential V_1. Applying relation (19) from [15] to the case of identical spheres, after the certain substitutions we find

$$H_{1\alpha} = -\frac{1}{h_\alpha}\frac{\partial V_1}{\partial \alpha} = -\sqrt{2}H_0(1-\cos\,\beta)^{3/2}\sum_{n=0}^{\infty}(2n+1)\frac{\exp[(2n+1)\alpha_0]+1}{\exp[(2n+1)\alpha_0]-1}\mathrm{P}_n(\cos\beta), \tag{21}$$

where h_α is the Lame coefficient [1]. The element of area in section S can be represented as $\mathrm{d}S = c^2 \sin\beta \mathrm{d}\beta \mathrm{d}\phi/(1-\cos\ \beta)^2$. The integration over φ in (19) reduces to multiplication by 2π. After the appropriate substitutions and rather voluminous calculations, we find the flux of induction

$$\Phi = 2\sqrt{2}\pi\mu\mu_0 c^2 H_0 \sum_{n=0}^{\infty} \mathrm{T}_n(n, \alpha_0, \beta_0), \tag{22}$$

where $\mu = \mu_i/\mu_0$ is the relative permeability of the material of the spheres, $\beta_0 = 2\mathrm{arctg}(c/b)$ is the value of the angle β at the points of the outer boundary of the S section, including at the points AA (Fig. 1). Here

$$\begin{aligned}\mathrm{T}_n(n,\alpha_0,\beta_0) = &-(2n+1)\tfrac{\exp[(2n+1)\alpha_0]+1}{\exp[(2n+1)\alpha_0]-1}\int_{\beta_0}^{\pi}\tfrac{\sin\beta P_n(\cos\beta)\mathrm{d}\beta}{(1-\cos\beta)^{1/2}} \\ &= \mathrm{D}_n(n,\alpha_0)\int_{-1}^{X_0}\tfrac{\mathrm{P}_n(X)\mathrm{d}X}{(1-X)^{1/2}},\end{aligned} \tag{23}$$

where $\mathrm{D}_n(n,\alpha_0) = (2n+1)\{\exp[(2n+1)\alpha_0]+1\}/\{\exp[(2n+1)\alpha_0]-1\}$, $X = \cos\beta$, $X_0 = \arccos\beta_0$.

The flux in the absence of spheres is equal to the circle area with radius b, i.e. $\Phi_0 = \pi b^2\mu_0 H_0$.

Thus, the μ_{eff} of a system of two identical spheres is described by an infinite series

$$\mu_{\mathrm{eff}} = \frac{\Phi}{\Phi_0} = 4\mu\frac{c^2}{b^2}\sum_{n=0}^{\infty}\mathrm{T}_n(n,\beta_0). \tag{24}$$

The rate of convergence of the series in the expressions given above, in general, depends on the geometric parameters of the system of spheres. Therefore, in some cases, in particular, for small values of the gaps between the spheres, when $c/b\to 1$, an additional study should be carried out to determine the reasonable number of terms in the corresponding sums necessary to achieve the desired accuracy of the calculations.

Effective permeability can be a convenient indicator of the degree of field distortion when a separate body or system of bodies is introduced into this field. It makes it possible to estimate the gain in the magnitude of the flux of magnetic induction through the central section of the system realized for the given geometry and properties of the body (or bodies) material. On the other hand, it is possible to judge from it the degree of use of these properties.

In this paper, the value of μ_{eff} is obtained for the case $\mu \gg 1$. Therefore, it can be assumed that the value (23) can be considered as an estimate of the permeability of the form

$$\mu_{\mathrm{f}} = \lim_{\mu\to\infty}\mu_{\mathrm{eff}}, \tag{25}$$

for a system of two identical spheres. The permeability of the form is the maximum achievable value of the μ_{eff} for a given geometry of the system. Having the dependence $\mu_{\mathrm{eff}}(\mu)$ for given proportions of the system geometry, it is easy to find such a value of the material permeability, the further increase of which will not lead to an increase in the value of μ_{eff}.

From the physical point of view, the limit transition $\mu\to\propto$ is equivalent to the transformation of the material of spheres into an ideal magnetic, on the surface of which tangential constituent fields, i.e. H_φ and H_β, become equal to zero. Thus, the structure of the field in the outer region will be determined by the single component $H_\alpha = -\mathrm{grad}V_1$, which is normal to the surfaces of the spheres. Remembering the principle of the dual permutability of Maxwell's equations, the illustration of field structures in a two spheres system for some particular cases can be found in [11].

5 Some Applications of Research Results

1. Consider a receiving ferrite antenna, in the simplest case containing a cylindrical core of material with high μ and low losses and a solenoidal winding placed on it [16]. When the antenna circuit is tuned into resonance, the voltage $U_A = e_A Q_A$ acts on the receiver input, where Q_A is the quality factor of the antenna circuit, and e_A is the EMF induced in the antenna by the received wave. It is known that $Q_A = W_A/P_A$, i.e. it is the ratio of the energy stored in the system to the energy lost in it in one oscillation period.

It has been experimentally established that, in the presence of a gap in the core, the quality factor increases. A qualitative explanation of this fact is the increase in the reactive energy stored in the system, while the losses are practically unchanged. The mathematical justification for this in the known literature is missing. The relationships obtained in the previous sections of the article allow for quantitative calculations and simulate the influence of the gap on the characteristics of the antenna. However, this research is beyond the scope of this paper.

2. If we increase the radius of the circle AA (Fig. 1), its area S grows rapidly, but the value of the tension H_z, averaged over this area, decreases. Therefore, we can assume that the magnetic flux through this circle at a certain value of its radius reaches a maximum.

However, detailed studies of these issues are beyond the scope of this paper.

6 Conclusions

The problem of the incidence of a plane electromagnetic wave on a system of two spheres having electrically small dimensions can be regarded as quasistationary. Scalar potentials obtained as a result of solving the Laplace equation are represented by series containing Legendre polynomials. Estimates show that for $\mu \gg 1$ the convergence of these series depends on the geometric dimensions of the system. In practical calculations, this may lead to the need for additional studies in order to determine the required number of terms in these series.

The concept of effective permeability of a two-spheres system is introduced. It is equal to the gain in the magnitude of the magnetic induction vector flux through a certain cross section of the system, which arises from its magnetic properties. The necessary ratios for effective permeability, referred to the central section of the system, are obtained. The results can be used, in particular, to quantitative description of the influence a gap in cylindrical ferrite cores of a frame antennas.

References

1. Morse, P.M., Feshbach, H.: Methods of Theoretical Physics, Part 2, 1298 p. McGraw-Hill, New York (1953). (Russian translation 1960)
2. Makarov, A.M., Luneva, L.A., Makarov, K.A.: To an estimate of the error of the Coulomb experiment. Bulletin of the N.E. Bauman's Moscow State Techn. Univ. Ser. "Natural Sciences" **1**, 88–95 (2007). (in Russian)
3. Hall, W.D., Beard, K.V.: Electrostatic force between a conducting sphere and a dielectric sphere. Pageoph **113**, 515–524 (1975)
4. Grover, S.N.: The electrostatic force between a conducting sphere and a charged dielectric sphere in arbitrarily oriented external electric field. Pageoph **114**, 521–539 (1976)
5. Ciric, I.R., Kotuwage Kumara, S.C.M.: Benchmark solutions for magnetic fields in the presence of two superconducting spheres. Mater. Sci. Forum **721**, 21–26 (2012)
6. Nainani, A., Roy, A., Singh, P.K., Mukhopadhyay, G., Vasi, J.: Electrostatics and its effect on spatial distribution of tunnel current in metal Nanocrystal flash memories. In: ICMTD-2007, pp. 251–254 (2007). http://dspace.library.iitb.ac.in/xmlui/bitstream/handle/10054/611/ICMTD_2007_p251-4.pdf?-sequence=3
7. Khachatourian, A., Chan, H.-K., Stace, A.J., Bichoutskaia, E.: Electrostatic force between a charged sphere and a planar surface: a general solution for dielectric materials. J. Chem. Phys. **140**, 074107-1–074107-10 (2014)
8. Aldridge, D.F., Oldenburdg, W.: Direct current electric potential field associated with two spherical conductors in a whole-space. Geophys. Prospect. **37**, 311–330 (1989)
9. Goyette, A., Navon, A.: Two dielectric spheres in an electric field. Phys. Rev. B **13**(10), 4320–4327 (1976)
10. Large, D.B.: Electric potential near a spherical body in a conducting half-space. Geophys. **36** (4), 763–767 (1971)
11. Stoy, R.D.: Solution procedure for the Laplace equation in bispherical coordinates for two spheres in a uniform external field: parallel orientation. J. Appl. Phys. **65**, 2611–2615 (1989)
12. Chaumet, P.C., Dufour, J.P.: Electric potential and field between two different spheres. J. Electrostat. **43**, 145–159 (1998)
13. Love, J.D.: The dielectric ring in a uniform, axial, electrostatic field. J. Math. Phys. **13**(9), 1293–1301 (1972)
14. Nakajima, Y., Sato, T.: Calculation of electrostatic force between two charged dielectric spheres by the re-expansion method. J. Electrostat. **45**, 213–226 (1999)
15. Koromyslov, V.A., Grigor'ev A.I.: On the polarization interaction between two closely spaced conducting spheres in a uniform electrostatic field. Techn. Phys. **47**(10), 1214–1218 (2002)
16. Krasnitsky, Y.A.: Electrically Small Frame and Ferrite Antennas, 140 p. TSI, Riga (2004). (in Russian)

Model of Wireless Data Network in GPSS Language

Aleksandr Krivchenkov(✉)

Transportation and Telecommunication Institute, Lomonosova str. 1, Riga 1019, Latvia
krivcenkovs.a@tsi.lv

Abstract. A simulation model of data transfer network when wireless technology is used was developed on the basis of simulation environment GPSS. A number of parameters are defined in the model for simulation network architecture, characteristics of traffic, wireless technology, characteristics of the radio channel. Examples of wireless technology 802.11 were used to get in simulation parameters as throughput, bandwidth, delays of packets, and distribution function for such delays, jitter and number of packets sent with errors. Comparing results of analytical and simulation models within a certain range of settings showed a good agreement (2%). The program text of the simulation model is available online.

Keywords: Wireless network simulation · IEEE 802.11 · GPSS model

1 Introduction

There are different wireless technologies that are used in data transmission networks. Some of them are based on standards IEEE 802.11 (Wi-Fi), 802.15.1 (Bluetooth), 802.15.4 (ZigBee). In networks design for many applications, it is very important to have correct knowledge about main relationships between network and traffic architecture, conditions of transmission and parameters of data transfer. Such research is difficult to perform in field experiments and analytical and simulation models are often used. Models for wireless networks are also very useful for educational purposes as they can demonstrate the influence of basic network mechanisms.

The informative published models [1–3] based on MATLAB® Simulink® environment simulate Physical and partially Data Link levels of the wireless channel. It is not sufficient for understanding the network processes. Here we need to apply models based on Queueing theory. Such models with necessary details are not widely represented in various simulation environments. For an example in [4], simulation model based on GPSS (General Purpose Simulation System) is used for analysing the general mechanisms of 802.11 wireless networks. Unfortunately, the program text of the model is not published.

In our consideration, data transmission network simulation model is discussed to determine the effects of different factors on the characteristics of wireless networks. We try to take into account the details of channel access mechanism and data rate adaptation on the Physical level. Often these factors are difficult to take into account in analytical

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_42

calculations. In published model [5] the basic wireless mechanisms for 802.11 are simulated. There is also an opportunity to analyse the dynamics of channel parameters when interacting objects are moving.

A model was developed for GPSS environment that is widely used for computer systems simulation and educational purposes. Some results of the simulations are published in [6]. In our model can be found: channel throughput, delays in packets transfer, the distribution function of delays, jitter, and the number of packets sent with errors in different conditions. In this paper, the main parameters of the model are described step by step. Comparing results of analytical and simulation models within a certain range of settings showed a good agreement (2%).

2 Network Simulation Model

2.1 Network and Traffic Architecture, Size of Frames

The set of communicating nodes and possible connections between them form the architecture of the network. The basic architectures of wireless technologies like 802.11, 802.15.1, 802.15.4 have common features (Fig. 1).

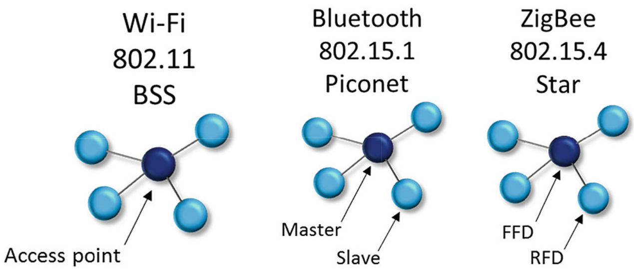


Fig. 1. Basic architectures for 802.11, 802.15.1, 802.15.4.

For those architectures, nodes can communicate with each other or with wired part of the network through special node as an Access point (802.11) or Master node (802.15.1). In our model, the main parameter that defines the architecture of network (and traffic in it) is the number of nodes (`Node_Count`).

In the model general (from all nodes) traffic is defined. Then it is distributed across all nodes with equal probability. The variable `Node_Select` randomly determines the number of the node that will be the next source of the message. So, each node is the source of traffic, the rate of this application traffic is set in “Message stream” section. For an example for application traffic 96 Kbps bitrate, if an application creates every millisecond message of 12 bytes length, bitrate = 8·12 bit/1 ms = 96 Kbps.

Messages generated in nodes (they are called application level messages) are coming on Transport level; it means that they are encapsulated in transport level packets. We do not divide traffic due to applications profile. Our profile rather Real-time application for what delays and jitter of packets are very important.

Messages can have two different sizes: `dl_Min` and `dl_Max`. The part of short messages is determined by the `Fraction_Short_Msgs` parameter. When generating a message its size (long or short) is determined randomly using variable `Msglength`:

```
Msglength VARIABLE
dl_Min+(RN2'G'Fraction_Short_Msgs)#(dl_Max-dl_Min)
```

The process of packets transfer through one of the wireless channels is simulated by occupation and subsequent processing time on a device named `Radio_channel`. This is the way for simulation the competition for common radio resource.

Table 1 shows the values of the parameter `Node_Count` for some architectures and network traffic (symbol «→» refers to the wireless communication channel, and "/"-wired).

Table 1. Parameter `Node_Count`, architecture and network traffic.

Architecture	Traffic	Node_Count
802.11 BSS	Host1 → AP/Host2	1
802.11 BSS	Host1 → AP → Host2	2
802.11 BSS	Host1 → AP → Host2, Host2 → AP → Host1	4

Application messages generation is performed in the block `GENERATE`. The distribution function of time intervals between packets is selectable.

The average time interval between messages at the application level:

```
*    Average interval between messages
; Gives the Bitrate from every Node in msec
Intermessage_Time FVARIABLE
8#(dl_Min#Fraction_Short_Msgs+dl_Max#(1000-
Fraction_Short_Msgs))/1000/Bitrate/Node_Count
```

In the simulation model it is assumed that at the Transport and Network levels protocols UDP (User Datagram Protocol) and IP (Internet Protocol) are used:

```
** Transport and Network level
dl_UDP    EQU 8 ; UDP transport protocol header in bytes
dl_IP     EQU 20; IP network protocol header in bytes
Pack_size  VARIABLE (dl_UDP+dl_IP+P$Msg_App_length); Pack
size coming on Link level in bytes
```

So, `Pack_size` is the variable that defines packet size coming on Link level. In the model, the situation when transport protocol TCP (Transmission Control Protocol) is used is not realized. It is so because we are focused on Real-time application profile. If TCP is used on transport level it leads not only to the changes in frame size but also to significant changes in transmission algorithm. And some parameters of transmission (for an example message delay) may be changed also significantly with respect to the results of our model.

The next overheard of frame size is defined by MAC (Media Access Control) frame structure for appropriate wireless technology. For 802.11 technology we have for `Frame_size` variable:

```
dl_LINK EQU 36 ;802.11 Frame overhard in bytes
****** Calculation of frame size on MAC level
Frame_size VARIABLE (dl_LINK+V$Pack_size) ;Frame size in
bytes
```

2.2 Frame Transmission Parameters

The transmission time of the frame is determined by the amount of time required for: access to the channel; transmission the frame; receive the acknowledgement frame (ACK). For 802.11 standard basic transmission time intervals are displayed in Fig. 2 where DCF (Distributed Control Function) for access to the channel is used.

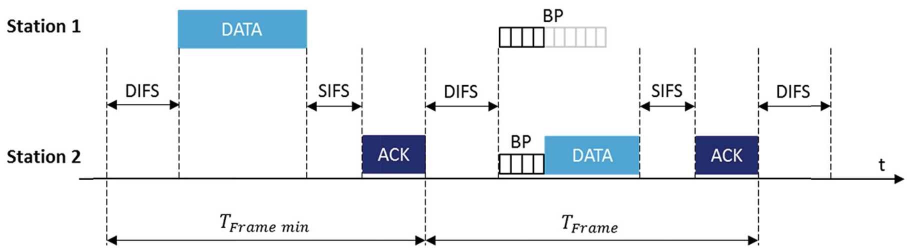


Fig. 2. Basic mechanism of frames transmission for 802.11 DCF.

For 802.11 technology frame transfer time in the simulation model is determined by the sum of the minimum `Msgtime` and some possible interval of delay (BP – Backoff Period). In next section the minimum (no BP interval) message transmission time is calculated:

```
;data transfer time in micro sec
Data_Frame_time  FVARIABLE Pream-
ble+(8#V$Frame_size+6)/V$Raw_rate+6
Msg_transf_time    VARIABLE
(V$DIFS+V$Data_Frame_time+SIFS+V$ACK_time)
*** Calculation of Frame transfer time
;data transfer time in micro sec
Data_Frame_time    FVARIABLE Pream-
ble+(8#V$Frame_size+6)/V$Raw_rate+6
* ** Transfering time due to Transactions model
Msgtime    VARIABLE
0.001#(V$DIFS+V$Data_Frame_time+V$SIFS+V$ACK_time)
```

Calculations for 802.11 depend on the appropriate specification: b,a,g,n,ac. Bitrate at the Physical level `Raw_rate` is determined by the bit rate index (a specific set of bitrates have all specifications).

The values of bitrates are set in `R_a_b matrix`. In the published text of model bitrates for 802.11a or 802.11g specifications are defined.

Additional delay time interval in process of access to the channel (BP – Backoff Period) appears only after the first unsuccessful attempt to transmit a frame. So, in the channel without concurrent conflicts (collisions) and distortion of the frame due to noise BP does not arise. In this case, frame transfer time will be equal to value `Msgtime`.

2.3 Simulation of Retransmissions in Channel

Retransmission of frames can occur either as a result of collisions in the channel, either due to errors in bits when transmitting on the no ideal (or noisy) channel. In the simulation model the probability of erroneous transfer defines a variable:

```
Error_Probability VARIABLE 1-(1-FN$Collis_prob)#(1-
V$Noise_prob)
```

The erroneous transmission of a frame that occurs due to the collisions (in the preceding text `FN$Collis_prob`) is determined by multiple access to the shared channel. We have taken into account substantive review [7], in which the analysis of different models of MAC-level algorithms was presented. It was shown that for a well-known model Bianchi which gives a numerical solution (in the state of saturation of the system), the probability of collisions is dependent only on the number of active stations (Fig. 3). Just this function is specified in the model.

In 802.11 technology frames retransmissions are accompanied by an increase of ‘competition window’, and it leads in turn to an increase in the transmission time of the frame. In the simulation model competition window defines a variable CW and is calculated as $CW_k = (CW_{min} + 1)2^k - 1$, where k is the number of retransmissions (in model variable `Retries`), and CW_{min} – the initial value of the window (variable `CW_min`).

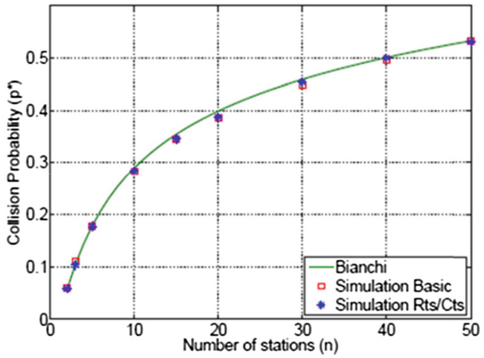


Fig. 3. The probability of collisions as a function of the number of stations: simulation and model Bianchi. Simulation results match the rate of 11 Mbps; others rates give the same probabilities.

Increase in this window leads to an increase in the transmission time of the frame.

The maximum number of retransmissions is not regulated in 802.11, but it may be of the order of 6–7, in the model it is defined by `Backoff_Limit`. If variable `Retries` exceeds the value of `Backoff_Limit` retransmissions are stopped and the error in the frame is fixed. The number of errors in frames is fixed into variable `Error_Count`. Specified parameters are collected in section of program code “Parameters for DCF function backoff for access to channel”.

2.4 Simulation and Parameters of SNR Influence on Errors in Channel

The probability of erroneous transmission of frames and frame retransmissions further can occur either as a result of collisions, either due to errors in. In such channel noise and interference are presented. Variable in the model `Error_Probability` takes also into account the probability of errors due to noise in variable `Noise_prob`:

```
Noise_prob VARIABLE V$PER
BER FVARIABLE 1/2#Exp(-Exp((V$SNR-
MX$R_a_b(P$R_index,3))/MX$R_a_b(P$R_index,2)))
PER FVARIABLE 1-(1-V$BER)^(8#V$Frame_size)
```

Thus, the probability of errors due to noise is considered equal to `PER` (Packet Error), which in turn is determined by the probability of errors when transferring bits BER and the total size of the frame in bits:

$$PER = 1 - (1 - BER)^{8 \cdot l}, \tag{1}$$

where the *BER* is determined by some function of *SNR* (Signal to Noise Ratio) and some parameters that depend on the Physical layer technology. The next section discusses the method of defining *BER* function.

We assume that in the channel operates Additive White Gaussian Noise (AWGN). As a quantitative characteristic of the impact of noise on the signal, the signal/noise ratio (*SNR*) is used.

Due to 802.11 device specifications characteristics of noise influence on error probability are used. In general, *BER* (and *PER*) can be expressed through *SNR* using some nonlinear function, which takes into account the type of modulation and signal encoding method, and in the case of *PER*-frame length as in (1).

From the analysis conducted in [4], direct computation method of *BER* is inefficient in use. In this regard, the method of exponential-logarithmic approximation is proposed, in which the bit error probability *BER* is estimated by the formula:

$$BER^{(\mathrm{i})}(SNR) = \frac{1}{2}\exp\left(-exp\left(\frac{SNR - \beta^{(i)}}{\alpha^{(i)}}\right)\right), \tag{2}$$

where the index i enumerates bitrate at the PHY level, $\alpha^{(i)}$, $\beta^{(i)}$ – variation parameters that depend also only on the bitrate.

We apply approximation method to 802.11a. Specification 802.11a regulates the work of Physical layer on eight speeds (6, 9, 12, 18, 24, 36, 48 and 54 Mbps), each of which is determined by one of the four schemes of modulation and one of three convolutional coders.

To calculate the parameters that are presented in the expression (2) a simulation of transmitter and receiver 802.11a, was used. Model for it [1] supports all speeds, regulated by the specification, and allows adjusting the SNR and varying the size of the transmitted frames to receive probability of bit distortion. In the model, the probability of bit BER is calculated.

Obtained in simulation datasets allow defining variation parameters for (2). Method of determining parameters is a non-linear regression. Obtained variation parameters are presented in Table 2 and are saved in the model text in section “MATRIX Bit rate R, a, b koef. for *BER* calculation”.

Table 2. Variation parameters for 802.11a defining the *BER*.

Bitrate [Mbps]	α	β
6	3.4952	–5.7829
9	3.7941	–3.8729
12	4.2123	–3.085
18	4.4505	–0.887
24	4.3979	2.2111
36	4.8863	5.05
48	4.3856	9.8807
54	4.8355	10.479

Substituting the variation parameters in expression (2) we get dependence *BER* from SNR for the entire set of bitrates. These dependencies are illustrated in Fig. 4. The resulting dependences *BER* from *SNR* and *PER* from *BER* are used in our simulation model to calculate the probability of error in frame transmission.

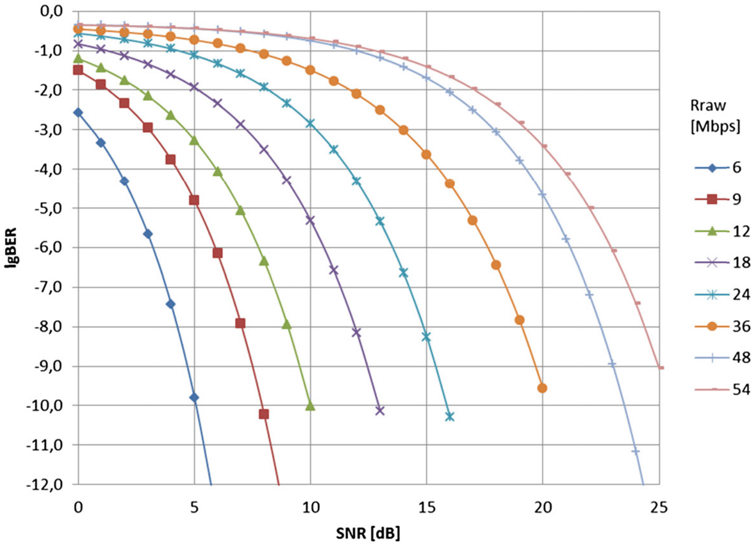


Fig. 4. Dependence of *BER* from *SNR* for various bitrates and 802.11a specification.

Taking advantage of the proposed methods, as well as published models [2] for 802.11g or [3] for 802.11n, one can obtain the appropriate dependencies for these technologies.

2.5 Simulation of Communicating Objects Mobility

In our simulation model assumption is done that changing the distance between communicating objects changes *SNR* in the communication channel. The next section identifies a few possible options for changing *SNR*:

```
*     Different types of SNR time dependence
;steady SNR in channel
Vsnr0   FVARIABLE 25 ;SNR equ 25 dB
;random distribution of SNR in range 0 - SNR_max
SNR_max   EQU 25         ;SNR_max EQU 55;Max SNR, example
0dBm-40dB-(-95dBm)=55dB
Vsnr1   FVARIABLE (0+(RN3@SNR_max))
* SNR is a function of time
SNR_ini   EQU 25
SNR_min   EQU 5
V_speed   EQU 10 ;Speed 10 m/sec
alpha   EQU 2 ;Gradient in free space
;receiver moves away from transmitter
Vsnr2 FVARIABLE (SNR_ini-
alpha#10/2.3026#Log(1+V_speed#AC1/1000))
```

For each node, the function of *SNR* time dependence is chosen in accordance with the value (number in column 4) in the matrix `Node_Rraw_s_e`.

For an example for Node 1the SNR function will be `Vsnr2` (receiver moves away from the transmitter) and for Node 2 – `Vsnr1` (random distribution of SNR in range 0 – `SNR_max`). In previous calculations we have taken into account that loss of signal power *L* (loss) at a distance of *d* is determined by the next relationship:

$$L = L_0 + \alpha \cdot 10 \log_{10} d, \tag{3}$$

where L_0 – loss in decibels on the first unit of length (in the first meter, if the distance d is measured in meters); α – gradient losses (`alpha` in the model). For losses in free space α = 2.

3 Examples of Simulation Experiments for 802.11

Simulation for 802.11 is held for multiple network structures. For an example in the first case, the model parameter `Node_Count` was equal to 2 (a device sends data to another via an access point). In the second case, the parameter `Node_Count` was equal to 4, which corresponds to four communication links (e.g., the same two devices, but the data are transmitted in both directions). The results of the simulations show that the variance of the results obtained with analytical models is approximately 2% when measuring bandwidth. The received results allow concluding a good coincidence of simulation and analytical models results in certain parameters ranges.

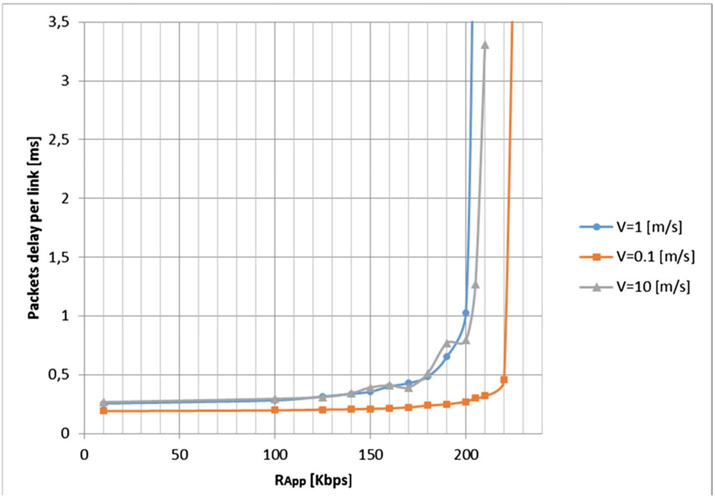


Fig. 5. Simulation results for two links when changing *SNR* in one of them due to the movement of a node.

In next series of simulations [6] estimate the impact of mobility of objects was done. In Fig. 5 graph shows data for the simulation of the interaction between 2 Nodes, one of which is moving relative to another with a speed v.

It was found that the influence of the DCF on bandwidth with a small number of Nodes (just 2) competing for the channel is negligible. Bandwidth varies significantly with variations in the level of *SNR* in the channel. In Fig. 5 bandwidth is reduced by 10% when *SNR* changes in the channel due to the movement. For fast movement (10 m/s) bandwidth remains slightly higher compared to the case when driving at an average speed (1 m/s). This feature is determined by the action of mechanism ARF (Auto Rate Fallback), which reduces the bitrate in the channel as *SNR* is reduced (when an object is moving from the access point). It was shown that parameters of the ARF mechanism set in the model significantly affect simulation results.

Examples demonstrate possibilities of the published model to simulate various mechanisms for network layers when wireless technologies are used.

4 Conclusions

Developed simulation model allows estimating the many factors effects on the characteristics of the data transmission network. Comparing the results of analytical and simulation models within a certain range of settings it is shown of the order of 2% agreement.

For a more in-depth study of *SNR* changes influence on the probability of errors was done. This made it possible to simulate communications between moving objects. In model functional dependence of *SNR* in communication distance between nodes is introduced in accordance with a theoretical model for path loss of radio waves.

Probabilities of *BER* and *PER* are evaluated on the basis of the application of Simulink models for relevant Physical layer specification.

Using presented model it is found that the impact of the DCF mechanism on the bandwidth with a small number of competing Nodes is negligible.

In the model, mechanism ARF (basic for bitrate changes in 802.11 objects) is realized and it is shown that this option significantly affects simulation results of bandwidth. Among other things published model demonstrates the significant dependence of bandwidth from the level of *SNR* in the channel. For moving objects in some simulations the bandwidth is reduced by 10% when objects are moving.

The set of experiments were performed to validate the model [6]. It is shown that basic effects are taken into account with accuracy about 5%.

References

1. Clark, M.: IEEE 802.11a WLAN model version 1.0.0.1 (2016). http://se.mathworks.com/atlabcentral/fileexchange/3540-ieee-802-11a-wlan-model
2. Saxena, A.: 802.11g WLAN PHY model version 1.3 (2012). http://se.mathworks.com/matlabcentral/fileexchange/38073-802-11g-wlan-phy-model?s_tid=srchtitle
3. Ogunfunmi, T.: IEEE 802.11n PHY Layer model version 1.0 (2008). http://se.mathworks.com/matlabcentral/fileexchange/22137-simulink-model-of-the-ieee-802-11n-phy-layer-model

4. Yakimov, M.: Thesis for the degree of candidate of technical sciences: analysis of mechanisms for increasing the efficiency of information transfer in high-speed local and urban wireless networks. The Russian Academy of Sciences. Institute for Information Transmission Problems. A.A. Kharkevich (2008). (in Russian)
5. Krivchenkov, A.: Wireless network GPSS model (2017). https://docs.google.com/document/d/1maIdb9ooLXAV9yis6gRCM4YH_jhAeJyEuIQ2xj75J3A/edit?usp=sharing
6. Krainyukov, A., Krivchenkov, A., Saltanovs, R.: Performance analysis of wireless communications for V2G applications using WPT technology in energy transfer. Procedia Eng. **178**, 172–181 (2017). http://www.sciencedirect.com/science/article/pii/S1877705817300851
7. Larroca, F., Rodríguez, F.: An overview of WLAN performance, some important case-scenarios and their associated models. Wirel. Pers. Commun. **77**, 131–184 (2014). https://iie.fing.edu.uy/personal/flarroca/publications/

Identification, Classification, Implementation and Cryptography Problems of Various Complex Systems

Automatic Gender and Emotion Recognition System as Important Factor for Safety Improvement

Jacek Mazurkiewicz(✉)

Department of Computer Engineering, Faculty of Electronics, Wrocław University of Science and Technology, Wybrzeze Wyspianskiego St. 27, 50-370 Wrocław, Poland
Jacek.Mazurkiewicz@pwr.edu.pl

Abstract. The paper presents the analysis and discussion of gender and emotion recognition based on human face picture. The research combines different features selection techniques with the set of softcomputing classifiers. The goal is: not very complicated, fast and sensitive approach to create the basis for safety systems with correct "on-line" gender and emotion recognition. The already known differences between the female and male face are the starting point for discussion. The second path is focused on already known for physiologists emotional states visible in human face. The classic classifiers are in use, but focus is on sensible correlation between the feature extraction and the actual classification. The significant set of the results are discussed and the best solutions are pointed. All tests were realized based on the mixture: well known base of face pictures and the set of own collection. The proposed solution can be an essential tool for the monitoring systems, safety guards and systems to point the dangerous situations based on video data.

Keywords: Face · Gender and emotion recognition · Softcomputing · Safety

1 Introduction

People are trying to automate many processes that used to be done by humans. Some of them are easy to implement, like a mechanical production line. There is a known set of steps resulting in making a certain good like a book or a chair. All of those work used to be done by humans and the only thing that changed was adapting those trivial mechanical tasks into software controlling production line. However things start to get complicated when it comes mimicking the way human brain works. Especially taking into account the fact that the actual way the most important organ in our body works is still not fully discovered. There are areas of neurology that are still to be explored.

Sometimes what seems to be an obvious and trivial task even for a 10-years old is not easy to achieve using computers. Traditional algorithms become too complicated or even impossible to be applied in certain applications. This is where a different approach can be made useful instead of conventional hard computing one. Fortunately since the early 90's a thing called soft computing started to be developed. What differs it from conventional

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_43

approach is the fact that it tolerates approximation as well as imprecision and uncertainty. In fact the main intent while creating this area of computer science was to imitate human brain to a possible extent [11].

One of those tasks in which soft computing comes suitable is gender and emotion state recognition. Each human individuals are different from each other, same applies to their faces. The way certain person shows emotions is unique which leads to a need for some approximation. The purpose of this paper was to create a software able to mimic human ability to detect emotion and gender based on a picture of a person's face. This kind of application can be useful in many areas starting from smile detection in cameras up to artificial intelligence implemented in humanoid robots that will accompany people's lives in not so distant future.

2 Emotion and Gender – Areas of Face

In order to start detecting emotional condition of person in the picture one needs to know what face features need to be extracted for further processing and what are emotions one is looking for (Fig. 1). Paul Ekman, professor of psychiatry from the University of California has done research on human emotions and facial expressions connected to them. According to his analysis there are 6 main expressions and a case where no emotion is shown which for the purpose of this thesis will be related to as a neutral emotion [4]. Therefore the list of all 7 emotions goes: happiness, sadness. Fear, disgust, surprise, anger, neutral. Also there are a lot of gender features on the face, such as [1–3]: size of the skull – male skull is bigger, slope of the forehead – male forehead is more sloped, shape of the eyebrows – female eyebrows are more round, and male eyebrows are bigger, depth of the eye socket – male eye sockets are much deeper, size of the jaw – male jaw is bigger, spacing between the eyes – female eyes are placed wider, size of the eyes – female eyes are bigger, size of the nose – male nose is bigger, size and color of the mouth – female lips are more red and fuller, colour of the skin – male are more red and female are more green, but this can be considerate only without any make-up and in perfectly equal light conditions.

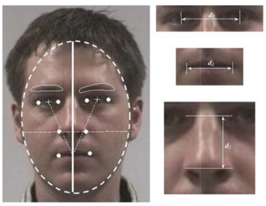


Fig. 1. Sources of gender and emotional features on the face [3].

System assumes that there are some existing algorithms to get the face from the picture without manually cropping of the photo. Because of the complexity of the problem there had to be put certain restrictions on people in the pictures, such as: Caucasians, age from 20 to 50, no visible diseases on the face, photo taken from front with ears visible, no plastic surgery – like the diseases it can affect gender features.

3 Gender Recognition

3.1 Proposed Approach

First step – taking the picture – is done manually, faces are cropped by hand and save as the pictures 133×169 px and 50×64 px. Eyes on the cropped photos should be on the same level. This part should be done by another system which would automates human faces extraction from camera image. Second step - transforming the picture – extracts the gender features from face. There are two different algorithms in the system which can do that: 2 color conversion, 8 color conversion. All of the previous works used complicated algorithm for gender features extraction and still did not achieved 100% accuracy, because of that we decided to use the easiest way of extraction and find out if it can provide the same or better accuracy. Also the eight colors conversion was only mentioned as possible improvement of two colors conversion so in the paper we wanted to check correctness of that statement. Algorithms was chosen also because of theirs low computational time and small memory consumption. Third step – calculating the results – is actual classifier. There are three different algorithms in the system which can do that: Multilayer Perceptron (MLP), k-Nearest Neighbors (kNN), Support Vector Machine (SVM). MLP is well known neural network frequently used in gender classification [9]. Can be easy tuned to perform better results by changing number of hidden layer and neurons in them. Multilayer Perceptron Classifier was not tested before with chosen gender features extraction algorithms. Also eight colour conversion algorithms was only mentioned as possible improvement of two colour conversion in [10] and have never been tested with any mention classifier. kNN is memory based reasoning algorithm. This have slightly different approach to classification than the approach in neural networks. It also has the best results of all known classifiers which manage gender classification [10]. k-Nearest Neighbors Classifier was only tested until now with Euclidean distance [10], in this thesis also eight another distances was tested. SVM is algorithm which mimics best the way in which human beings classify gender of other humans. Because of kernel trick this can also classify not-linear separable data [13]. It also has the best results [9] from all tested classifier on FERET database from which half of the learning and testing set was taken for tests provided for this paper. Fourth step – displaying the results – is simple show of the persons gender. Male is coded as 0 and female as 1.

3.2 Experiment Outline

Every classification algorithm was tested with every feature extraction algorithm attached. Classification algorithms was fine tuned during the tests to find parameters that will suit the best to presented problem. All tests were conducted on total number of 51 photos taken from FERET database [13] and 49 pictures taken by ourselves, there were 45 female and 55 male. For fine tuning the algorithms for every parameters set there were conducted three test: 75 photos in learning set (33 female and 42 male) and 25 photos in testing set, 50 photos in learning set (28 female and 22 male) and 25 photos in testing set, 25 photos in learning set (13 female and 12 male) and 25 photos in testing set. After finding the best parameters classifiers was learned with learning set which grew from 25 to 75 photos with 5 picture threshold. Every person was represented by one photo. All tests were conducted with 25 pictures in testing set with 12 female and 13 male.

3.3 Gender – Results

Simulation results are available on Figs. 2, 3 and 4.

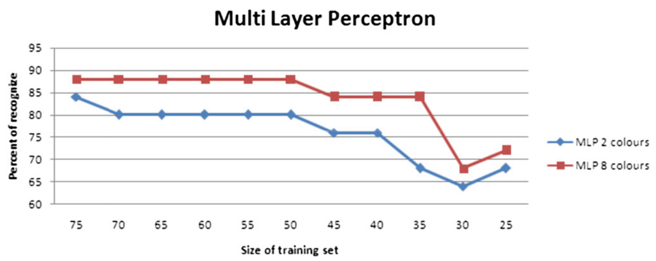


Fig. 2. Multilayer Perceptron best results.

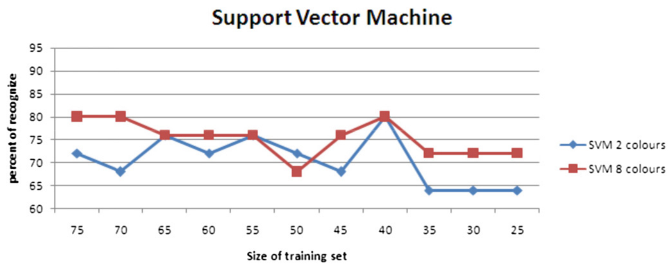


Fig. 3. Support Vector Machine best results.

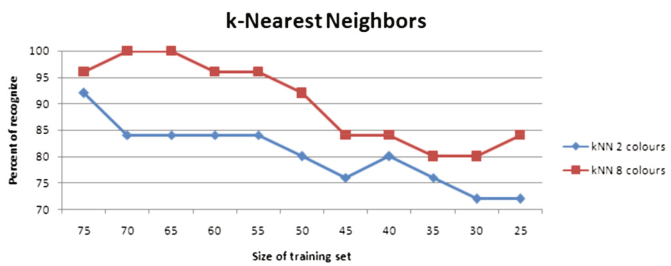


Fig. 4. k-Nearest Neighbours best results.

Algorithm for MLP is implemented with backpropagation and logistic activation function. This activation function was chosen because its answer vary between 0 an 1 and expected results are in set {0.1}. Expected value of error after learning is set to

0.01, below that we can observe overfitting. There are 3 layers: input with 3 200 neurons (size of input images in pixels), hidden with changing number of 100 and 200 neurons during tests – more neurons in hidden layer did not provide better results and computational time for learning started to be unacceptable. Output with 1 neuron (gender are coded as 1 bit). Learning rate vary from 0.05 to 0.5 with 0.05 threshold. Learning was conducted till mean square error reach 0.01 or number of iteration reach 20 000. SVM-learning was conducted with Sequential Minimal Optimization method of finding separating hyperplane. Classifier was tested with 5 different kernel functions provided by MATLAB build-in function svmtrain: Linear, Quadratic, Polynomial, Gaussian RBF, MLP. What was surprising RBF kernel did not provides good results (as it was in Shaknarovich, Viola and Moghaddam work [13]), in fact gave the worst accuracy of classification from all tested kernels. To tune the kNN classifier number of neighbors varied from 1 to 15 (testing set contain only 25 photos, more neighbors would not provide any improvement), and test was conducted for nine different distances which can be found in build-in function of MATLAB – knnsearch. kNN with city block distance achieved best results from all tested distances for two colour conversion but almost the worst for eight colors conversion. Accuracy of classification is comparable with best results achieved in previous works in gender recognition subject. Chebyshev distance turns out as the worst distance that can be used in the gender classification with chosen algorithms of features extraction. Linear correlation distance has as good as city block and Euclidean distance results for big databases, but do not classify correctly when the size of database is decreasing. Cosine – like in linear correlation distance cosine distance achieve good results only for big databases. Euclidean distance was found to be the best distance for eight color conversion and second best for two color conversion. Hamming distance provides the same results like cosine distance. Jaccard similarity coefficient distance has the best results for small databases but worse than Euclidean an Minkowski distance for big databases in eight colors conversion. Minkowski distance was found to be the same good as the Euclidean distance, but is more computationally demanding so was not chose as the best. Spearman's rank correlation distance has almost the worst results for both features extraction algorithms. For Multilayer Perceptron best results was conducted for network with 200 hidden neurons – learning rate 0.1 for two colour conversion as feature extraction and 0.3 for eight colour conversion as feature extraction. Eight color conversion algorithm turned out as more suitable for this neural network than two color conversion. Number of correct classified photos is almost stable between 75 and 50 pictures in learning set, and fall noticeable when the learning set decrease to only 30 photos. For SVM the best results was conducted with quadratic kernel function. Here also eight color conversion was more suitable algorithm of gender features extraction.

This classifier is really dependent of what pictures are in learning set. Sometimes removing some noisy photos from this set leaded to bigger number of correctly classified pictures. For kNN the best results was conducted with city block distance and 14 neighbors for two colors conversion and Euclidean distance and 6 neighbors for eight colors conversion. Like in the rest of classifiers eight colors conversion is more suitable algorithm of gender features extraction. kNN need large database to classify correctly, but it reaches even 100% accuracy. It has the best results of all tested classifiers. The results was better than Nazir et al. [10] work, where Discrete Cosine Transformation was used to extract gender features from face.

4 Emotion Recognition

4.1 Extracting Face Features

3 Haar-based classifiers implemented are used for the purpose of this thesis: face classifier [5], eyes classifier [5], mouth classifier [12]. In order to make the detection faster, in the first step, the algorithm is looking for a face. After finding it, searching for eyes and mouth is done only in area detected as face. Face and eyes classifiers work good and mark correct image areas in 95% of cases so no further modifications were required here. Mouth classifier however was much more problematic. Even though several kinds of those were tested, none of them was giving satisfactory results. In the Fig. 5 areas detected as mouth by the classifier are shown.

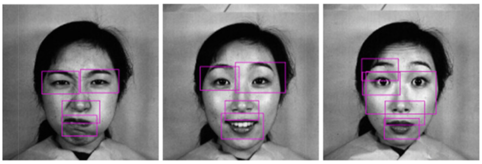


Fig. 5. Areas detected by mouth classifier.

It can be seen that the real mouth area is detected properly however it also marks an inconsistant number of other areas. Due to the fact that to minimize time of execution, eyes and mouth are searched only in area marked as face, the highest vertical coordinate of found edge of area is almost always the part of the correct mouth section. This is how the right mouth area is found. Neural network needs a fixed value of inputs and outputs so such values should have been determined. The optimal ones were chosen experimentaly. Best results were acquired when size of eye images were 20×30 pixels and mouth 30×25. Pictures below show the whole image preprocessing that is done starting from an input image up to creating 3 new images of both eyes and mouth. Pixel values are represented in 8-bit grayscale which translates into [0.255] range of values. However for the purpose of neural network it is better when input values belong to [0.1] or [–1.1] range. In this particular case [–1.1] gave noticeably better results during testing so it was decided to pick such range. Pixel values are read row by row, image by image starting from left eye picture, converted using formula presented above and saved into text file.

4.2 Neural Network Topology

Number of inputs equals number of pixels from all rectangles containing eyes and mouth. The eyes images are sized 20×30 pixels and mouth 30×25. These results in total amount of 1950 input neurons with values from range [–1.1]. Each output neuron represents one of 7 emotions: happiness, sadness, fear, disgust, surprise, anger, neutral. Output neurons values are from range [0.1] or [–1.1] – both cases were tested. The model of neural network chosen is multilayer perceptron with one input layer, one hidden layer and one output layer. Several activation functions were tested as well as different numbers of neurons in hidden layer. Neural network was created using FANN library. Samples are read from text file which is organized in the way that inputs of each sample are followed by outputs desired values. Since the goal is to find the main emotion shown in the picture, output neuron asigned to correct emotion is given value 1, the rest are all given value 0 or –1. Learning and testing is done using different datasets and separate text files.

4.3 Emotions – Results

The sources of testing samples used for the purpose of training and testing neural network are: the Japanese Female Facial Expression (JAFFE) Database (147 photographs in grayscale) [8], own, self-taken photographs of different people (105 photographs in color). Both databases provide photographs taken frontal with one face per picture. First source contains photographs taken in laboratory conditions – similar lightning and distance between a person and a camera. Second source provides photographs taken in varying lightning conditions and distances from camera which allows to test how the system behaves when given more naturally taken photographs. Every person was asked to pose each emotion 3 times.

Output values of neural network are from range [0.1] or [–1.1]. In the training dataset an output neuron representing correct emotion shown by the person in the photograph is given maximum value from range, rest are given minimum value. The actual output values obtained by the process of testing the network may of course take any value from a set range. The main task of emotion recognition software is to find a main emotion that was shown by the person in the photograph. Therefore the main found emotion is the highest value of 7 output neurons assigned to each emotion. Some emotional states however are ambiguous so it was decided to expand detection and detect second emotion as well. Effectiveness relates to percentage of correctly detected emotions in first try – output neuron assigned to the right emotion has the highest value from all output ones. Semi-effectiveness relates to percentage of correctly detected emotions in first try or second try – output neuron assigned to the right emotion has the highest or second highest value from all output ones.

Activation functions that will be tested for the purpose of measuring effectiveness of recognition are listed below. They are part of FANN library [7]:

– FANN_SIGMOID – also called hyperbolic tangent, one of the most popular activation functions. Used for output range: [0.1].
– FANN_SIGMOID_SYMMETRIC – symmetric version of above function. Used for output range: [–1.1].
– FANN_SIGMOID_ STEPWISE – approximation (linear) of sigmoid, which is faster but less precise. Used for output range: [0.1].

– FANN_SIGMOID_SYMMETRIC_STEPWISE – symmetric version of above function. Used for output range: [–1.1].

Above activation functions were tested using samples taken in laboratory conditions with different amounts of hidden neurons in order to find the most effective network (Figs. 6, 7, 8 and 9).

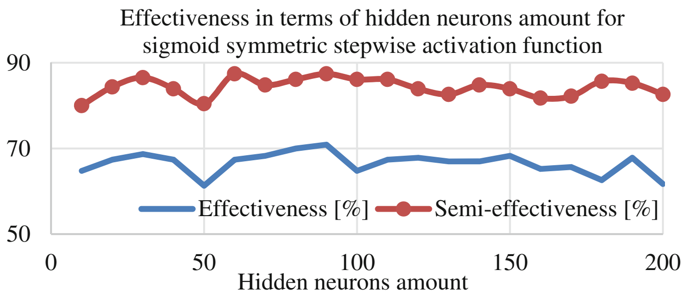


Fig. 6. Effectiveness in terms of hidden neurons amount for stepwise symmetric function.

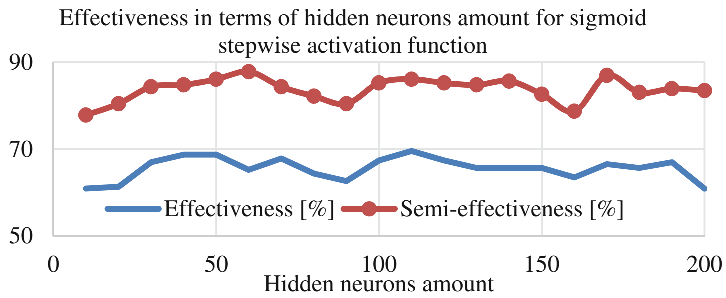


Fig. 7. Effectiveness in terms of hidden neurons amount for stepwise function.

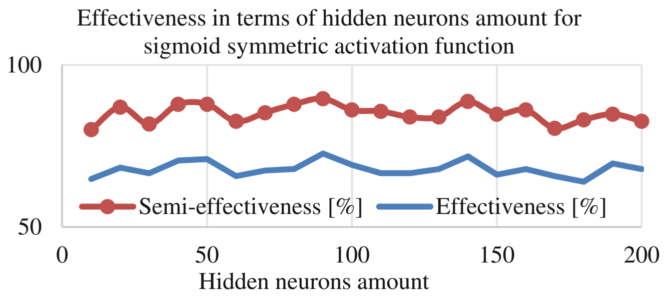


Fig. 8. Effectiveness in terms of hidden neurons amount for sigmoid symmetric function.

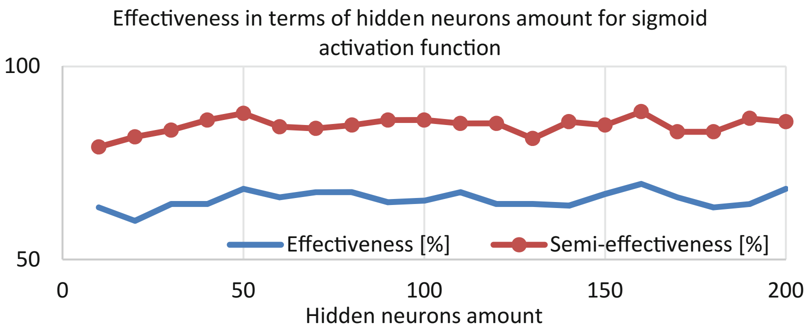


Fig. 9. Effectiveness in terms of hidden neurons amount for sigmoid function.

Training is done with a desired error value set to 0.0001. There are 100 samples in training dataset and 47 in testing one. People of whom pictures were taken are the same in both datasets however all pictures of them are different. During creation of neural network weights are assigned randomly therefore outcomes may vary [6]. This is why for a precise evaluation every test has been done 10 times and the outcome shown below is the average value of 10 tries.

For both sigmoid and stepwise sigmoid activation functions better results are obtained in symmetric case- when outputs are from range [–1.1] instead of [0.1]. The best results (72.6%) are obtained for sigmoid activation function when outputs are in range [–1.1] and the amount of hidden neurons equals 90. This case also gives the best semi-effectiveness ratio (89.57%). Sigmoid activation function is however a bit slower than it's stepwise approximation but the average value of time difference is 0.4862 s so it not a significant difference. In conclusion, the most effective setup of neural network is: outputs from range [–1.1], sigmoid symmetric activation function, 90 hidden neurons. Recognition ratio of all emotions can be seen in Table 1.

Table 1. Recognition ratio of all emotions.

Emotion	Recognition ratio	
	Effectiveness [%]	Semi-effectiveness [%]
Neutral	75.00	88.75
Happiness	58.33	95.00
Sadness	77.50	95.00
Fear	51.11	75.56
Surprise	85.71	94.29
Disgust	72.50	97.50
Anger	77.50	90.00

5 Conclusions

The way how people show emotions vary depending on race and personal form of expression of each individual. However humans can identify the mood of others even if they do not know them therefore the main patterns for showing certain emotion must repeat. This is why this task requires high amount of training dataset consisting images of people from different races, ages and sexes. While there are classification tasks where there is one sure answer to every question, like car or mobile phone model recognition when it is known for sure a certain object should be classified as one brand or another. In this paper the problem is different because sometimes even humans have problems with detecting emotion shown in the picture. This is why achieving 100% recognition ratio is almost impossible. In terms of feature extraction Haar-based classifiers can achieve fast and accurate extraction of face regions required for further processing. 79% recognition ratio which was obtained by neural network in this thesis is a promising result. This could be improved with more training samples. Photographs do not need to be taken in laboratory conditions and thanks to geometric approach used for finding face features there are no constraints regarding color or lightning as long as face is visible. The easy algorithms of gender features extraction can work really great when we want to classify gender of people that are the most common in some particular region, like in this case only Caucasians. It works with most successful classifier with 100% accuracy, this is why there had been no need of developing other way of picture preprocessing.

References

1. Baluja, S., Rowley, H.: Boosting sex identification performance. Int. J. Comput. Vis. **71**(1), 111–119 (2007)
2. Burton, A., Bruce, V., Dench, N.: What are the Differences between Men and Women? Evidence from Facial Measurement Perception (1993)
3. Cottrell, G., Metcalfe, J.: EMPATH: face, emotions and gender recognition using holons. In: Neural Information Processing Systems, pp. 564–571 (1990)
4. Ekman, P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. University of California, San Francisco (2003)
5. http://docs.opencv.org/3.2.0/d7/dbd/group__imgproc.html. Accessed May 2017
6. http://leenissen.dk/fann/fann_1_2_0/r2030.html. Accessed June 2017
7. http://leenissen.dk/fann/html/files/fann_data-h.html. Accessed June 2017
8. http://www.kasrl.org/jaffe.html. Accessed May 2017
9. Khan, S.A., Nazir, M., Riaz, N., Naveed, N.: Computationally intelligent gender classification techniques: an analytical study. Int. J. Sig. Process. Image Process. Pattern Recogn. **4**(4), 145–156 (2011)
10. Nazir, M., Ishtiaq, M., Batool, A., Jaffar, A., Mirza, M.: Feature selection for efficient gender classiffcation. In: WSEAS International Conference (2010)
11. Rosenblatt, F.: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC (1962)

12. Santana, C.M., Suárez, D.O., Tejera, H.M., Artal, G.C.: ENCARA2, real-time detection of multiple faces at different resolutions in video streams. J. Visual Commun. Image Represent. **18**(2), 130–140 (2007)
13. Shakhnarovich, G., Viola, P., Moghaddam, B.: A unified learning framework for real time faces detection and classification. In: IEEE Conference on Automatic Face and Gesture Recognition, pp. 14–21 (2002)

Implementation Efficiency of BLAKE2 Cryptographic Algorithm in Contemporary Popular-Grade FPGA Devices

Jarosław Sugier(✉)

Department of Computer Engineering, Faculty of Electronics,
Wrocław University of Science and Technology,
Janiszewskiego St. 11/17, 50-372 Wrocław, Poland
jaroslaw.sugier@pwr.edu.pl

Abstract. This paper explores hardware implementations of BLAKE2 cryptographic hash function in the newest popular-grade Spartan-7 device from Xilinx. Specifically, it discusses a particular modification in organization of the cipher hardware eliminating the involved data paths for distribution of message bits among round units – which is accomplished by application of block memory modules for repetitive storage of the message inside each round instance. The idea was applied in four different organizations of the algorithm: the typical iterative one and three high-speed loop-unrolled architectures with 2, 4 and 5 rounds instantiated in hardware. Together with standard (without RAM) implementations this produced a set of 8 test cases: the implementation results allowed to evaluate the proposed modification which, like in our previous studies which used older device families, led to outstanding reductions in FPGA array utilization and also to some improvements in performance parameters.

Keywords: Spartan-7 family · Implementation efficiency · Resource utilization Block RAM · Loop unrolling

1 Introduction

Malfunctions caused by security violations are so common in operation of contemporary IT systems that application of appropriate cryptographic methods is a necessity in system design and maintenance. This paper deals with hardware implementations of one such method – the BLAKE2 hash function.

In particular we discuss a specific modification in hardware realization of this algorithm which eliminates the complicated data paths distributing message bits among the round units by repetitive storage of the message inside each round instance. The idea – which was initially presented in [9] and then extended in [8] – is applied to implementations of the algorithm in four different organizations: the standard iterative one and three high-speed loop-unrolled architectures with 2, 4 and 5 rounds instantiated in hardware. Together with standard (without RAM) implementations this produced a set of 8 test cases.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_44

The paper is organized as follows. After introducing the BLAKE2 method in the next section, in Sect. 3 we present organization of cipher implementations and evaluate required capacity of memory resources needed for realization of the proposed modification. Section 4 contains implementation results and their evaluation.

2 The BLAKE2 Hash Algorithm

The original BLAKE algorithm [1] was a candidate proposed in the SHA-3 contest where it successfully qualified to the final round with other 4 algorithms, being praised for its high cryptographic strength and good implementation efficiency, especially in software. In 2013 the authors published specification of an evolved version of the method – called BLAKE2 – with modifications aimed mainly towards its simplification and optimization because the SHA-3 evaluation showed that the original proposal offered an unnecessary large security margin. In this work we will consider the BLAKE2s version of the cipher [2] which works with 32-bit words and can produce a 256-bit hash (the larger BLAKE2b variant operates on 64b words and generates 512b digests).

The algorithm begins with splitting the input message m into 512b blocks $m^0 \ldots m^{N-1}$ (if needed, the last block is null-padded) and the hash output $h(m)$ is iteratively computed in the HAIFA [3] iteration scheme:

$$
\begin{gathered}
h^0 := \mathrm{IV} \oplus \mathrm{P}, \\
\text{for } i = 0 \ldots N-1, \\
h^{i+1} := \text{compress}\left(h^i, l^i, m^i\right), \\
\text{return } h^N,
\end{gathered}
$$

where IV is a 64-byte constant pattern adopted from the SHA-2 standard, P – a block of hash parameters (e.g. requested digest length or tree-hashing parameters), l_i denotes number of data bytes in $m_0 \ldots m_i$ blocks and compress() is a compression function processing a single message block. Like in other hash algorithms based on Merkle-Damgård construction computing the digest of a free-length message stream consists in repetitive application of this function and its implementation in hardware is the actual challenge in realization of the whole algorithm.

Internal processing of the compression is organized around a state – a 4×4 matrix of words $v_0 \ldots v_{15}$ which are initialized, among others, with the current chain hash value h_i. The compression consists in repeating $n_r = 10$ rounds over the state, with each round applying so called G functions twice:

$$G_0\left(v_0, v_4, v_8, v_{12}\right);\ G_1\left(v_1, v_5, v_9, v_{13}\right);\ G_2\left(v_2, v_6, v_{10}, v_{14}\right);\ G_3\left(v_3, v_7, v_{11}, v_{15}\right); \tag{1}$$

and

$$G_4\left(v_0, v_5, v_{10}, v_{15}\right);\ G_5\left(v_1, v_6, v_{11}, v_{12}\right);\ G_6\left(v_2, v_7, v_8, v_{13}\right);\ G_7\left(v_3, v_4, v_9, v_{14}\right). \tag{2}$$

Particular function $G_i(a, b, c, d)$ transforms four state words and uses two of the $m_0 \ldots m_{15}$ message words as the auxiliary input. The transformation is defined as:

$$
\begin{aligned}
a &:= a + b + m_{\sigma r(2i)} \\
d &:= (d \oplus a) >> 16 \\
c &:= c + d \\
b &:= (b \oplus c) >> 12 \\
a &:= a + b + m_{\sigma r(2i+1)} \\
d &:= (d \oplus a) >> 8 \\
c &:= c + d \\
b &:= (b \oplus c) >> 7
\end{aligned} \tag{3}
$$

where the operators +, ⊕, and >> denote, respectively, 32b addition, bitwise xor and vector rotation, and the indices of the message words $m_{\sigma r(2i)}$ and $m_{\sigma r(2i+1)}$ are taken from ten permutations $\sigma_0 \div \sigma_9$ defined statically in the standard.

3 Hardware Implementations of the Cipher

3.1 Organizations of Data Processing

Hardware implementation of hash functions has been a subject of intensive studies for many years and there is an extensive literature discussing possible solutions (see for example [4–6]). In this work we are investigating high speed architectures which, with application of loop unrolling techniques [4], strive for high performance at the cost of large hardware utilization. In the following analyses after taxonomy used in [4] an architecture with k unrolled rounds will be denoted as xk while the basic iterative one – as x1. Like in our previous work [9] the test suite will comprise the following 4 organizations:

- x1: the basic iterative architecture with one round implemented in hardware and the state being passed though it repeatedly in 10 clock cycles (i.e. each complete round is computed in one clock tick);
- x2: modification of the above with a combinational cascade of two rounds implemented in hardware with total computation done in 5 clock cycles (with each clock tick the state is propagated through two rounds);
- x4: the cascade is built from 4 rounds and 3 clock cycles are required for complete computation (the final result is taken from the second round to get $n_r = 10 = 2 \times 4 + 2$);
- x5: as in the previous case but with 5 rounds unrolled in hardware and the computation takes 2 cycles.

The idea of using block memory resources for repetitive storage of message words in order to eliminate their involved distribution among cipher rounds was initially proposed in [9] for the original BLAKE algorithm but it can be directly applied also to its successor. It consists in assigning one dual-port block RAM module for each instance of the G_i function and storing the m_i words directly within as it is illustrated in Fig. 1.

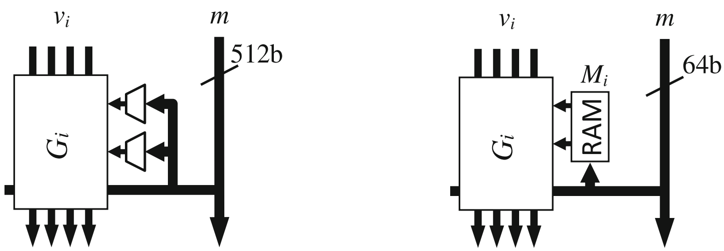


Fig. 1. Distributing and multiplexing the message bits for a G_i module in the standard implementations (left) versus storing them in a dual-port RAM unit assigned to the module (right).

Xilinx Spartan FPGA devices which are used in this study offer so called block RAM modules and these were used for this purpose. Each module can store 16 kb of data and can be configured with different depth vs. width ratios – in organizations from 16 k × 1b to 512 × 32b. For our application the last case – 512 × 32b – is a suitable one with each m_i word occupying exactly one memory location. Additionally, the dual port functionality of the modules allows to read concurrently two different m_i words in one clock cycle as they are required for computation in the G_i function.

3.2 Evaluating Memory Requirements

When introducing the proposed method of message storage we have assumed for simplicity that all 16 message words are stored in each memory module regardless the fact whether they all are actually needed in computations of a particular G function. Such a simplification was fully justified in the case of FPGA implementations because in those devices capacity of available RAM units is not configurable so optimizing the number of stored message words according to the actual demand would not improve neither size nor performance (pre-loading delay) of the implementation. In this subsection we will adopt another perspective and exactly analyze how much memory is needed in each G module – this could be a crucial information in ASIC implementations where size and organizations of RAM units is fully adjustable.

Distribution of the message words depends on the σ permutations and these are presented in Table 1 [2]. Let's consider operation of some cipher round instance Rj – in x1 this would be the only instance $R0$, in x2 – $R0$ or $R1$, etc. During calculations of the r'th round the instance selects the σ_r permutation and this determines which m_i words will be needed by its G functions. Looking at the columns assigned to each G_i instance and considering which round numbers will be computed by the Rj (depending on the architecture not all round numbers are computed in all Rj instances) we can find how many message words will be needed in each G_i module – hence what capacity of the associated memory module is required.

Let's μ_i denote number of message words appearing in Eq. (3) of specific G_i function and M_i – the associated memory module. In the basic x1 architecture all the rounds are computed in the one (and only) $R0$ module so all the permutations affect the G instances contained within. Because not all word indices appear in the first two columns of the

Table 1. The σ permutations and their elements assigned to the instances of the G function.

	G_0		G_1		G_2		G_3		G_4		G_5		G_6		G_7	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
σ_0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
σ_1	14	10	4	8	9	15	13	6	1	12	0	2	11	7	5	3
σ_2	11	8	12	0	5	2	15	13	10	14	3	6	7	1	9	4
σ_3	7	9	3	1	13	12	11	14	2	6	5	10	4	0	15	8
σ_4	9	0	5	7	2	4	10	15	14	1	11	12	6	8	3	13
σ_5	2	12	6	10	0	11	8	3	4	13	7	5	15	14	1	9
σ_6	12	5	1	15	14	13	4	10	0	7	6	3	9	2	8	11
σ_7	13	11	7	14	12	1	3	9	5	0	15	4	8	6	2	10
σ_8	6	15	14	9	11	3	0	8	12	2	13	7	1	4	10	5
σ_9	10	2	8	4	7	6	1	5	15	11	9	14	3	12	13	0

table which are assigned to the G_0 unit (the missing ones are 3 and 4) so the associated M_0 module needs to store $\mu_0 = 14$ words, without m_3 and m_4. Actually, most of the other M_i modules in the x1 configuration must store 14 words, with the exceptions of M_4 (in columns 8 and 9 only index 3 is omitted and $\mu_4 = 15$) and M_7 ($\mu_7 = 13$, without m_6, m_7 and m_{12}).

In the unrolled architectures the round instances compute specific (and not all) rounds so the reductions are greater. In the extreme x5 case each Rj instance uses just two permutations and there are some G modules where they generate $\mu_i = 3$ (e.g. the $R2$ applies σ_2 and σ_7 and the G_0 instance receives m_{11}, m_8 in the first permutation and m_{13}, m_{11} in the second one).

Table 2. Evaluating actual memory requirements in all four architectures.

Round instance		Permutations	Range of $\mu_0 \div \mu_7$	In round $\Sigma_i\, \mu_i$	Total $\Sigma_R\, \Sigma_i\, \mu_i$	Total [kb]
x1	$R0$	$\sigma_0\sigma_1\ldots\sigma_9$	$13 \div 15$	112	112	3.50
x2	$R0$	$\sigma_0\sigma_2\ldots\sigma_8$	$7 \div 10$	69	142	4.44
	$R1$	$\sigma_1\sigma_3\ldots\sigma_9$	$8 \div 10$	73		
x4	$R0$	$\sigma_0\sigma_4\sigma_8$	$5 \div 6$	45	150	4.69
	$R1$	$\sigma_1\sigma_5\sigma_9$	$4 \div 6$	44		
	$R2$	$\sigma_2\sigma_6$	$2 \div 4$	30		
	$R3$	$\sigma_3\sigma_7$	$3 \div 4$	31		
x5	$R0$	$\sigma_0\sigma_5$	4	32	156	4.88
	$R1$	$\sigma_1\sigma_6$	4	32		
	$R2$	$\sigma_2\sigma_7$	$3 \div 4$	31		
	$R3$	$\sigma_3\sigma_8$	$3 \div 4$	30		
	$R4$	$\sigma_4\sigma_9$	$3 \div 4$	31		

The results of this tedious evaluation of the μ_i values in all the architectures are presented in Table 2. For every organization xk and every round instance Rj the second column lists the permutations applied within, the third one gives min-max range of the μ_i values and the fourth – total capacity of the M_i units. The last two columns summarize memory for the whole design (in all rounds): the fifth expresses the total storage in a number of 32b words and the last one in kilobits.

4 Implementation Results

4.1 Implementing the Designs

The implementations were accomplished for a design where the main BLAKE2 unit was equipped with rudimentary input/output registers providing means for iterative hashing of the message in 512b chunks. Eight such designs (four architectures implemented twice: in the standard way and with the proposed RAM utilization) were automatically synthesized and implemented by Xilinx Vivado software for the Spartan-7 XC7S50 FGGA484-2 device [10]. The same approach was applied in our previous works on BLAKE and BLAKE2 [7–9] so an already existing test platform was uniformly extended to accommodate the BLAKE2 version on the new Spartan-7 platform, keeping the ability to produce comparable results. The only but unavoidable change was in the used implementation software: the previous results for Spartan-3 and 6 had to be obtained with the ISE package while the new 7 Series family (and only this family) is supported exclusively in the newer Vivado suite.

Table 3. Parameters of the implementations.

Case	T_{clk}	Levels of logic	Routing delay	LUTs	Slices	F7 Muxes	F8 Muxes
Standard organizations							
x1	22.8	51	50%	4 530	1 339	1 024	512
x2	42.7	88	52%	8 080	2 128	1 984	992
x4	83.5	181	52%	15 302	3 994	3 968	1 984
x5	90.6	204	51%	14 499	4 134	1 216	544
Modified with RAM							
RAMx1	21.1	48	41%	2483	716	0	0
RAMx2	39.5	89	43%	4 015	1 140	0	0
RAMx4	76.3	172	48%	7 343	2 076	0	0
RAMx5	88.0	201	49%	11 360	3 078	0	0

The parameters of all 8 test cases are included in Table 3. Performance aspect is represented in the second column by the value of the minimum clock period as it was estimated after static timing analysis of the final, fully routed design. The two next columns provide parameters which illustrate effectiveness (or difficulties) of the routing implementation: for the longest combinational path in the design the third column gives number of logic elements it contains and the fourth – percentage of its propagation delay

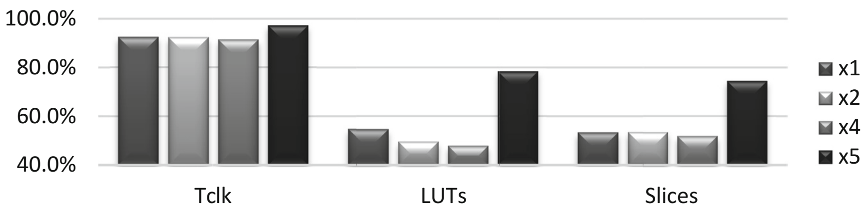


Fig. 2. Speed and size of the proposed RAM-based implementations as percentages of the standard cases.

incurred by the routing resources (and not logic elements). Any significant rise in the latter parameter indicates problems with routing of connections between logic elements of the array. Size characteristics are reported in the following columns: number of utilized LUT generators is given as the estimation of used logic cells, while the slices indicate occupied nodes of the FPGA array. The last reported elements – F7 and F8 multiplexers – are the elements needed in implementation of wide boolean functions which, if have more than 6 arguments, must be split between multiple LUT elements.

Before more detailed analyses will start, one peculiarity should be commented regarding implementation of the x5 case in the standard way (i.e. without RAM). As it can be seen, this design turned out to be actually *smaller* than the corresponding x4 case looking at the LUT elements and only marginally larger in the slice number – something which is definitely not expected as this should be the largest case with five round instances in silicon. The situation can be explained looking at the utilization of the F7 and F8 multiplexers: in x5 this utilization is much lower than the trend in x1–x4 cases would suggest which indicates that the Vivado synthesis tool took different optimization strategy in combining the involved processing of *G* functions with the multiplexers feeding the m_i words. Such irregularities caused by different optimization techniques could also be observed in parameters of this and other ciphers in older Spartan-3 and (especially) Spartan-6 devices [7–9].

No such anomaly can be seen in implementations which use the proposed RAM modification: size of the x5 design increases quite proportionally compared to the x4 one. It is worth noting that in these designs the implementation tools did not use the F7 and F8 multiplexers at all.

4.2 Size and Speed Effects of the Proposed RAM Application

Based on the data from Table 2 we can compare size and speed of implementations modified in a way proposed in this paper against the results of the standard approach. Such a comparison is the purpose of Fig. 2: the minimum clock period T_{clk} (speed) and the number of LUT generators or slices (size) for implementations with RAM are expressed as percentages of corresponding values of standard realizations – for all four organizations. As one can see, in all cases the percentages are below the 100% level, i.e. the RAM implementations were faster (shorter T_{clk}) and smaller than their traditional counterparts.

Like in the previous studies included in [8, 9] made for the older platforms, the modification brings exceptional size reductions also for BLAKE2 cipher in Spartan-7 devices. Leaving aside the x5 case for a moment due to the reasons presented above, the x1–x4 RAM architectures were implemented with 55–48% of LUTs and 54–52% of slices of their standard counterparts. In the previous studies the reductions were even few percent bigger for some cipher/platform combinations but in overall variation in the reduction factor was bigger. The new family produces the most stable size improvements.

Such stability (still within the x1–x4 range) is even more demonstrated in performance parameter: in all these instances the clock period was reduced exactly to 92% (±0.6%) while in previous studies the similar average reduction was achieved with variance at least an order of magnitude bigger. This also suggests better robustness and stability of the implementation tools in the new Vivado software. Another aspect supporting this trend is better efficiency of routing resources in the new Spartan-7 device: as the Table 2 showed the routing part of the longest delay remains within 41–49% in the RAM designs and within 50–52% in the standard ones, showing no signs of routing congestion which often plagued older platforms, especially Spartan-6.

Returning to the x5 case we can further explain consequences of a different synthesis optimization in this item: while it indeed reduced the size of the standard x5 design making the LUT and slice ratios in Fig. 3 noticeably bigger (as the same was not accomplished in the RAM x5 case) it also affected the performance since the reduction of the x5 clock period was the smallest one – albeit still down to 97%.

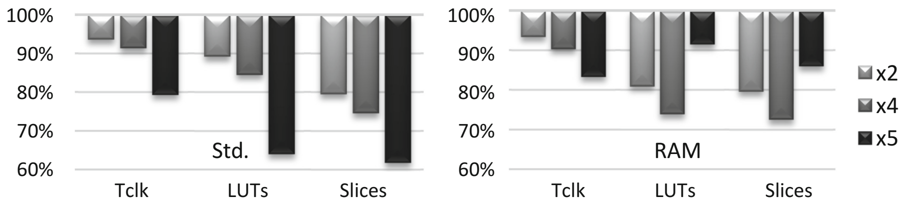


Fig. 3. Parameters of the unrolled architectures as percentages of the estimations drawn from the x1 case.

4.3 Scalability of the Loop Unrolling Mechanism

Scalability in implementations of the loop unrolled architectures is the ability to keep size and speed in linear proportion to the number of rounds instantiated in hardware. As the previous studies have shown e.g. in [7] some contemporary cryptographic algorithms may exhibit significant weaknesses in this aspect, mainly due to their very involved and irregular internal organization which is difficult to map to the FPGA array in larger (highly unrolled) organizations. As a result they often tend to occupy disproportionately large amount of resources and operate slower due to routing congestion problems.

Taking the parameters of the x1 implementation it is possible to calculate expected clock period and size parameters of the unrolled xk cases like it was described e.g. in [7]. Figure 3 presents this kind of results as the ratios actual_value/estimation for the cases investigated in this paper. The lower the ratio, the faster (shorter Tclk) or the smaller (lower number of LUT or slices) was the unrolled design in relation to what could be expected from its x1 case. The value of 100% is the threshold separating “better than” (<100%) from “worse than” (>100%) the expected.

Once more, all the ratios being below the 100% threshold means that in all cases the unrolling produced better (faster and smaller) implementations than it could be expected from the x1 cases, both for the standard and the RAM-modified implementations of the algorithm. Of all the cipher/platform combinations (BLAKE | BLAKE2/S3 | S6 | S7) which were included in the previous studies this is the first time when such a situation occurs: scaling especially in the clock period was more problematic for both versions of BLAKE in the older device families [8]. This is another confirmation of good stability and superior potential of the new Spartan-7 resources.

Differences among the data presented in Fig. 3 can help in drawing additional conclusions. Again the standard x5 implementation stands out with significantly higher reductions but this is as expected due to already discussed singularity of this case. Putting it aside, the remaining results look more consistent with rather subtle variations. On average the reductions are bigger in the RAM versions than in the standard ones, although the differences are not as dominant as they were on the Spartan-3 and 6 devices [8]. On the older platforms especially the standard implementations, being larger and putting more stress on the FPGA resources, exhibited routing congestion which could seriously impair efficiency of the unrolling mechanism. No signs of these problems were identified on the new Spartan-7 platform.

5 Conclusions

This work is a continuation of our previous studies which proposed a specific modification in hardware implementations of the BLAKE/BLAKE2 hash algorithm – a modification which eliminates the need for involved data paths distributing message bits among the cipher rounds by using block memory units provided as auxiliary FPGA resources. To the already tested, older Spartan-3 and Spartan-6 devices a new Spartan-7 platform (introduced in 2017) was added. As previously, the tests included 4 different organizations of the cipher: the standard iterative one and three high-speed loop-unrolled architectures with 2, 4 and 5 rounds instantiated in hardware. Together with standard (without RAM) implementations used for comparison this produced a set of 8 test cases.

The implementation results confirmed again that the modification outstandingly enhanced size of all the tested architectures also on the new platform: on average, occupation of the Spartan-7 array was reduced by half while the improvements in speed, although not so spectacular, were also verified. These advances are comparable with the ones achieved with the older platforms but what characterized the new results was their better stability and predictability, especially in the cases of the more sized, loop unrolled architectures which often created problems in the older devices. Potential of the new Spartan-7 architecture was confirmed particularly in its routing capabilities which were sufficient in all test cases and did not produce congestion problems which plagued specifically the Spartan-6 implementations.

References

1. Aumasson, J.P., Henzen, L., Meier, W., Phan, R.C.W.: SHA-3 proposal BLAKE, version 1.3. https://www.131002.net/blake/blake.pdf. Accessed May 2017
2. Aumasson, J.P., Neves, S., Wilcox-O'Hearn, Z., Winnerlein, C.: BLAKE2: simpler, smaller, fast as MD5. In: Jacobson, M., Locasto, M., Mohassel, P., Safavi-Naini, R. (eds.) Applied Cryptography and Network Security, ACNS 2013. LNCS, vol. 7954, pp. 119–135. Springer, Heidelberg (2013)
3. Dunkelman, O., Biham, E.: A framework for iterative hash functions: Haifa. In: 2nd NIST Cryptographich Hash Workshop, vol. 22 (2006)
4. Gaj, K., Homsirikamol, E., Rogawski, M., Shahid, R., Sharif, M.U.: Comprehensive evaluation of high-speed and medium-speed implementations of five SHA-3 finalists using Xilinx and Altera FPGAs. In: The Third SHA-3 Candidate Conference, p. 368. IACR Cryptology ePrint Archive (2012)
5. Gaj, K., Kaps, J.P., Amirineni, V., Rogawski, M., Homsirikamol, E., Brewster, B.Y.: ATHENa – Automated Tool for Hardware EvaluatioN: toward fair and comprehensive benchmarking of cryptographic hardware using FPGAs. In: 20th International Conference on Field Programmable Logic and Applications, Milano, Italy (2010)
6. Junkg, B., Apfelbeck, J.: Area-efficient FPGA implementations of the SHA-3 finalists. In: 2011 International Conference on Reconfigurable Computing and FPGAs (ReConFig), pp. 235–241. IEEE (2011)
7. Sugier, J.: Implementation efficiency of BLAKE and other contemporary hash algorithms in popular FPGA devices. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) Proceedings of 11th International Conference Dependability and Complex Systems DepCoS-RELCOMEX, AISC, vol. 470, pp. 457–467. Springer, Heidelberg (2016)
8. Sugier, J.: Improving FPGA implementations of BLAKE and BLAKE2 algorithms with memory resources. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) Dependability Problems and Complex Systems, Proceedings of 12th International Conference Dependability and Complex Systems DepCoS-RELCOMEX. AISC, vol. 582, pp. 394–406. Springer, Heidelberg (2017)
9. Sugier, J.: Simplifying FPGA implementations of BLAKE hash algorithm with block memory resources. Procedia Eng. **178**, 33–41 (2017)
10. Xilinx, Inc.: 7 Series FPGAs Data Sheet: Overview. DS180.PDF. www.xilinx.com. Accessed May 2017

Performance Comparison of Observer Design Pattern Implementations in JavaScript

Artur Zochniak and Tomasz Walkowiak(✉)

Faculty of Electronics, Wroclaw University of Science and Technology, Wrocław, Poland
zochniak@protonmail.com, tomasz.walkowiak@pwr.edu.pl

Abstract. Websites serving dynamic content must access remotely stored data to present it in a browser, but before that an appropriate document has to be built. The process of building a WWW document is done fully on the client-side. That means that its duration can be measured and results compared among different solutions. To ensure the best possible user experience it should be as short as possible. Authors present performance comparison of observer design pattern implementations in two JavaScript frameworks: AngularJS, EmberJS. Different types of data observer implementations are described. Authors implemented an exemplar application in all analyzed frameworks and tested their performance in a function of size of input data. The fastest solution is shown and reasons of differences is analyzed. The presented results allow to build fast response web pages.

Keywords: Web application · JavaScript framework · Performance analysis

1 Introduction

Modern Internet has a lot to offer in terms of volume of available data, but not only amount of available data is important, another important factor is the method how the data is presented. In World Wide Web there are two kinds of web pages: static and dynamic ones. Static web pages are served as a result and browser has only to present them on the user side whereas dynamic web pages are constructed on the client side. This paper describes approaches to rendering dynamic web pages in web browsers and discusses performance of frameworks easing building modular single page applications (SPA) as well as provides example automated performance testing environment design and, at the end, takes into consideration what business benefits might come from wisely choosing technology for building or maybe, better to say, what we can avoid by doing that (the problem is discussed in [1, 2]).

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_45

2 JavaScript Frameworks

2.1 Web Pages

Web pages represent content available on some URL (Uniform Resource Locator). After user requests a resource address, it is being translated to server address via DNS (Domain Name System), from which browser downloads main document (written in HTML language) and later performs subsequent requests asking for additional resources required by the website to function properly. Additional resources could be images, they can define how the web page looks like (CSS) or provide additional scripting for the web page. In this paper we focus on scripts being executed on the webpage which happen to be written in JavaScript. It is worth mentioning that JavaScript is just a language and does not enforce any architecture on the implementation.

2.2 JavaScript

JavaScript allows creating web pages with dynamic content. Dynamic content encourages interactivity and this increases user experience, thus software manufacturers tend to introduce as much interactivity as possible to satisfy user needs. Multiple libraries created for easing creating web page applications have evolved into sets of rules being known as appropriate workflows increasing productivity of implementing recurring product designs as well as contain architectural guidelines for creating web applications reminding desktop application, so called SPAs (single page applications) are released as a software frameworks.

2.3 Single Page Application

Single page applications are web pages which load application code at the launching time and later only needed data or required code is downloaded from the server. Of course, if there is a process of downloading new data from the server, then there has to be event of handling this process. There are multiple solutions to implement recognition of the change of data and this is where design patterns come into play.

2.4 Observer Design Pattern

Design patterns are verified solutions to a general problems in software development. Observer is one of design patterns and its role is to notify interested parties about an event (in case of web pages interested the interested party is a view rendering component).

Observer design pattern in traditional programming languages can be implemented with two classes, first of which is a subject class, and second is observer. At first observer has to subscribe to the subject it is interested in listening to changes of, and once it is done, whenever subject is modified – it may notify its observers about the fact of a change.

Observer design pattern is simple in implementation, yet it is very powerful design pattern, used not only in building the view, but also in general data processing like spreadsheets where changing a cell on which depends a value of another cell has to be propagated to recalculate subsequent cells [3]. Observer design patter, also known as publish-subscribe [4] distributed messaging solutions like Apache Kafka[1].

In the next section, we present multiple approaches solving problem of re-rendering layer of presentation – view-layer in JavaScript frameworks – which highly depend on the detection of data change.

2.5 Types of Data Observer Implementations

Once data is fetched from the server – browser has to present given data somehow to the user. There are few methods how this can be done, those methods differ from each other by complexness of implementation, easiness of using from the point of view of developer and, from the point of interest of this paper, most importantly – performance overhead.

Manually building the view. This is the simplest solution and basically requires creating the whole webpage once something changes. This approach has a drawback of limiting user experience, because the whole state of the user-input is lost. This solution also do not require any JavaScript running on the client side, so has the highest backward compatibility.

Manual notification about changes. This method requires developer to be aware of all the relation between model and the view, not only direct, but also indirect ones. Its main idea is that once webpage is built – further manipulations the web page web page document are performed via DOM (document object model). Trivial implementation of updating partial web page content asks the server for new data and replaces and/or appends received data to a predefined slot on the web page, usually pointed by a CSS selector, XPath or any other unambiguous method of pointing to an element or elements. The advantage of this solution is that it may not require any dependencies and is very simple (only request handler and element selector are needed).

Extended version of this approach is present in some frameworks in such a way that values of expressions presented in rendered view just have to have list of dependent expressions (which in the end have to be explicitly defined by the developer), so this is actually only a wrapper or syntactic sugar around manual implementing handling of received data (EmberJS discussed in this paper partially requires providing explicit list of dependency in case of computed values).

Dirty checking for any change. This implementation of observer design pattern has to be provided with a list of expressions to check for changes during application living cycles and once change was detected – given expression and recursively all expressions dependant on changed one – get recalculated and changes are represented in the view. This approach is convenient for the developer, but in pessimistic case, where there is a lot of expressions to be watched – such implementation might be slow in determining changes to data. One of techniques for increasing performance in case of many watched expressions is taking out some of watched elements and taking them back into account

[1] http://kafka.apache.org, last accessed 2017/07/01.

for single cycle in life of an application. One might find more information on the topic of increasing performance of such a implementation in [5].

Immediate recalculation and intermediate virtual caching layer. This method is just about to get rid of all problems related to DOM legacy implementation issues such as the fact that insertion in DOM is synchronous, as explained in [6], and even if there were significant improvement to DOM implementation [7] – since it is synchronous – it is still a subject to improvements. Virtual-DOM, because this is how this approach is called, acts as a proxy between JavaScript and the DOM. One might think about it like catches all the modifications to the DOM and applies them to the real DOM all at once (and this is the reason why virtual-DOM implementation has to have a differing algorithm). While such a solution might work really fast (and often does [8]). Even contributors of the frameworks known as the fastest ones agree, that comparing frameworks relying on synchronous DOM manipulations versus those making use of asynchronous rendering – due to implementation differences – makes a little sense [9], thus in this publication we do not take into consideration implementations introducing intermediate layer between JavaScript and DOM.

2.6 Relations Between Frameworks

JavaScript, as in its first versions, did not intend to limit programmers' abilities by hiding properties or any form of encapsulation (like it is done in other languages, like Java or C++). Developers have found a way of implementing private fields by technique named closuring (which requires creating function returning a function, where private variables are inaccessible from the outer world after outer function execution is finished and returned inner function). Such an openness of the language leads to various ways of achieving the same thing and many existing frameworks have developed different ways of modifying properties. Figure 1 presents different approaches to modifying properties.

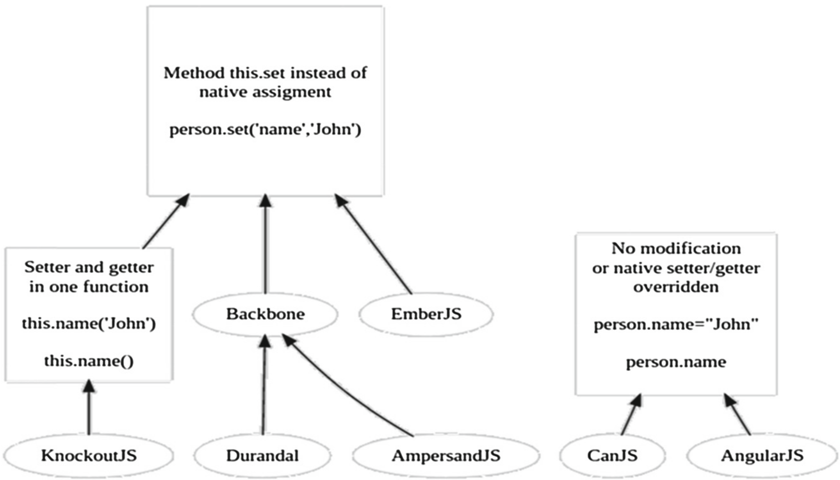


Fig. 1. Diagram of relation between frameworks.

As the figure shows, there are at least 3 ways of changing value:

- via native setter and getter, which directly modify property: object.name = "John";
- via object.set method, which takes property name as first argument, as value as a second: object.set('name', 'John');
- and via function, which dependently on a context it's being called in, might be getter or a setter: object.name("John"), (assert object.name() === "John");

First are native setters and getters and those are the most intuitive way of accessing object's properties, but before ECMAScript 2015 standard it was hard to implement notification of a change of a any property due to lack of defineSetter method (which was added to the standard in 2015[2]).

Third implementation uses something similar to functional currying, where property name isn't passed as first argument of the setter, but instead is accessible and modifiable via its dedicated method, which happens to be named as property value; it is also important that such a method allowing getting the value and setting a new value has to implement caching of current value, because in such a case property name cannot hold any value since it is already a function.

After ECMAScript15 was widely implemented in the browsers – it became obsolete to seek for a different way of implementing observer design patterns, because native API offered it out-of-the-box. More on reactive data structures one can find in [10].

3 Testing Environment

A simple web application was developed to test JavaScript frameworks' performance. It is a web page which uses loops for iterating over test cases. The webpage inserts floating frame (iframe) after every test and inner webpage (within floating frame) informs parent frame that the test has finished and test runner can execute next test by setting an attribute on the document body DOM element, which is periodically checked for changes by the outer frame (the one holding the benchmarking environment).

Example test cases are loaded by entering a dedicated webpage parameterized with variables passed in GET request: framework (either AngularJS or EmberJS), action (adding elements one by one) and number of elements. There are two implementations in separate folders: in-angular and in-ember, but benchmarking tool is designed to allow further extensions for different frameworks. Tool is also prepared for measuring other actions, but when visiting a webpage – often data gets only populated, therefore there is no reason for considering other actions in simple case.

3.1 Test Cases

Scripts running tests were chosen to represent both small, medium and big datasets that is each framework is tested against 100, 1 000, 10 000 and 50 000 elements. Such a

[2] Ecma International, "ECMAScript® 2015 Language Specification", Proxy Objects http://ecma-international.org, last accessed 2017/07/01.

variety of test cases ensure that both simple and complicated data sets sizes are covered. In terms of testing it is also very important to measure only rendering time itself, not the process of rendering the whole webpage.

3.2 AngularJS 1.4

Angular is a framework following MVC approach and AngularJS's architecture include modules, providers, service, factory, value, controller, scope, directive and others like filter or injector. In sample implementation, we use default attribute directives to initialize an app with ng-app (which makes AngularJS load main module and then initializes all the providers, this includes services and controllers), then appropriate controller is given to an HTML element via ng-controller, within which there is ng-repeat, which handles data present in the model.

This application controller offers adding new "Todo" entries and its implementation is no more complicated than just appending new elements to a list.

Angular works in such a way that it checks all the time and if something changed – it tries to recalculate all expressions and does that if among all watched expressions there are no more changes in value before and after given cycle of main cycle loop and this is clearly visible in the source code, but for the convenience of the reader we provide equivalent simulating framework's implementation:

```
function angularJS_expression_watching_simulation() {
let exprA = function() {return 5;};
let exprB = function() {let c = 0; return () => c++/3|0;}.call();
let exprC = function() {return exprA() + exprB();};

let watchedExprs = [exprA, exprB, exprC]
let cached = [null, null, null];
while(isAppRunning)
    let dirty;
    do {
        dirty = false;
        for(let j = 0; j < watchedExprs.length; ++j) { /* (A) */
            let value = watchedExprs[j]();
            if (value !== cached[j]) {
                cached[j] = value;
                console.info(i+": expr"+['A', 'B', 'C'][j], "updated to", value);
                dirty = true;
                break; /* (A) */
            }
        }
    } while(dirty);
}
}.call()
```

Lines containing exprA, exprB, exprC simulate expressions used in the view layer. Next is caching mechanism and finally loop using dirty-checking for recalculating changes as long as there is some new value. exprA is just a constant expression, exprB is non-linear expression, which value increases every 3rd execution; exprC is a sum of exprA and exprB and cache ought to be updated only if exprA or exprB change value (what is true during every 3rd application cycle run).

This solution is exactly we have discussed in Sect. 2.5 in paragraph about dirty-checking. It has a lot computational overhead and is in the rank of $O(n^2)$ in complexity, but its advantage is that it does not require any attention for the developer point of view, because main loop detects and digests all the changes to the data.

3.3 EmberJS

EmberJS is also a MVC (model-view-controller) framework. In its architecture authors provided following main classes: Application, View, Model, Controller and Router. Application is the top-level structure building the web application and requires root-element of an application, view, controller and a model. View represents them on the webpage, controller takes care of all actions performed by the user and model stores data needed for the view to be rendered. Router on the other hand provides possibility of exchanging views and controllers based on location.hash given in the URL after #.

Ember takes also a different approach on setting it up from AngularJS, because it is convention based, this means that developer has to learn the convention and there is no necessity of explicitly establishing relations between view, controller, model or router as long as they follow EmberJS's convention, which expects to find those objects under 'this.${Applicationname}Controller', 'this.${Applicationname}Route'. Default expected names can be overwritten by special attributes like 'controllerName'. More of this can be found in this documentation entry[3].

EmberJS's observer's implementation requires understanding EmberJS.Observable, EmberJS.ComputerProperty and EmberJS.Evented, but since implementation's basic is simple, we provide following simulation of explaining how EmberJS observer design pattern implementation is done:

[3] Ricardo Mendes, EmberJS Team, "EmberJS Documentation: Getting Started: Core concepts", https://guides.emberjs.com, last accessed 2017/07/01.

```
(function EmberJS_observer_simulation() {
    let exprs = {'A': null, 'B': null, 'C': null};
    function updateExprA() {
        exprs.A = 5;
        exprC();
        return exprs.A;
    }
    let tmp = 0;
    function updateExprB() {
        exprs.B = (tmp++/3)|0;
        updateExprC(); // exprC depends on exprB, thus exprC gets recalculated
        return exprs.B;
    }
    function updateExprC() { exprs.C = exprs.A + exprs.B; return exprs.C; }
    for(let i = 0; i < 10; ++i) console.log("exprB:' + exprB(), ", exprC: " + exprs.C);
})()
```

In code above functions updateExprA, updateExprB, updateExprC reflect values rendered onto the view and their specification is exactly the same like in AngularJS's case and values are cached in a dictionary exprs. On the other hand implementation is different and requires triggering recalculation of exprs.C every time when expressions it depends on change, in this case those are exprs.A and exprs.B.

EmberJS observer design pattern implementation bases not on native getters and setters, but requires using customized setters. This procedure allows ember to explicitly learn what values must be recalculated thanks to Ember.Observable#set implementation[4].

Ember takes an interesting approach in rendering the view – its low level library Backburner limit direct operations on the DOM and does that by merging all changes to the DOM and after each cycle, if there were multiple modifications to the same element – Backburner makes sure that they are applied only once, further explanation can be found in talk given by the author[5].

3.4 Results

Ember, compared to AngularJS, performs almost twice as better, it is visible that the more the elements, the more AngularJS struggles when amount of data increases and this is strongly connected with the difference how AngularJS and EmberJS handle changes to the data. Benchmarking application containing implementations in those two frameworks was deployed to Heroku, it was executed, results were collected and are presented in Table 1.

4 EmberJS Team, "EmberJS Documentation: Ember.Observable", https://emberjs.com.

5 Erik Bryn, "Backburner.js and the Ember Run Loop. 2016, slide 14", http://erikbryn.com, last accessed 2017/07/01.

Table 1. Performance analysis results.

Nr of el.	Rendering time [ms] for AngularJS 1.4	Rendering time [ms] for EmberJS
100	273	323
1 000	793	557
10 000	7 310	2 714
50 000	23 657	11 869

Those results show that it is possible that even simple application provided with lots of data might handle it better or worse. Accordingly to [2] rendering time is very important, because it is only of factors determining user's satisfaction, what directly influences number of users coming back or staying on given webpage. This also influences business potential and gives a change to generate profit. There are also other consequences like comfort of developer while implementing features, but since this criteria is hard to measure, we will not discuss it here. But the easier and user-friendly environment – the bigger the chance that developer will do his or her best to polish the outcome as much as possible.

3.5 Other Approaches

In web development community, there are also other approaches to rendering view and few of them are discussed below.

ShadowDOM was described in Web Components specification[6] and is about encapsulating logic of visible scope of styles and attributes of DOM-element, but not only from the point of view of organization of implementation of given HTML component, but also works on the low-level in the browser, what leads to increased performance. The topic of ShadowDOM is widely discussed in [11].

ReactJS is a library for building user interface which is using virtual-DOM for rendering and updating generated view. ReactJS takes a functional approach, in which view is just a function of state and every change to the state re-renders component and its sub-componenets, puts them in cache, which later virtual-DOM applies differences to the real DOM [12].

VueJS, like ReactJS, uses an intermediate layer before publishing changes to the DOM (similar to EmberJS's Backburner for merging modifications to the DOM), but the place where VueJS really outstands other frameworks is the building the list of dependencies, because it uses recursive traversal of properties by "touching" to build deep list of dependency [13].

[6] World Wide Web Consortium, Hayato Ito, "ShadowDOM", https://www.w3.org, last accessed 2017/07/01.

4 Summary

At first, we have touched a bit of what web pages are, how they are rendered, how JavaScript was changing the web and what, what frameworks are, what is observer design pattern and how it is related to frameworks. We have described and explained how benchmarking environment was designed and then we came up with two example applications written in two frameworks: AngularJS and EmberJS, what gave us fundamental data to compare those two in terms of performance of mentioned design pattern. Later we had shown that it is clearly visible that EmberJS has better performance because of its internal architecture, but at the cost of developer convenience (using custom getters). We hope that data and conclusion presented in this paper can be useful in wide range of problems from choosing technology to designing projects offering presenting data in a product in which information has to be provided in real-time to the user transportation, GPS tracing, visualization software or any other project requiring low-latency for showing results. Online benchmarking tool is publicly available on Heroku[7].

References

1. Everts, T.: Google: 53% of mobile users abandon sites that take longer than 3 seconds to load. http://soasta.com. Accessed 01 July 2017
2. Everts, T.: Time Is Money. The Business Value of Web Performance. O'Reilly Media, Sebastopol (2016)
3. Wijmo Team: Wijmo Event Class's addHandler documentation. http://wijmo.com/5/docs
4. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional, Boston (1994)
5. Czerwinksi, S.: Optimizing AngularJS: 1200 ms to 35 ms. Chapter "Optimization 2". http://blog.scalyr.com/2013/10/angularjs-1200ms-to-35ms. Accessed 01 July 2017
6. Garsiel, T., Irish, P.: How browsers work: behind the scenes of modern web browsers. https://html5rocks.com. Accessed 01 July 2017
7. Dutton, S.: Big boost to DOM performance – WebKit's innerHTML is 240% faster. https://developers.google.com/web/updates/2012/04/Big-boost-to-DOM-performance-WebKit-s-innerHTML-is-240-faster. Accessed 01 July 2017
8. Horie, L.: TodoMVC Performance Benchmarks. https://github.com/lhorie/todomvc-perf-comparison. Accessed 01 July 2017
9. You, E.: TodoMVC Benchmark. https://vuejs.org/perf/. Accessed 01 July 2017
10. Dulisz, D.: Computed properties and dependency tracking. https://monterail.com. Accessed 01 July 2017
11. Glazkov, D.: What the heck is ShadowDOM. https://glazkov.com/2011/01/14/what-the-heck-is-shadow-dom/. Accessed 01 July 2017
12. Banks, A., Porcello, E.: Learning React: Functional Web Development with React and Redux. O'Reilly Media, Sebastopol (2017)
13. Fritz, C.: VueJS Documentation: reactivity in depth/how changes are tracked. https://vuejs.org. Accessed 01 July 2017

[7] http://pwr-js-obs-perf.herokuapp.com/.

Multistage Hammerstein–Wiener System Identification with the Help of Binary Excitation

Marcin Biegański(✉) and Grzegorz Mzyk

Department of Control Systems and Mechatronics, Faculty of Electronics, Wrocław University of Science and Technology, Wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland
{Marcin.Bieganski,Grzegorz.Mzyk}@pwr.edu.pl

Abstract. In the paper, we consider a problem of identification of Hammerstein-Wiener (N-L-N) system. The proposed strategy consists of 4 steps. Firstly, we identify the finite impulse response of the linear block in the presence of random input and random noise using both parametric and non-parametric identification tools (least squares and kernel estimate). Secondly, special binary sequences are generated to recover second nonlinear characteristic. Next, the interactive signal between dynamic linear and second nonlinear block is estimated. Finally, we recover first nonlinear characteristic using standard algorithm for Hammerstein system identification (least squares and singular value decomposition).

Keywords: Hammerstein–Wiener system · Nonlinear system identification
Nonparametric identification · Least squares · Kernel estimates
Binary sequences

1 Introduction

Block-oriented structures [5, 12] play a crucial role in automation and robotics. For the last four decades scientists have been developing more detailed algorithms for system identification, especially for the most known structures – Hammerstein (N-L) and Wiener (L-N) systems. Hammerstein system consists of two fundamental blocks – one static nonlinear characteristic (N) followed by one dynamic linear filter (L). This structure is well described in the literature and considered fully identifiable [6]. Wiener system, on the other hand, consists of the same two blocks, but in the reverse order. Even though the only difference between these two cascade structures is just the reverse order of simple blocks, the latter system has theoretical analysis provided only in some special conditions [7].

The clear majority of methods are devoted to Hammerstein [8, 9] and Wiener [11, 15] cascade structures, but approximation capabilities of these systems are still considered insufficient. The latest research has been made into even more complicated structures such us Hammerstein-Wiener (N-L-N) [1, 3, 14, 17], Wiener-Hammerstein (L-N-L) [13] and block-oriented system with feedback [1, 10].

The article elaborates the Hammerstein–Wiener (N-L-N) structure. The paper is organized as follows. In the first section, identification problem is depictured in detail.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_46

Then, four stages of the identification algorithm are proposed. Firstly, parameter vector of the dynamic linear block is identified with the use of combination of parametric (least squares) and nonparametric (kernel estimate) methods. Secondly, parameters of output nonlinear characteristic are estimated with the use of kernel-based method. Thirdly, special signal is generated in order to evaluate the process between dynamic and second nonlinear blocks. Finally, parameters of input nonlinear characteristic are identified with the use of least squares method as in Hammerstein identification process. Functionality of the method is presented in the following section on the case of polynomial characteristics and three-dimensional finite impulse response filter. Finally, some conclusions are drawn to underline uniqueness and identifiability conditions as well as advantages and disadvantages of the proposed method.

2 Statement of the Problem

Consider a Hammerstein–Wiener (N-L-N) system as a block-oriented structure that consists of one linear dynamic block represented by finite impulse response (FIR) filter surrounded by two nonlinearities. System shown in Fig. 1 is described by the following equation

$$y_k = \eta\left(\sum\nolimits_{j=0}^{q} \gamma_j^* \mu\left(u_{k-j}\right)\right) + z_k, \tag{1}$$

where $\mu(u)$ and $\eta(x)$ are the input and output nonlinear characteristics of the static blocks described by unknown vectors of parameters a^* (see (2)) and b^* (see (3)) of known orders m and n, respectively.

$$\begin{gathered} \mu(u) = \mu(u, a^*) = a_1^* f_1(u) + a_2^* f_2(u) + \cdots + a_m^* f_m(u), \\ a^* = \left(a_1^*, a_2^*, \ldots, a_m^*\right)^T, a^* \in \mathbb{R}^m, \end{gathered} \tag{2}$$

$$\begin{gathered} \eta(x) = \eta(x, b^*) = b_1^* g_1(x) + b_2^* g_2(x) + \cdots + b_n^* g_n(x), \\ b^* = \left(b_1^*, b_2^*, \ldots, b_n^*\right)^T, b^* \in \mathbb{R}^n. \end{gathered} \tag{3}$$

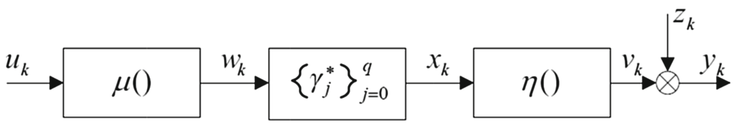


Fig. 1. Hammerstein-Wiener (N-L-N) system.

Moreover $f_i, i = 1, \ldots, m$ and $g_j, j = 1, \ldots, n$ are base functions and their form is *a priori* known.

Dynamic linear block is described by unknown parameter vector γ^*. It is assumed that order q of this vector is given.

$$\gamma^* = \left(\gamma_0^*, \gamma_2^*, \ldots, \gamma_q^*\right)^T, \gamma^* \in \mathbb{R}^{q+1} \tag{4}$$

The system can be described as follows

$$y_k = v_k + z_k, \tag{5}$$

$$v_k = \eta\left(x_k\right), \tag{6}$$

$$x_k = \sum\nolimits_{j=0}^{q} \gamma_j^* w_{k-j}, \tag{7}$$

$$w_k = \mu\left(u_k\right). \tag{8}$$

It is assumed that static nonlinear characteristics are both Lipschitz functions, i.e. are uniformly continuous with limited derivatives. First characteristic is strictly monotonous, whilst signals w_k, x_k and v_k are inaccessible for direct measurement.

System is excited in two separate experiments, with two different types of signals, both independent and identically distributed random variables. $u_k^{(1)}$ is a process with Lipschitz probability density function that has finite variance $\sigma_{u^1}^2$ whereas $u_k^{(2)}$ is a binary signal, i.e.

$$u_k \sim \mathcal{B}(0, 1) \tag{9}$$

with probabilities

$$P\left(u_k = 1\right) = \mathrm{P}\left(u_k = 0\right) = \frac{1}{2}.$$

Moreover, input characteristic $\mu\left(u_k^{(2)}\right)$ is presumed as follows

$$\mu(0) = 0, \mu(1) \neq 0.$$

For the clarity of presentation and without any loss of generality we take $\mu(1) = 1$.

Noise signal z_k is assumed to be i.i.d. symmetrically distributed random variable which is independent from the input signal and has finite variance σ_z^2. Furthermore $E\left\{z_k\right\} = 0$ and $Me\left\{z_k\right\} = 0$.

Additionally, there is one technical assumption: $G^* = \sum_{j=0}^{q} \gamma_j^*$ which is the steady-state gain of the linear dynamic block is constant and equals 1.

3 The Algorithm

The aim of the algorithm is to minimize the following mean squared error criterion

$$Q(\gamma, a, b) = E\left(y_k - \hat{y}_k\right) \to \min_{\gamma,a,b}, \tag{10}$$

where

$$y_k = y_k(\gamma^*, a^*, b^*), \tag{11}$$

$$\hat{y}_k = \hat{y}_k(\gamma, a, b) \tag{12}$$

are the system output dependent on true parameter vectors, and model output dependent on estimated parameter vectors, respectively.

The algorithm consists of the following four stages:

- recovery of the vector γ^* of finite impulse response of dynamic block with the use of the combined parametric-nonparametric estimate (kernel regression and least squares) (Sect. 3.1.),
- generation of special binary sequence with the aim of estimation of second individual nonlinear characteristic, conducted with kernel-based method (Sect. 3.2.),
- generation of additional signal that has the same expected value as interactive x_k process to support identification of input nonlinearity (Sect. 3.3.),
- employing the results from first three stages, parameters a^* are identified with the use of Hammerstein identification method [8] (Sect. 3.4.).

3.1 Identification of Linear Block

Consider the regression vector

$$\phi_k = \left(u_k^{(1)}, u_{k-1}^{(1)}, \ldots, u_{k-q}^{(1)}\right)^T \tag{13}$$

and assume local linearity near the point $u_0 = 0$. Using kernel-censored least squares method the finite impulse response vector γ^* is estimated [2, 13, 14]

$$\tilde{\gamma} = \left(\sum\nolimits_{k=1}^{N} \phi_k \phi_k^T K\left(\frac{\Delta_k}{h}\right)\right)^{-1} \cdot \left(\sum\nolimits_{k=1}^{N} \phi_k y_k K\left(\frac{\Delta_k}{h}\right)\right), \tag{14}$$

where

$$\Delta_k = \|\phi_k\|_\infty = \max_{j=0,1,\ldots,q} \left|u_{k-j}^{(1)}\right| \tag{15}$$

and

$$K\left(\frac{\Delta_k}{\mathrm{h}}\right) = \begin{cases} 1, as\left|\Delta_k\right| \le h \\ 0, elsewhere \end{cases} \tag{16}$$

is the kernel function. h is the bandwidth parameter that must be fit with e.g. cross-validation method [16] in order to minimize influence of bias and variance of this estimate. To fulfil the technical assumption result is normalized:

$$\hat{\gamma} = \frac{\tilde{\gamma}}{\tilde{G}}, \text{ where } \tilde{G} = \sum\nolimits_{j=0}^{q} \tilde{\gamma}_j. \tag{17}$$

3.2 Estimation of Second Nonlinear Static Characteristic

In this stage system is excited with $u_k^{(2)}$ binary signal. Let us introduce the binary representation of number $i - 1$ with d_i vector, $i = 1, 2, \ldots, N_0$, $N_0 = 2^{q+1}$, i.e.

$$\begin{aligned} d_1 &= (0, 0, \ldots, 0, 0, 0)^T, \\ d_2 &= (0, 0, \ldots, 0, 0, 1)^T, \\ d_3 &= (0, 0, \ldots, 0, 1, 0)^T, \\ &\vdots \\ d_{N_0} &= (1, 1, \ldots, 1, 1, 1)^T. \end{aligned}$$

With use of this representation we can form points

$$x_{[i]} = d_i^T \gamma^* \tag{18}$$

that are dependent on both vectors d_i and parameter vector γ^* estimated in the previous stage of the algorithm. As result we obtain the grid of deterministic points

$$x_{[1]}, x_{[2]}, \ldots, x_{[i]}, \ldots, x_{[N_0]} \tag{19}$$

that represent all possibilities of non-measurable signal $x_{[i]}$. In these points, we can estimate output nonlinearity. In the passive experiment, we select input sequences that directly match corresponding vector d_i. The proposed kernel-based estimate of $\eta\left(x_{[i]}\right)$ has the following form

$$\hat{\eta}\left(x_{[i]}\right) = \frac{\sum_{k=1}^{N} y_k \delta\left(\phi_k, d_i\right)}{\sum_{k=1}^{N} \delta\left(\phi_k, d_i\right)}, \tag{20}$$

where

$$\delta\left(\phi_k, d_i\right) = \begin{cases} 1, & as\ \phi_k = d_i \\ 0, & elsewhere \end{cases} \tag{21}$$

is the kernel function. In this estimator, there is no bandwidth parameter because input sequence must exactly match one of the vectors d_i. Denominator in the proposed estimator averages points $\hat{\eta}\left(x_{[i]}\right)$ in local clusters. Probability of the perfect match is constant and equals

$$P\left\{\delta\left(\phi_k, d_i\right) = 1\right\} = \frac{1}{N_0} = \frac{1}{2^{q+1}}. \tag{22}$$

The results of this step are given by the set of N_0 pairs

$$\left\{ \left(x_{[i]}, \hat{\eta}\left(x_{[i]} \right) \right) \right\}_{i=1}^{N_0}. \tag{23}$$

Using this set of pairs, we can easily find the best fitting parameters by the least squares method:

$$\hat{b} = \left(\Psi^T \Psi \right)^{-1} \cdot \Psi^T \zeta, \tag{24}$$

$$\Psi = \left(\psi_1^T, \psi_2^T, \ldots, \psi_{N_0}^T \right)^T, \tag{25}$$

$$\psi_i = \left(g_1\left(x_{[i]} \right), g_2\left(x_{[i]} \right), \ldots, g_n\left(x_{[i]} \right) \right)^T, \tag{26}$$

$$\zeta = \left(\hat{\eta}\left(x_{[1]} \right), \hat{\eta}\left(x_{[2]} \right), \ldots, \hat{\eta}\left(x_{[N_0]} \right) \right)^T. \tag{27}$$

3.3 Estimation of Interactive Signal x_k

Owing to assumption concerning strict monotonicity of the input nonlinear characteristic we obtain reversible function that can be used to recover x_k signal. The proposed approach is shown in Fig. 2.

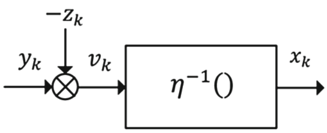


Fig. 2. Reverse flow of output nonlinear block.

Estimation is described by the following equations

$$y_k = \eta\left(x_k \right) + z_k, \tag{28}$$

$$x_k = \eta^{-1}\left(y_k - z_k \right). \tag{29}$$

The fact is, we are not able to measure noise signal, but for every single point $x_{[i]}$ we get the cluster of random values and its variety depends solely on the presence of disturbance process. Probability density function of this process can be estimated based on the value of deviation from the average value $\hat{\eta}\left(x_{[i]} \right)$ e.g. with the kernel-based method. As result it is possible to structure $\varsigma()$ function with whom we can to create additional signal r_k. Estimation shown in Fig. 3 is described as follows.

$$r_k = \varsigma\left(y_k \right) \tag{30}$$

$$\varsigma(y) = E\{x_k | y_k = y\} = \int_{-\infty}^{\infty} \eta^{-1}(y - z) f(z) dz \tag{31}$$

$$\begin{aligned} Er_k &= E\varsigma(y_k) = E \int_{-\infty}^{\infty} \eta^{-1}(y_k - z) f(z) dz \\ &= E \int_{-\infty}^{\infty} \eta^{-1}(\eta(x_k)) f(z) dz = Ex_k \cdot \int_{-\infty}^{\infty} f(z) dz = Ex_k \end{aligned} \tag{32}$$

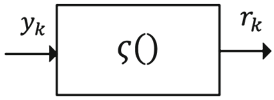


Fig. 3. Estimation of x_k signal.

As result of this step without knowledge of distribution function $f(z)$ we obtained signal r_k that has the same expected value as x_k process.

3.4 Extraction of First Nonlinear Characteristic

In this section, due to multistage approach we can use all previously identified signals and parameters to apply separable least squares method, exactly as in simpler Hammerstein system. As result we can secure parameter vector $\hat{a}$. We do not have to use singular value decomposition as we extracted parameters $\hat{\gamma}$ in Sect. 3.1.

Signal x_k can be described as

$$x_k = \phi_k^T \cdot \theta^*, \tag{33}$$

where

$$\phi_k = \left(f_1(u_k), \ldots, f_m(u_k), f_1(u_{k-1}), \ldots, f_m(u_{k-1}), \ldots, f_1(u_{k-q}), \ldots, f_m(u_{k-q})\right)^T \tag{34}$$

is the regression vector and

$$\theta^* = \left(\gamma_0^* a_1^*, \ldots, \gamma_0^* a_m^*, \gamma_1^* a_1^*, \ldots, \gamma_1^* a_m^*, \ldots, \gamma_q^* a_1^*, \ldots, \gamma_q^* a_m^*\right)^T \tag{35}$$

is the vector of triple products.

After introducing generic vectors $\Phi_N = \left(\phi_1^T, \phi_2^T, \ldots, \phi_{N0}^T\right)^T$ and $X_N = \left(x_1, x_2, \ldots x_N\right)^T$ triple products estimate $\hat{\theta}$ is represented by the following equation

$$\hat{\theta} = \left(\Phi_N^T \Phi_N\right)^{-1} \cdot \Phi_N^T X_N. \tag{36}$$

With parameter vectors $\hat{\theta}$ and $\hat{\gamma}$ we only need to divide corresponding elements to obtain parameters $\hat{a}$.

4 Conclusions

In the paper, a new strategy for Hammerstein-Wiener system identification was proposed. Our idea consists of both parametric (least squares) and nonparametric (kernel estimate) methods of system identification. Moreover, the system is excited with two types of signals to provide specific conditions for every single stage of the algorithm. The idea of using binary sequence to recover parameter vector of output nonlinear characteristic seems to be working even without knowledge about probability density function of the noise signal. This problem of including disturbance process as an input for the subsystem is well known in literature as error-in problem [4]. Dividing problem into four separate stages significantly reduces dimensionality of the system identification. Effectiveness of this method is strictly dependent on the length of finite impulse response of the dynamic block. In consequence, proposed strategy is rather recommended for short memory dynamic filters and more elaborated nonlinearities.

Acknowledgement. The work was supported by the National Science Centre, Poland, Grant No. 2016/21/B/ST7/02284.

References

1. Bai, E.W.: An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. Automatica **34**(3), 333–338 (1998)
2. Bai, E.W.: Non-parametric nonlinear system identification: an asymptotic minimum mean squared error estimator. IEEE Trans. Autom. Control **55**(7), 1615–1626 (2010)
3. Biegański, M.: Kernel-based identification of Hammerstein-Wiener systems. Master's thesis, Faculty of Electronics, Wrocław University of Science and Technology (2017)
4. Chen H.F., Zhao, W.: Recursive identification and Parameter Estimation, pp. 165–288. CRC Press (2014)
5. Giri, F., Bai, E.W.: Block-Oriented Nonlinear System Identification. Lecture Notes in Control and Information Sciences, vol. 404. Springer, London (2010). https://doi.org/10.1007/978-1-84996-513-2
6. Greblicki, W., Pawlak, M.: Nonparametric System Identification. Cambridge University Press, Cambridge (2008)
7. Greblicki, W.: Nonparametric input density-free estimation of the nonlinearity in Wiener systems. IEEE Trans. Inf. Theor. **56**(7), 3575–3580 (2010)
8. Hasiewicz, Z., Mzyk, G.: Combined parametric-nonparametric identification of Hammerstein systems. IEEE Trans. Autom. Control **48**(8), 1370–1376 (2004)
9. Hermansson, A., Syaffie, S.: Hammerstein system identification by non-parametric instrumental variables. Int. J. Control **82**(3), 440–455 (2015)
10. Mzyk, G.: Application of instrumental variable method to the identification of Hammerstein-Wiener systems. In: Proceedings of the 6th International Conference MMAR, pp. 951–956 (2000)
11. Mzyk, G.: A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems. IEEE Trans. Circuits Syst. II Wxpress Briefs **54**(10), 897–901 (2007)
12. Mzyk, G.: Combined parametric-nonparametric Identification of Block Oriented Systems. Lecture Notes in Control and Information Sciences, vol. 454. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-319-03596-3

13. Mzyk, G., Wachel, P.: Kernel-based identification of Wiener-Hammerstein system. Automatica **83**, 275–281 (2017)
14. Mzyk, G., Biegański, M., Kozdraś, B.: Multistage identification of an N-L-N Hammerstein-Wiener system. In: Proceedings of the 22nd International Conference MMAR (2017)
15. Wachel, P., Mzyk, G.: Direct identification of the linear block in Wiener system. Int. J. Adapt. Control Signal Process. **30**(1), 93–105 (2016)
16. Wand, M.P., Jones, H.C.: Kernel Smoothing. Chapman and Hall, London (1995)
17. Wills, A., Schön, T.B., Ljung, L., Ninness, B.: Identification of Hammerstein-Wiener models. Automatica **49**(1), 70–81 (2013)

Towards CLARIN-PL LTC Digital Research Platform for: Depositing, Processing, Analyzing and Visualizing Language Data

Marcin Pol[1], Tomasz Walkowiak[2(✉)], and Maciej Piasecki[1]

[1] Faculty of Computer Science and Management, Wroclaw University of Science and Technology, Wrocław, Poland
marcin.pol@pwr.edu.pl

[2] Faculty of Electronics, Wroclaw University of Science and Technology, Wrocław, Poland
Tomasz.Walkowiak@pwr.edu.pl

Abstract. The paper presents a new functionality of CLARIN-PL Language Technology Centre (LTC). LTC Platform is developed as a research place for processing, visualizing and depositing language data. It can connect and support the research workflow, enabling scientists to increase the efficiency and effectiveness of their research in connection to CLARIN services. The platform is a free and open source web application. Researchers can use it to collaborate, document, archive, share, and register research projects, materials, and data.

Keywords: Natural Language Processing · Language technology infrastructure Web based application · CLARIN

1 Introduction

CLARIN-PL Language Technology Centre has been created as a standard CLARIN B-type centre[1] and the Polish node of the CLARIN research infrastructure. However, as CLARIN-PL follows a bi-directional development model [1] in which the requirements of the users is an driving important factor, we have started gradual expansion of the basic blue-print of the B-type centre towards a platform supporting digital research paradigm in language data analysis. As a result, Language Technology Center Platform (LTCP) was created. It is aimed to support researchers and students in the fields of Humanities, Social Sciences and also Computer Science in work with natural language engineering and text mining. The platform brings researchers into a manageable, secure cloud environment. It is a tool that promotes open, centralized workflows by enabling capturing of different aspects and products of the research lifecycle, including developing a research idea, designing a study, storing and analyzing collected data. It facilitates also publishing resources, tools and reports or papers in Internet with persistent identifiers (DOIs). In this paper we present the present state and foreseen lines of development of LTCP.

[1] https://www.clarin.eu/sites/default/files/CE-2012-0037-centre-types-v07_0.pdf.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_47

2 Related Works

The idea of a platform supporting research can be visible in several systems built inside and outside CLARIN, e.g.:

1. WebLicht [2] is a tool implemented by CLARIN-D for setting up natural language processing pipelines by SS&H users. It supports users by GUI in creating and executing chains of web services for different simple tasks in text analysis.
2. Open Science Foundation is a complex system which integrates many open systems and supports researchers in managing research projects (e.g. storing, organising experiments, archiving results, etc.).
3. The CLARIN Language Resource Switchboard (LRS) is an application accessible by web page that gives functionality similar to WebLicht. However it is aimed at linking language data with the tools in semi-automatic way prior to processing in pipelines powered by WebLicht.
4. Treex::Web [3] is a web application offering ready-to-use chains of web services tools. Users can also create their own chains. The results can be visualised or directly stored.
5. Multiservice [4] combines several mature offline linguistic tools in a common online platform.

None of the above is a complete, open source Natural Language Processing (NLP) platform with upload, process, annotation and visualization applications. Moreover, existing NLP processing solutions have several limitations. First of all, these tools allow to process only serial chains of tools, which is not enough for more complex text mining tasks [5] and lack machine learning (ML) tools. In addition, they use very exhaustive formats and data in the intermediate steps of processing pipelines have to go between the central system and a webservice (for example deployed in other data center). The experiments showed that the performance overhead for such data transfer is up to 500%. Additionally, existing solutions have limited scalability. LTCP is intended to alleviate these problems and limitations.

3 Users, Requirements and Functionality

3.1 Starting Point

Researchers can use LTCP to manage projects and collaborations, register their intermediate results their work and, during the later phase of a project, to deposit the official outcomes of the project. Repositories support mainly the final, depositing phase. Our platform support data and materials depositing and archiving sharing by Lindat version of D-Space[2] or their sharing in the earlier phases by NextCloud[3]. Both applications were modified and expanded with new functionalities.

LTCP will help researchers during the whole project lifecycle. The platform brings research into a secure cloud environment which is manageable by the project researchers. It is developed and maintained by the CLARIN-PL LTC, builds and supports scientific research communities, and develops research tools and infrastructure to enable managing and archiving research.

The core functionality of the LTCP is the possibility to create and develop NLP projects connected to CLARIN tools and applications. Users will be able to set up a project for a particular paper or specific experiment or for the work of an entire lab. To create a project, users have to set up a free account with the D-Space (they can login via federation identity using shibboleth[4]). Once logged in to the D-Space, users can upload files to NextCloud or D-Space. The latter is used to make projects/resources publicly available, while NextCloud was added in response to the users' demands (and hesitation in too haste using of D-Space) for private or working projects/resources. In the case of depositing in D-Space CMDI meta-data are required, in NextCloud strongly suggested (e.g. as a source of data for supervised training or experiment management). In case of D-Space usage, unique, persistent IDs., that does not need to be the case in NextCloud storage. Moreover, CLARIN-PL D-Space is connected to the unique long-term persistent archiving system developed by one of the CLARIN-PL partners. NextCloud supports controlled access, so project members can be assigned different permissions: read only, read and write, and administrator.

While the spirit of open science encourages making projects publicly available, NextCloud provides an option to make all or parts of a project public and export data to the D-Space. In general, private projects are not browsable. All exported public projects are visible by CLARIN Virtual Language Observatory[5]. In addition to using a unique, persistent URLs, DOIs, LTCP enables sharing in a variety of additional ways. A promoted one is on the basis of a license selected from the CLARIN Licence Category Calculator[6] (D-Space offers direct connection to the calculator). LTCP provides practically unlimited storage for projects (at least concerning SS&H projects), but individual files are limited up to 1.5 gigabytes each by the NextCloud constraints.

It will be possible to efficiently search linguistically preprocessed text resources/ corpora via Kontext (or a similar system), and they can be already upload to the web-based annotation editor Inforex offering rich functionality for group working and annotation project management. Extraction of various statistics from users' text data sets can

[2] https://github.com/ufal/clarin-dspace, it can be used also for storing intermediate project results, but this results in their wide visibility that can be considered as some scientific noise, and is often not welcomed by the researchers.

[3] https://nextcloud.com/.

[4] https://shibboleth.net/.

[5] https://vlo.clarin.eu/.

[6] https://www.clarin.eu/content/clarin-license-category-calculator.

be performed by LEM[7] system, as well as extraction of collocations and terminology can be done with the help of MeWeX[8] system (it also supports manual editing the extracted collocations as a MWE lexicon). Moreover, users can also directly use WebSty for analysing similarities and clustering of texts in Polish or English. Subsequent paragraphs, however, are indented.

3.2 Web Services and Predefined Pipelines

Natural Language Processing and Machine Learning tools are now widely available. However, they are developed in different technologies (like Python, Java, R, C++) and use variety of different formats of data that are very often not compatible with each other. Many language tools, especially for languages different than English, are hard to be installed without skills in programming. They have usually a large number of different parameters to be set-up. Interpretation of the parameters requires mostly knowledge from Computational Linguistics. Therefore, their integration and building its own processing workflow is not a simple task, especially for researchers without an experience in computer engineering. CLARIN proposes making linguistic tools available in Internet (e.g. as Web Services and simple Web Applications) and developing research web based applications [6], as means to diminish the above mentioned problems. Thus, users can process data online and do not need to bother about technicalities, but still need to understand processing mechanisms on the level of their presentation.

However, building a multi-user, web system generates a next set of problems connected with the system availability and performance. The system should be scalable, responsive and available all the time. Language tools have excessive CPU/memory consumption and a number of users or tasks at a given time is unpredictable.

Existing solutions, such as multilingual WebLicht [2] and Multiservice [4] for Polish, have several limitations. First of all, these tools allow to process only serial chain of tools, which is not enough for more complex text mining tasks [5] and lack machine learning tools. Moreover, they use very executive formats and the data in the intermediate steps of processing pipelines have to go between central system and the webservice (for example deployed in other data center). The experiments [7] showed that the performance overhead for such data transfer is upto 500%. Additionally, existing solutions have very poor scalability. That's why authors developed their own solution as presented in Sect. 4.

3.3 Searching of the Linguistically Preprocessed Text Corpora

LTCP can help during the research lifecycle. The platform brings research into a manageable, secure cloud environment. The Web-based project is a tool that promotes open, centralized workflows by enabling capture of different aspects and products of the

7 LEM: http://ws.clarin-pl.eu/lem.shtml – offering extraction of statistical description of text corpora, comparison of text and corpora, as well as a connection to the expanded WebSty system for stylometry.

8 MeWeX: https://mewex.clarin-pl.eu/.

research lifecycle, including developing a research idea, designing a study, storing and analyzing collected data. It provides possibilities of writing and publishing reports or papers with Persistent IDs. It is developed and maintained by the CLARIN-PL LTC, builds and supports scientific research communities, and develops research tools and infrastructure to enable managing and archiving research. The NSF provides and supports a variety of NLP tools and services to assist in the research process.

The core functionality of the LTCP is the possibility to create and develop NLP projects. Very simply, a project functions as a workspace, with the design of a particular project depending on users and the type of research workflow that they are trying to manage and preserve. Users might wish to set up a project for a particular paper or specific experiment or for the work of an entire lab. To create a project, users must set up a free account with the D-Space (they can login via federation identity using shibboleth[9]). Once logged in to the D-Space, users can upload files to D-Space (making projects publicly available) or NextCloud (private projects), add meta data (CMDI format). Due to using D-Space the uploaded files receive a unique, persistent id (DOIs)[10]. CLARIN-PL D-Space is connected to the prototype of a long-term persistent archiving system.

LTCP together with NextCloud built into it is intended to be collaborative, and users can easily add contributors to projects. The NextCloud supports controlled access, so project members can be assigned different permissions: read only, read and write, and administrator.

While the spirit of open science encourages making projects publicly available, NextCloud has implemented an option to make all or parts of a project public and export to the D-Space. In general, private projects are not browsable. Exported public projects are visible by CLARIN Virtual Language Observatory[11]. In addition to using unique, persistent URLs, DOIs, the NSF promotes sharing in a variety of additional ways. A primary one is the option to add a license (D-Space with connection to the Clarin Licence Category Calculator[12]). The Platform provides unlimited storage for projects, but individual files are limited to 1.5 gigabytes (GB) each (using NextColoud).

4 Architecture

4.1 Overview

LTCP software architecture consists of three layers: web applications, repositories and core services. Web applications are aimed to communicate with users to perform given set of tasks in LTCP (Fig. 1). Web applications could be developed in two different styles as Single Page Applications that communicate with other parts of LTCP by REST services or as multitier applications where the server side mediates communication with LTCP components. For example: WebSty and LEM are SPAs whereas Inforex and

9 https://shibboleth.net/.
10 http://handle.net/.
11 https://vlo.clarin.eu/.
12 https://www.clarin.eu/content/clarin-license-category-calculator.

Mewex are 3-tier applications developed in PHP. The second layer consists of two repositories: D-Space and NextCloud. They allow to store corpora and results of processing. Additionally D-Space repository is an authorization manager, which gives users possibilities to use all applications from the first layer with one login and password. The third tier delivers core services of LTCP. It includes single authorization for the LTCP, access to pipelines of language and machine learning tools (NLPServices) and a monitoring module. The aim of the monitoring module is to control the LTCP state and allow a fast reaction on faults or system overloads. Monitoring will be performed a full form on different levels, starting from hardware, through virtual machines, language and machine learning tools up to user web applications.

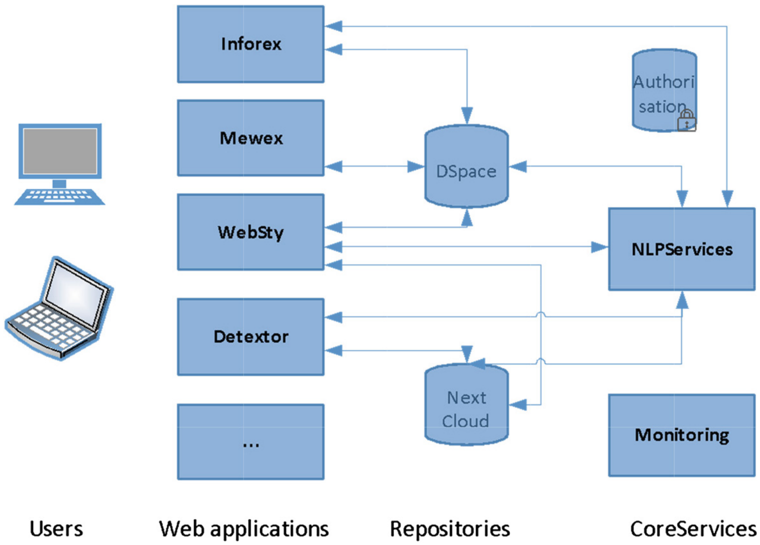


Fig. 1. LTCP architecture.

4.2 Private Cloud

The LTCP is deployed in a private cloud. Hardware consists of nine servers in a mixed rack/blade architecture. Each server has from 192 to 224 GB of RAM, which gives a total of almost 2 TB. Each server has two Intel (R) Xeon (R) CPUs E5-2665@2.40 GHz, which let you run up to 16 threads per processor. In total, it gives power of 324 processes in parallel. Data storage subsystem is built on IBM Storwize V7000 with redundant dual-active intelligent FC 8 Gb controllers and dual-active iSCSI controllers. Storage is using RAID10 volumes. Data is protected by backup with deduplication mode. System is configured to create complete data snapshot every Sunday. All system is protected by UPS [8].

Servers are managed by XENServer that allows to run and manage a large number of virtual machines. Virtualization makes the LTCP management more convenient and efficient. The resources (memory, CPU, disk) could be attached to any machine on demand

and changed according to needs. Operating systems are independent from the hardware in the virtual environment so they can be easily moved to another server as a reaction to any failure or resource shortage. Moreover, virtualization provides a disaster recovery mechanism ensuring that when a virtualized system crashes, it will be restored as quickly as possible.

4.3 Single Authorization

LTC authorization is accomplished through the private D-Space federated login. It is done by generating a random string of 129 characters token, using a cryptographic generator. Then the token is assigned to the authorization process and it is stored in the user's D-Space database. To allow Clarin applications to use the token, it is placed in the HTTP cookie named "clarin-en-token". It is available throughout the wildcard clarin-pl domain. When user is running a federated application such as NextCloud, the presence of the previously mentioned cookie is checked. If the cookie is in the browser, verification is performed by calling D-Space microservice. It is checking if the token is associated with the logged in user. D-Space returns to the application the user name and then logs into the application that requests the login. In the case of an error in the verification or absence of the cookie, the d-space will be redirected user. The scheme is depicted in Fig. 2.

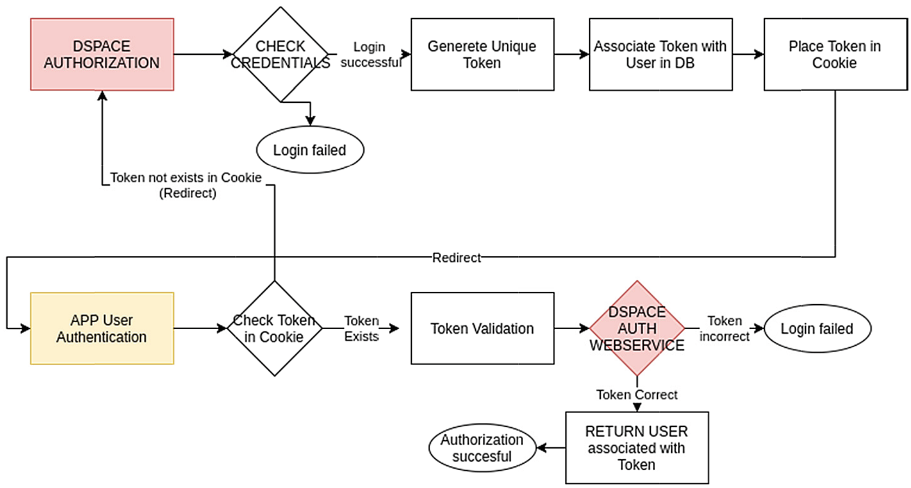


Fig. 2. Authorization schema.

4.4 NLP Services

The processing of texts by a chain of NLP and ML tools is done by NLP Services which architecture is presented in Fig. 3. Each of tool is run as a separate microservice [9]. It allows to keep in memory large knowledge models. Communication between microservices is done by a queening system. In the implementation, we have used the

AMQP[13] protocol for lightweight communication mechanisms and open source RabbitMQ[14] broker for a queuing system. AMQP protocol has clients for a large number of different software platforms as required by technologies used by language and ML tools. Each NLP microservice collects tasks from a given queue and sends back messages when results are available.

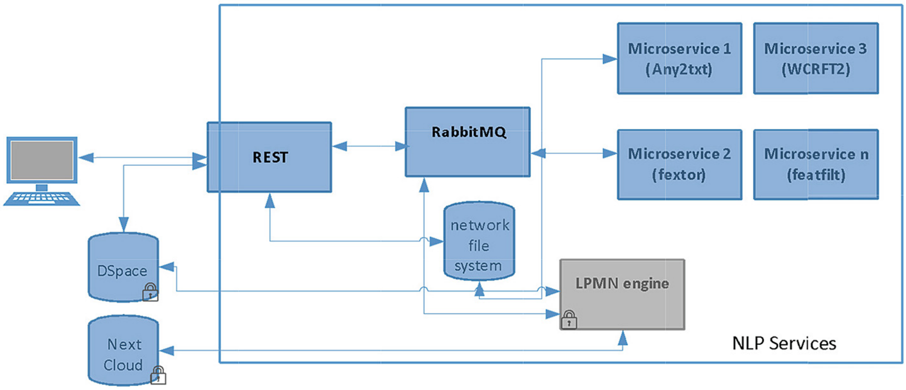


Fig. 3. NLP Services architecture.

In the proposed architecture (Fig. 2) additional server grants the access from the Internet. It works as a proxy for the LTCP delivering REST API for other system components (for example web applications). Such an approach allows for easy integration with almost any kind of applications. The exchange of data between microservices, i.e. inputs to NLP tools and results of their processing, is done by a network file system. It makes integration of NLP tools easier since they are mostly designed in the manner that they expects a file as in input and produces files as an output.

Almost each (more frequently used) NLP microservice is deployed on a separate virtual machine. Therefore, it is easy to scale up the system just by duplicating a virtual machine as a reaction to a high number of requests for a given microservice [5].

4.5 Language Processing Modelling Notation

Having language tools implemented as microservices there is a need to describe cooperation of them to realize specific tasks. Therefore, we have developed [5] a human readable orchestration [10] language, that allows to describe different text processing tasks. It was called Language Processing Modelling Notation (LPMN).

The engine for running workflows described in LPMN was developed and is a core element of NLP Services (Fig. 3). It allows to process large corpus of text in a batch like mode.

The LPMN statement for DSpace preprocessing pipeline is presented in Fig. 4.

13 https://www.amqp.org/.
14 https://www.rabbitmq.com.

```
urls([{“name”:’Cyrulik.ccl”,“url”:“https://clarin-
pl.eu/rest/bitstreams/23849/retrieve”},
      {“name”:“1918.ccl”, “url”:“https://clarin-pl.eu/rest/bitstreams/23860/retrieve”},
      {“name”:“1932.ccl” “url”:“https://clarin-pl.eu/rest/bitstreams/23864/retrieve”}])
|any2txt|wcrft2({“morfeusz2”:false})|liner2|wsd|dir|makezip|todspace(/1111/2222/)
```

Fig. 4. Exemplar LPMN for preprocessing pipeline.

5 Conclusions and Further Development

CLARIN-PL LTC Platform (LTCP) focuses on NLP research tools. It gives openness, unique identifiers and research data management. The high degree of flexibility means that it will be possible to easily customize projects to fit a variety of needs, from small ones to large research collaborations. SS&H researchers can process large corpuses of text and easily publish or share results. LTCP brings easy deposit and sharing functionality like Open Science Framework with “WebLicht style” webservices processing and exporting results to web annotation application or search tools. The main reason in building LTCP is in offering SS&H researchers a more direct way of using CLARIN services in all phases of their projects, a kind of CLARIN specific open science framework (that does not exclude the use of third party solutions in any moment, but depositing). We plan to fully implement LTCP during this year and next expand its functionality to connect more tools as microservices and applications like for example Mtas[15]. So far, the platform is focused mainly on Polish, but most of the tools are able to work with English and work on German is in progress, and will continue expanding LTCP to support better multilinguality.

Work financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education

References

1. Piasecki, M.: User-driven language technology infrastructure – the case of CLARIN-PL. In: Proceedings of the Ninth Language Technologies Conference, Information Society – IS 2014, Ljubljana, Slovenia (2014). http://nl.ijs.si/isjt14/proceedings/isjt2014_01.pdf
2. Hinrichs, E., Hinrichs, M., Zastrow, T.: WebLicht: web-based LRT services for german. In: Proceedings of the ACL 2010 System Demonstrations, ACL Demos 2010, pp. 25–29, Stroudsburg, PA, USA. Association for Computational Linguistics (2010)
3. Sedlák, M.: Treex::web. In: LINDAT/Clarin 3.2. (2014). http://hdl.handle.net/11858/00-097C-0000-0023-44AFC
4. Ogrodniczuk, M., Lenart, M.: A multi-purpose online toolset for NLP applications. In: LNCS, vol. 7934, pp. 392–395. Springer, Heidelberg (2013)
5. Walkowiak, T.: Language processing modelling notation – orchestration of NLP microservices. In: Advances in Dependability Engineering of Complex Systems: Proceedings of the Twelfth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, 2–6 July 2017, Brunów, Poland, pp. 464–473. Springer (2018)

[15] http://www.ep.liu.se/ecp/136/002/ecp17136002.pdf.

6. Wittenburg, P., et al.: Resource and service centres as the backbone for a sustainable service infrastructure. In: Calzolari, N., et al. (eds.) Proceedings of LREC 2010, Malta, pp. 60–63. ELRA (2010)
7. Walkowiak T.: Web based engine for processing and clustering of Polish texts. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) Theory and Engineering of Complex Systems and Dependability. Advances in Intelligent Systems and Computing, vol. 365. Springer, pp. 515–522 (2015)
8. Walkowiak, T., Pol, M.: The impact of administrator working hours on the reliability of the Centre of Language Technology. J. Pol. Saf. Reliab. Assoc. **8**(1), 167–173 (2017)
9. Dragoni, N., Giallorenzo, S., Lluch-Lafuente, A., Mazzara, M., Montesi, F., Mustafin, R., Safina, L.: Microservices: yesterday, today, and tomorrow, CoRR, vol. abs/1606.04036 (2016)
10. Peltz, C.: Web services orchestration and choreography. Computer **36**(10), 46–52 (2013)

Sustainable Transport Interchange

Assessing the Design and Operation of Riga's International Coach Terminal

Maria Tsami[1], Evelina Budilovich (Budiloviča)[2(✉)], Vissarion Magginas[1], Giannis Adamos[1], and Irina Yatskiv (Jackiva)[2]

[1] Traffic, Transportation and Logistics Laboratory, University of Thessaly, Pedion Areos, 38334 Volos, Greece
[2] Transport and Telecommunication Institute, Lomonosova Street 1, Riga 1019, Latvia
evelina.budilovich@gmail.com

Abstract. This paper aims at identifying the performance level of Riga's International Coach Terminal, by assessing its design and operation. Based on an extended state-of-the-art review of best practices, a list of representative indicators, grouped into eight groups, were used and evaluated by terminal's users. These groups deal with mobility provision, way-finding information, time and movement issues in the terminal, accessibility, comfort, station image and attractiveness, safety and security, and handling of emergency situations. The objectives of the research are to: (a) point out the level of user's satisfaction from the current terminal operation, infrastructure and services, and (b) correlate the above attributes with the overall terminal assessment. For the data collection, a face-to-face and internet-based questionnaire survey was conducted, with users stating their perceptions and level of satisfaction, related to the terminal infrastructure, operation and services, as classified in the aforementioned eight groups of indicators. A decision-tree approach was applied to indicate the key performance indicators in users' assessment formulation for the case study. Research findings reveal the most significant parameters that need to be modified in order to increase users' satisfaction, which will gradually increase the overall image and attractiveness of the terminal and the usage of its services.

Keywords: Urban transport · Interchanges · Design · Operation Decision tree

1 Introduction

In recent years, there seem to be an increasing interest towards the enhancement of sustainable urban mobility. This is evident, as more and more EU policy is being formulated, with the most important relevant act being the 2013 Urban Mobility Package, which consisted of a number of initiatives, action plans and legislation documents [1] and continued the work that had begun with the 2011 White Paper towards a single European Transport Area [2] and its predecessor, the 2001 White

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_48

Paper on European Transport Policy [3]. One of the main goals of a Sustainable Urban Mobility Plan (SUMP), as it is defined in the aforementioned Mobility Package, is the reduction of private car use and its substitution by public transport means. In order for this to be possible, it is necessary to provide passengers the opportunity to have a seamless journey while using a multitude of transport means [1]. Transport interchanges have been identified as one of the most important components of intermodal transport. These facilities allow seamless travel to be carried out more easily [4]. As a result, transport interchanges help towards increasing public transport use [5–7].

Lately, there has been a fair amount of research on the design and assessment of intermodal transport interchanges [4, 8–10]. The most recent, comprehensive approach to the interchange design and operation problem came as part of the EU funded City-HUB project, which aimed at the formulation of a comprehensive business model and a set of guidelines for these interchanges, this way promoting public transport and by extension sustainable urban mobility [11].

The aim of this paper is the assessment of the design and operation of the Riga's International Coach Terminal (RICT). The terminal was assessed based on a number of indicators, which were divided into eight groups: travel information, way-finding information, time and movement issues, access to and from the station, comfort, station attractiveness, safety and security and emergency situation handling. The categories and indicators that the respondents were asked to rate were chosen based on the findings of a state-of-the-art review.

2 State-of-the-Art Review

Regarding safety, all necessary measures and precautions should be taken, in order to ensure passengers' well-being. This includes areas at the interchange's surroundings where a possible accident involving passengers and vehicles could happen, as well as the minimization of the possibility of an accident happening inside the station. This goal can be achieved by implementing traffic control measures where necessary, while forcing a distance between passengers and vehicles could also be beneficial [12]. At the same time, the feeling of security has become even more relevant nowadays due to the higher number of terrorism incidents and is closely related to psychological factors, so well-lit open spaces and the avoidance of isolated areas should work positively on the passengers' perception of it [13]. Another factor that could contribute towards this goal is the presence of specialized staff offering this particular service [14].

Safety and security is closely related to emergency situation handling. An interchange has to abide by all safety standards and have a clear plan to deal with these situations. An adequate number of emergency exits; lights and fire alarms combined with the existence of a well thought out evacuation plan are deemed necessary [11].

The feeling of comfort is dependent upon factors like the number of available seats and how sheltered an interchange is [13]. Similarly, the general condition of the interchange (cleanliness, temperature, noise) also affects a passenger's sense of comfort or discomfort [12]. Another element affecting comfort is the waiting time utilization opportunities, which can also contribute, by providing passengers the choice to accommodate other needs or wishes (for example shopping or eating) during an intermodal trip [13]. Amenities that are offered at the interchange can also help create comfortable conditions during an individual's travel. The number of automated teller machines (ATMs), Wi-Fi access and information screens are important parameters to consider when it comes to passenger comfort [12].

Attractive interchange design is closely related to comfort and can contribute towards creating a positive mood to passengers by providing added aesthetic value to the facilities and by creating a pleasant ambiance [11].

Another critical factor is the total time needed for a complete trip with the use of intermodal public transport [14]. The more time a passenger needs to reach his/her destination, the more reluctant he/she is to switch from his/her private vehicle to public transport means [6, 15–17]. This leads to the conclusion that better coordination between the different transport modes combined with reduced distances that a passenger needs to walk to facilitate the transfer between modes or reach his desired facility, will alleviate the problem. The problem of movement inside the station is also closely associated with the time issue. Overcrowding can cause discomfort to the users of the facilities, while also hindering their movements inside the interchange and can result in reduced movement speed and even missed connections [13]. Another movement issue in an interchange is the transfer between different levels, the number of which should be kept to a minimum. If, however, the existence of more than one level is unavoidable, they should be accessible through an adequate number of lifts and/or escalators [12].

The area of ticketing and information provision is also a critical one [16]. Ticketing is recognized to have a significant effect on the use of public transport [4, 5]. Integrated ticketing has the potential to raise the use of public transport by making an intermodal movement easier and faster [18]. While on a multimodal trip, the availability of accurate and in time information is imperative in order to have a seamless journey. Moreover, accurate information can help travelers use their time more efficiently, as well as give them a general sense of the interchange's facilities, which is especially helpful to people with disabilities, since it can influence their choice of transport modes [13].

Another important element for a successful transport interchange is the ease of access to and from the station. To make the interchange easy to access, the access/egress zones should be composed of multiple routes which provide safe passage for pedestrians and cyclists, as well as facilities for disabled passengers and those who choose to access the interchange by use of motorized vehicles [12].

Way-finding information is also considered very important in the design of a transport interchange. With the efficient use of navigation aids it becomes possible for non-frequent users of the interchange to find their way through the facilities more easily. Signing, in particular, is especially helpful to new users of the interchange, and helps them transfer between the different modes, find the services they are looking for [12].

3 Method

3.1 Survey Set-up and Data Collection

Data was collected through a questionnaire survey conducted both online and face-to-face. The questionnaire was divided into three sections [11]. The first one consisted of general trip information questions, such as origin and destination, trip stage, travel purpose, trip duration and means of transport used to travel to and from the terminal. The second section was dedicated to passenger satisfaction questions. In this part each passenger was required to assess a number of indicators for each of the eight groups. Passengers were asked to rate each indicator in a Likert scale, with 1 being the lowest possible score and 5 being the highest. In this section, passengers also chose the three most important aspects of an interchange from a list of eight items. These items were information, waiting areas, safety and security, services, shops and cafes, transport communication between different modes, access to the interchange and other (in which case the respondents were asked to specify their answer). The last section contained questions aiming at personal information, such as gender, age, education level, employment status and net-income per month. For the purpose of this particular study the questionnaire was also translated into Latvian and Russian.

3.2 Decision Tree Approach

A decision tree was used to model how the performance evaluation of the selected indicators affects the overall satisfaction level of the terminals. Decision trees are commonly applied for decision analysis, facilitating the identification of a strategy most likely to reach a goal [19–21]. In this case, the goal is to have satisfied users and therefore it is crucial to indicate all these elements that lead to a high level of overall satisfaction. In the framework of this study, the Weka J48 tree was used, as an open Java implementation of the C4.5 algorithm [19, 20]. C4.5 is a well known, supervised learning algorithm, used for data mining and classification problems in machine learning. The algorithm (C4.5) learns a mapping from attribute values to classes, applied to classify new, unseen instances. Beginning with the root node that represents the entire dataset the algorithm splits data into smaller subsets, that denote the partitions of the original dataset that satisfy specified attribute value tests. This process continues until all instances in the subset fall in the same class and therefore the tree growing is terminated.

In total 37 attributes (Table 1) were analyzed using the J48 classifier tree with the pruning values of 0.25 and as test mode the evaluation on training data. The developed tree had 51 nodes and 26 leaves (end nodes).

Table 1. List of attributes and coding.

Attribute	Code
Availability and ease of use of travel information at the terminal	A1
Availability of travel information (timetables, routes, delays) before your trip	A2
Accuracy and reliability of travel information displays for bus/trains at the terminal	A3
Ticket purchase (ticket offices, ticket machines, etc.)	A4
Signposting to different facilities and services	B1
Signposting to transfer between transport modes in all parts of the terminal	B2
Information and assistance provided by staff	B3
Transfer distances between different transports modes	C1
Co-ordination between different transport operators or transport services	C2
Use of your time (transferring & waiting) at the terminal	C3
Distance between the facilities and services	C4
Ease of movement due to number of people inside the terminal	C5
Ease of access to the terminal	D1
Ease of access from the terminal	D2
General cleanliness of the terminal	E1
Temperature, shelter from rain and wind, ventilation, air conditioning	E2
General level of noise of the terminal	E3
Air quality, pollution (e.g. emissions from vehicles)	E4
Number and variety of shops	E5
Number and variety of coffee-shops and restaurants	E6
Availability of cash machines	E7
Availability of seating	E8
Availability of mobile phone signal and Wi-Fi	E9
Comfort due to the presence of information screens	E10
The surrounding area is pleasant	F1
The internal design of the terminal (visual appearance, attractiveness, etc.)	F2
The external design of the terminal (visual appearance, attractiveness, etc.)	F3
Safety getting on and off the transport mode (train, bus, taxi, bicycle, etc.)	G1
Safety whilst inside the terminal	G2
Feeling secure in the transfer & waiting areas (during the day)	G3
Feeling secure in the transfer & waiting areas (during the evening/night)	G4
Feeling secure in the surrounding area of the terminal	G5
Lighting	G6
Information to improve your sense of security	H1
Signposting to emergency exits	H2
Location of emergency exits in case of fire	H3
Overall score of user satisfaction with the service in the terminal	I

4 Analysis and Results

4.1 Sample Characteristics

The survey, organized by the Transport and Telecommunication Institute in cooperation with Riga's International Coach Terminal, was realized in Spring 2017, including a pilot implementation (17% of total sample), a face-to-face survey (45% of total sample), an online survey (15% of total sample) and a focus group (students) survey (23% of total sample). Achieving a response rate of 95%, the final sample was determined to 239 users.

The 62% of the respondents are women and the rest 38% men. The 35% of the users are between 18–25 years old, the 30% of them between 41–65, the 28% between 26–40, the 3% of them younger than 17, the 3% older than 66 years old, and the rest 1% preferred not to answer this question.

Regarding the education level of the respondents, it was observed that the majority of them (55%) is highly educated, the 24% has received secondary level of education, the 15% holds a secondary professional level diploma, and the rest 6% are primarily educated.

In addition, the majority of respondents (41%) live in households with 1–2 people, the 22% in households with 3 people and the rest 37% in households with more than four people. Focusing on the employment status of respondents, it was indicated that the 64% of them are employed, the 24% are students, the 3% are unemployed and the rest 9% respondents stated a different status.

Lastly, regarding the monthly net-income of respondents, the 28% of them have income 200–499 EUR, the 27% 500–799 EUR, the 28% greater than 800 EUR, and the rest 17% lower than 200 EUR.

4.2 Results

The developed decision tree is depicted in Fig. 1. Tables 2 and 3, present the summary of results and the detailed accuracy by class, respectively.

Based on the tree results, the tree developed is highly accurate (almost 90% correctly classified instances). Results showed that there are eleven paths leading to a perception of satisfaction and two paths leading to the best possible satisfaction. These two paths are the following:

- "The surrounding area is pleasant" >3, "Number and variety of coffee-shops and restaurants" >2, "The external design of the interchange (visual appearance, attractiveness, etc.)" >2, "Temperature, shelter from rain and wind, ventilation, air conditioning" >4, "Ease of movement due to number of people inside the interchange" >4, "The surrounding area is pleasant" >4, "Information and assistance provided by staff (e.g. at customer information points)" >3.
- "The surrounding area is pleasant" >3, "Number and variety of coffee-shops and restaurants" >2, "The external design of the interchange (visual appearance, attractiveness, etc.)" >2, "Temperature, shelter from rain and wind, ventilation, air conditioning" >4, "Ticket purchase (ticket offices, machines etc.)" ≤ 3.

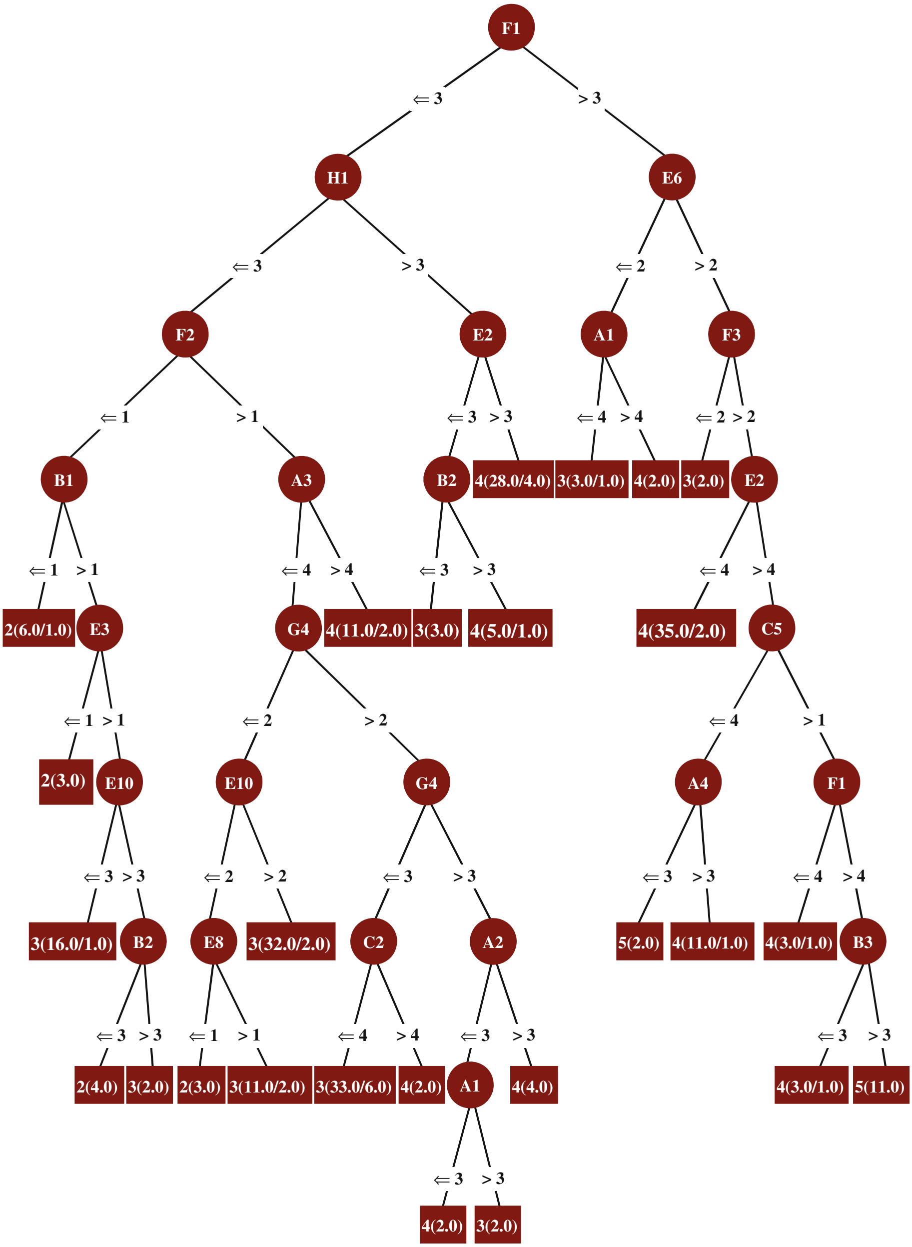


Fig. 1. Decision tree development.

Table 2. Overviews of results.

Parameter	Value
Correctly classified instances	89.5397% (214)
Incorrectly classified instances	10.4603% (25)
Kappa statistic	0.83333
Mean absolute value	0.0708
Root mean squared error	0.1882
Relative absolute error	27.294%
Root relative squared error	52.3692%
Total number of instances	239

Table 3. Detailed accuracy by class.

	TP rate	FP rate	Precision	Recall	F-measure	MCC	ROC area	PRC area	Class
	0.000	0.000	0.000	0.000	0.000	0.000	0.989	0.167	1
	0.714	0.005	0.938	0.714	0.811	0.804	0.967	0.839	2
	0.958	0.084	0.885	0.958	0.920	0.865	0.967	0.921	3
	0.931	0.087	0.887	0.931	0.908	0.839	0.961	0.923	4
	0.650	0.000	1.000	0.650	0.788	0.794	0.958	0.805	5
Weighted average	0.895	0.071	0.896	0.895	0.891	0.839	0.964	0.902	

5 Conclusion

The goal of this paper was to capture the viewpoint and preferences of travelers on different aspects that define the design and operation of sustainable urban transport interchanges. Towards this direction, an on-site face to face and online questionnaire survey were conducted at the Riga International Coach Terminal in Latvia, and useful feedback was gathered from 239 users about their habits, preferences and satisfaction.

For the analysis of the data, a decision tree approach was selected. In this case, due to the nature of the data, a classification tree was developed, using one of the most prevalent relevant algorithms, namely C4.5. The number and kind of the attributes classified using the classification tree were chosen in accordance with the questions included in the face-to-face and online survey questionnaires. Due to sample size and the high number of attributes, the tree's leaf size was relatively small. This comes as a result of the complex nature of a transport interchange, since to fully evaluate a user's satisfaction, one has to test the perceived quality of every facility and offered service in the interchange. Moreover, the high number of splits leads to more accurate predictions of passenger satisfaction.

This methodology works towards helping the decision makers understand the users' perspectives; predict the most important factors contributing to their satisfaction, by studying the developed decision tree, and act accordingly to accommodate their needs.

Interpreting the decision tree development, it can be concluded that the surrounding area atmosphere is the most important indicator for users (parent node of the tree). Both paths that have the highest evaluation scores deal with indicators that improve the general perception of quality in the interchange (number and variety of coffee-shops and restaurants, external design of the interchange (visual appearance, attractiveness, etc.), temperature, shelter from rain and wind, ventilation, air conditioning). The main difference between the two paths is that the first includes more indicators and is even more focused on the perceived hospitability of the interchange facilities and the quality of offered services.

The tree paths are indicative of the indicators that need higher attention, from the decision makers' point of view, in order to increase the overall perceived quality and usage of the interchange. An improvement policy based on these conclusions could be implemented in order to achieve this goal and tackle issues that according to the users interfere with their general satisfaction with the station. Future research can address the issue of improvement priority between the attributes of the two paths leading to the highest satisfaction rate.

Acknowledgements. This paper is based on the research and work that has been conducted in the framework of the ALLIANCE project (http://alliance-project.eu/), which has received funding from the European Union's Horizon 2020 Research and Innovation Programme. The authors would like to thank both the consortium of the project and the European Commission.

References

1. European Commission: Together Towards Competitive and Resource-Efficient Urban Mobility /COM/2013/913 final/, Brussels (2013)
2. European Commission: White Paper - Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system /COM/2011/0144 final/, Brussels (2011)
3. European Commission: White Paper- European transport policy for 2010 – Time to decide / COM/2001/370 final/, Brussels (2001)
4. Terzis, G., Last, A.: Urban Interchanges - A Good Practice Guide, Final Report (2000)
5. Monzon, A., Alonso, A., Lopez-Lambas, M.: Key factors affecting the efficiency of transport interchanges. In: 13th World Conference on Transport Research (WCTR), Rio de Janeiro, Brazil (2013)
6. Di Ciommo, F., Vassallo, J., Oliver, A.: Private funding of intermodal exchange stations in urban areas. Transp. Res. Rec. J. Transp. Res. Board **2115**(12), 20–26 (2009)
7. Brons, M., Givoni, M., Rietveld, P.: Access to railway stations and its potential in increasing rail use. Transp. Res. Part A Policy Pract. **43**(2), 136–149 (2009)
8. MIMIC: Mobility Intermodality and Interchanges (1999)
9. PIRATE: Promoting Interchange Rationale, Accessibility and Transfer Efficiency, Sheffield (2001)
10. KITE: A Knowledge Base for Intermodal Passenger Transport in Europe, Vienna (2009)
11. City-HUB: Deliverable D3.2, Guide for Efficient and Smart Design (2013)
12. Palmer, D., Millard, K., Harmer, C., Spousta, J., Kostiainen, J.: Making a successful interchange in reality. In: CITY-HUBs: Sustainable and Efficient Urban Transport Interchanges (2016)

13. Nathanail, E., Adamos, G., Tsami, M.: Why interchanges?. In: CITY-HUBs: Sustainable and Efficient Urban Transport Interchanges (2016)
14. Crozet, Y., Joly, I.: Budgets Temps de Transport: Les Societes Tertiaires Confrontees a la Gestion Paradoxale du Bien le Plus Rare. Les Cahiers Scientifiques du Transport **45**, 27–48 (2004)
15. Mackie, P.J., Jara-Diaz, S., Fowkes, A.S.: The value of travel time savings in evaluation. Transp. Res. Part E. Logistics Transp. Rev. **37**(2–3), 91–106 (2001)
16. Wardman, M., Hine, J.: Cost of Interchange: A Review of the Literature. Institute of Transport Studies, University of Leeds (2000)
17. Hine, J., Scott, J.: Seamless, accessible travel: users' views of the public transport journey and interchange. Transp. Policy **7**(3), 217–226 (2000)
18. Nathanail, E.: Measuring the quality of service for passengers on the Hellenic railways. Transp. Res. Part A. Policy Pract. **42**(1), 48–66 (2008)
19. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Monterey, Wadsworth (1984)
20. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
21. Tsami, M., Nathanail, E.: A decision tree application in transit quality of service in the city of Volos. In: 2nd Conference of Sustainable Urban Mobility, 5–6 May 2014, Volos, Greece (2014)

Mechanism for Investment in the Transport Infrastructure Development in Latvia

Irina Kuzmina-Merlino[1(✉)], Oksana Skorobogatova[1], Niels Schmidtke[2], and Fabian Behrendt[2]

[1] Transport and Telecommunication Institute, Lomonosova street 1, Riga 1019, Latvia
{Kuzmina.I,Skorobogatova.O}@tsi.lv

[2] Fraunhofer Institute for Factory Operation and Automation IFF, Sandtorstraße 22, 39106 Magdeburg, Germany
{niels.schmidtke,fabian.behrendt}@iff.fraunhofer.de

Abstract. Well-developed and efficient infrastructure is crucial for ensuring the effective functioning of the economy. A considerable lack of industry financing certainly affects the assessment of the quality of the overall infrastructure in a global context; according to the report by the World Economic Forum, Latvia languished in 51st position in its global league table for overall quality of the transport infrastructure provision [1]. In countries that have a high rating on the quality of the overall infrastructure, for example, Germany and Great Britain, appropriate investment mechanisms have been developed. They set benchmarks for decision-making and evaluation of investment efficiency. The goal of this study is to develop a conceptual approach to the creation of an investment mechanism for Latvia, based on the experience of Germany and other countries having a high rating of the quality of the overall infrastructure. In general, the results of the study can be useful in developing financial mechanism for investing in the development of transport infrastructure in Latvia as an integral part of the investment policy for the industry development.

Keywords: Transport infrastructure · Investment mechanism
Investment Decision-Making

1 Transport Infrastructure Investment Growing Needs

By the economic nature transport infrastructure is a critical component in economic development of any country at all levels of income. It supports a country's economic growth and human well-being. International experts forecast a significant increase in the need for investment in transport infrastructure in the future. The total amount of infrastructure investment required to sustain economic growth in OECD countries to 2030, is estimated to be above USD 50 trillion [2]. According to the research by Oxford Economics transport infrastructure investment is projected to increase at an average annual rate of about 5% worldwide over the period of 2014 to 2025. Western Europe's share of global transport infrastructure spending is forecast to change from 11% in 2014 to 10% in 2025. For the most of Central Eastern European countries, transport

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_49

infrastructure investment is expected to increase an average of nearly 10% annularly from 2014 to 2025 [3]. It is estimated that USD 350–770 billion in transport is needed annually by developing countries between 2015 and 2030 [4]. Global infrastructure investment needs in road and railroad infrastructure, airports and seaports estimated by McKinsey Global Institute in 2013 is equal to USD 23.8 trillion; EUR 1.5 trillion needed for investment in European transport [5].

The data presented above show very high transport investment needs in the future due to its rapid growth. However, the growing need for funding the transport infrastructure does not mean that this need is provided with financial resources. An unsatisfied need for transport financing is observed in countries with developed market economies as well. For example, in the USA, according to the National Surface Transportation Policy and the Revenue Study Commission, the shortage of capital necessary for improvement the transport infrastructure in the United States between 2008 and 2035 will be 71% [6]. It should also be noted that inadequate financing of transport infrastructure is among ten most serious economic risks for all countries, since the welfare and sustainable development of any region depend on the state of infrastructure.

An analysis of available economic and scientific literature has made it possible to identify problems in the financing of transport infrastructure, which in most cases are common to many countries; the most urgent problem is the gap between the need for financing the infrastructure and the available financial resources. World experience shows that a successful solution to the problem of overcoming the lack of investment for development and innovative renewal of transport infrastructure is possible through the development of investment mechanisms setting the benchmarks for decision-making and performance measurement.

The goal of this study is to develop a conceptual approach to the creation of an investment mechanism for Latvia as a means to make an effective investment decision, based on the experience of Germany and other countries having a high rating of the quality of the overall infrastructure.

The article focuses on studying the following issues:

1. Transportation infrastructure investment performance in Latvia and Germany.
2. How are the investments financed? How to bridge infrastructure financing gap?
3. Mechanism for improving the management process of investment decision making in transport infrastructure.

The results of the study can be useful in building mechanism for investing in the development and support of transport infrastructure in Latvia as an integral part of the country’s investment policy for the transportation industry development.

2 Research Methods

The choice of methodology used in this study has been determined by the logic of solving the research issues and by the necessary to achieve the research aim. In this study, the authors used analytical and logical-structural approaches, statistical and graphical methods of data processing and presentation, as well as the method of comparative economic analysis.

Analysis of the official data published by the state institutions of the Republic of Latvia and the Federal Republic of Germany has proved the priority of transport sector development for their economies. The data that characterizes transport infrastructure investment performance was taken from the official publications of the World Forum and the Organization for Economic Cooperation and Development, OECD. To evaluate the results of the development of transport infrastructure in the international aspect, the calculation methodologies of Global Competitiveness Index (GCI) has been implemented.

At the first stage of the study the authors carried out a preliminary analysis of the general characteristics of the transport infrastructure investments to identify a gap between the need for financing the transport infrastructure and the actual amount invested. A comparative analysis of the dynamics of investment in infrastructure and investment returns is presented graphically for the period 2000–2015; analysis of these indicators allows general comprehension of the amount of finance invested in the development of transport infrastructure and the efficiency of these investments in Germany and in Latvia.

Understandably, the public sector alone cannot meet such a huge annual infrastructure investment requirement. The role of private sector investment (public-private partnership) is shown as findings taking from case studies of successful private sector involvement in transport infrastructure financing in the leading countries.

The final part of the article is devoted to the authors' attempt to develop a conceptual approach to the development of the methodology of investment decision-making, in which the investment mechanism is viewed as a necessary tool connecting the state investment policy and effective decision-making in the field of transportation infrastructure financing.

3 Literature Review

3.1 Transport Infrastructure Investment Performance in Germany and Latvia

This part of the article considers the indicators characterising the investment activity and the results of investing the transport infrastructure of two countries: Latvia and Germany. The analysis of these indicators allows general comprehension of the amount invested in the development of transport infrastructure and the efficiency of these investments.

As the Federal Republic of Germany is located in Central Europe and its companies have substantial economic interdependencies Germany is depending on an efficient freight transportation system [7]. On the one hand, regarding the Logistics Performance Index (LPI) Germany is ranked first in 2016 [8]. On the other hand, regarding the infrastructure Germany is only in the eighth place with declining tendency in the evaluation of the "infrastructure" pillar of the Global Competitiveness Index (GCI, initially collected in 2006) (see Fig. 1) [9]. A federal and state commission established in 2012 already showed this development, that EUR 7.2 billion is lacking every year for renewal infrastructure and deferred investments (any new and expansion projects are not included in this amount) [10].

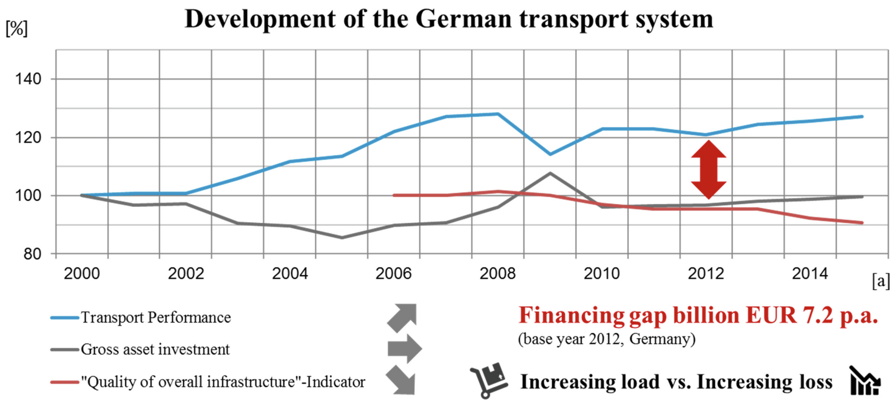


Fig. 1. Development of the German transport system (developed by the authors [9–11]).

In the last 15 years, there has been a continuous increase in transport performance (tkm) with investments remaining approximately the same over this period [11]. The German transport system is facing an increasing loss, which has already been shown by the mentioned commission. This criticism of the state of infrastructure and of the investment restraint of the public sector has meanwhile shown its effects. Since the elections in 2013, the Federal Government responded to the loss and started an investment boom. Two investment programs in the amount of up to EUR 15 billion contribute to higher investments for the transport sector [12, 13].

The transport and storage sector is one of the most promising sectors in the Latvian economy, and today it contributes 9.5% of GDP [14]. The main source of financing the transport sector in Latvia is EU funds. According to the Ministry of Transport of the Republic of Latvia investments in the transportation sector (years 2004–2006 and 2007–2013) amounted in total of EUR 2012.54 million, including financing of 1459.35 million EUR from the EU funds; of EUR 32.77 million from TEN-T; and state, municipalities and private funds together provided EUR 520.42 million. For the period 2014–2020 it is planned to draw funds from the European sources in the amount of EUR 1.3 billion for the purposes of maintaining and developing the transport infrastructure [15]. The OECD data shows that in 2015 total investment in infrastructure development increased more than 12.5 times comparing with 2000 (see Fig. 2).

In Latvia, the priority in transporting goods is given to roads and the railway. It is interesting to see the distribution of investment between these decisions. If in 2000, only 25% of the total amount of allocated investment was on roads, then in 2015 the allocated EUR 412 million is divided almost equally. At the same time, it should be noted that despite the positive dynamics of investments, the indicator “Quality of overall infrastructure” in 2015 was only 4.4. This is the lowest index for the last 6 years.

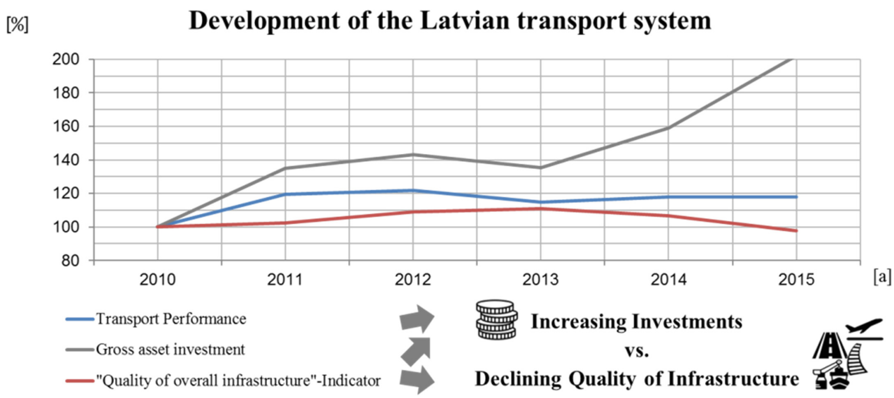


Fig. 2. Development of the Latvian transport system (developed by the authors, [9]).

Analysis of the statistical data leads to the conclusion about the need to develop a conceptual approach for the selection of investment projects and making an effective investment decision.

3.2 Infrastructure Funding and Financing: How to Bridge Financing Gap?

The sources of infrastructure financing. Traditional sources of funding and financing the transport infrastructure are the state budget of each country and various funds. Along with various proposed measures, the European Commission stipulates mobilizing finance for investment combining public and private investment in key transport infrastructure projects. In some countries (Netherland, Iceland, Austria, Finland, and others) with well-developed pension systems pension funds are well-resourced and potentially larger investors in transport infrastructure. Analysing OECD Report "Strategic Transport Infrastructure Needs to 2030" [16] some options of financing schemes of transport infrastructure that could help balance long-term needs and economic advantages of investing in infrastructure against short-term pressures and the costs and consequences of not investing have been founded. In particular, there has been determined the role of the diversified infrastructure funds that could play an important role in delivering project and program funding; such long-term funds are likely to have:

– ear-market, multiple sources of funding – e.g. budgets, fuel taxes, user charges, savings;
– some cross-financing from road taxes and revenues to rail/public transport infrastructure;
– any funding reviews signalled well in advance.

The analysis of economic literature showed that, in addition to the state budget funds, European funds and European investment bank, potential sources of investment into transportation field can be the following asset pools: bank loans, pension funds, insurance companies, transportation network companies (TNCs), sovereign wealth funds (SWF), private equity, venture capital, venture capital funds, individual investors.

Private sector involvement can also help manage the transition to user-pays/self-financing investments. For example, in Australia has been a significant rise in purely private spending, from 15% of total investment in the mid-1980s to more than 50% in 2013 [17]. As it has been noted above, the main source of financing the transport infrastructure in Latvia is European funds; but the amount of the European financing has been decreasing over the years. Therefore, it is extremely important to find alternative sources of financing the transport projects and to use the experience of other countries.

Public - private partnerships (PPPs): case from Germany. Regarding the commission's work [10], the existing finance gap for the overall transport infrastructure (see Fig. 1) has to be closed as well in the area of Federal Highways and roads (about EUR 1.3 billion p.a., base year 2012) [10]. Funding from public investments are not sufficient to improve the state of the German transport system (see declining "Quality of overall infrastructure"-indicator). Therefore, within the commission's work financing instruments have been developed that are both, tax-based and user-based instruments. There is a high coverage potential especially by using the user financing principle [10]. Based on causation additional revenues can be completely and for that specific purpose reinvested in the German road infrastructure [18].

The establishment of a road fund (fund structure) at the national level, which combines revenue from taxes and user fees, as well as an associated financing responsibility, represents a potential financing concept. It enables the use of financial resources for the maintenance and operation of the road network, a higher flexibility in financial disposition as well as a long-term and sustainable link between the performance, cost and financing side. Budgetary funds can be gradually used to expand user financing until a complete financing cycle has been reached [10]. It is conceivable that funds mainly result from traditional tax financing. These can be supplemented by financing instruments such as a toll for heavy good vehicles (HGV) or buses and can be even organizationally enhanced by Public Private Partnership-Models (PPP).

Within PPPs the contracting authorities (public sector) conclude a long-term contract with a private partner as contractor. The goal is to achieve high efficiency and user orientation, i.e. a faster realization of the measure and shorter traffic impairments during the operation and maintenance phase. Furthermore, innovations are encouraged due to functional requirements to create space for ideas of the contractors.

Three different PPP business models are distinguished in the Federal Republic of Germany (see Fig. 3). PPPs model is recognized as the most effective, on the basis of which the investment mechanism operates. PPPs are beneficial for both the state and private business: the state can receive additional revenues in various forms: state general revenue, state fuel taxes, state license permit and fees; private business receives development, strengthening of reputation, growth of well-being of firms.

At the same time, it is necessary to conclude that PPPs require a solid legal and regulatory framework to protect the interests of all partners, public and private, including by ensuring that public policy objectives and contract provisions are met, and providing a stable and sustainable business environment.

Public Private Partnership-Models for Federal Highways

A-model (Expansion)	**V-model (Availability)**	**F-model (Special constructions)**
Since 2005	Since 2009	Since 1994
Expansion, operation, maintenance and pro rata financing	Expansion, operation, maintenance and (pro rata) financing	Construction, operation, maintenance and (pro rata) financing (bridge, tunnel, mountain passes)
Forwarding of the collected toll (risk)	Construction contract, payment independent of traffic volume	Collecting a toll from all users, toll rate setted by customer
User orientation (dependence on traffic volume)	User orientation (dependence on availibility)	User orientation (dependence on traffic volume)
Term of 30 years	Term of 20 to 30 years	-
Pilot projects: A8 Augsburg – Munich A1 Bremen - Hamburg	Pilot projects: A9 Lederhose – federal state boundary Bavaria	Pilot projects: B103n Rostock B103 Lübeck

Fig. 3. Public private partnership-models for federal highways [19].

3.3 Investment Mechanism and Its Role in the Decision-Making

In accordance with "European Transport Policy", represented in the so-called "White Paper", in the context of significantly limited funds of national budgets of the EU member states, there is a serious problem of choosing the format of financing transport infrastructure construction and project upgrading occurs [20].

Among the reasons there are: (1) State budget reducing; (2) Confused roles across the state, municipality and for the private sector; (3) Disordered investment decision-making mechanism; (4) Policy and capital market under-developed in facilitating more private investment; (5) High cost of project delivery, and others. Among the common problems listed above, there is a need to streamline (to improve) the mechanism for making investment decisions.

The investment mechanism is analytical tools and processes permitting: (1) to identify the investment needs, (2) to estimate the projects capable of meeting these needs, and then (3) to determine the most efficient ways of financing these projects. In essence, the investment mechanism is a means for making an efficient investment decision [21].

4 Research Result

This part of the article presents the authors' efforts to develop a general conceptual approach to the construction of an investment mechanism able to help to direct the activities of the authorities in issues of specifying the prioritization and selection of projects which are consistent with the objectives of the state investment policy and could provide the most favourable price-quality ratio. The experience of Great Britain and Germany was taken into consideration in the process of developing the conceptual model of the investment mechanism. An example of an investment mechanism working at the government level is the HM Treasury 'Green Book' issued by the Ministry of Finance of the UK [22].

Any investment solution passes through the ROAMEF cycle, including the following stages: Rational, Objectives, Appraisal, Implementation, Evaluation, and Feedback (see Fig. 4).

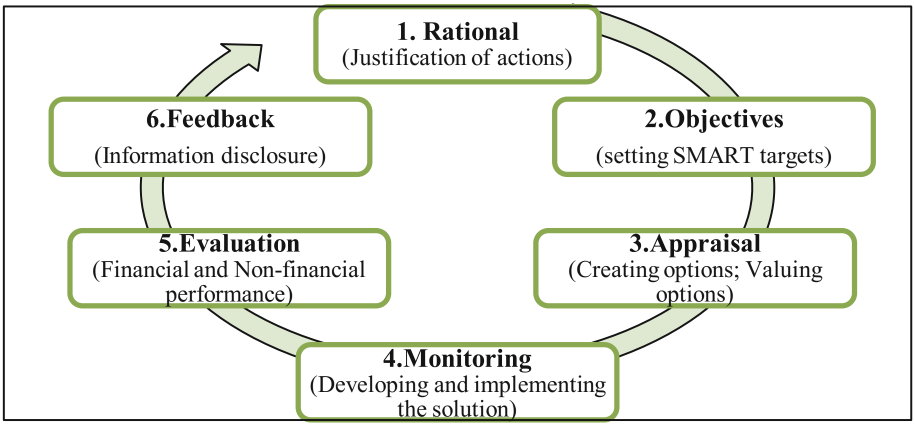


Fig. 4. ROAMEF cycle (developed by the authors, based on the 'Green Book' [22]).

Applying the logical-structural method of research and basing on the results of the theoretical research presented in this article, the authors developed a conceptual approach to the creation of an investment mechanism as a means for making an effective investment decision in accordance with the investment policy of the state (see Fig. 5). Figure 5 shows a typical level structure of mechanisms, the first link of which are mechanisms at the government level, then the level of ministries, and the level of project estimation and reporting closes the structure.

Application of the ROAMEF cycle approach at each level establishes the relationship between the objectives of the investment policy and resources, on the one hand, and the indicators of project implementation and reporting, on the other.

At the level of the government of the country (the 1st level), the investment mechanism determines the general policy in the context of the state strategy for the various government agencies applying for funding; for example, the amount of funding necessary for the development and maintenance of the transport sector is determined (see Fig. 5). As a matter of principle the planning authority as well as the legislature's competences for the interests of transport policy belongs to the national government, with EU White Paper setting out the guidelines for European transport policy [20].

Basing on the national priorities, officials of the Ministry of Transport determine the financial mechanism for investing in the projects of transport infrastructure development (the 2nd level). The officials rely on the developed criteria (public policy, performance indicators, and needs of the society) in the process of working.

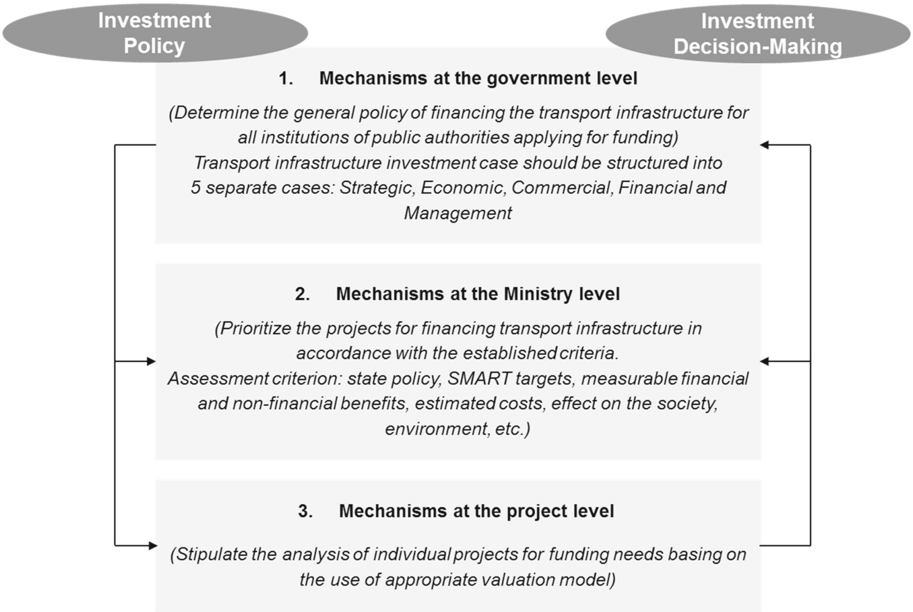


Fig. 5. Conceptual approach to building investment decision-making mechanism (developed by the authors basing on PwC model [23] and 'Green Book').

At the project level (the 3rd level), the individual projects compete for receiving financing for the development of transport to solve a specific transport problem. At this level, the models of assessing the viability and long-term nature of the project are of fundamental importance.

Figure 6 shows a concept intending to support the decision-making regarding investments in transport infrastructure in Latvia. On the one hand, the cost-benefit-analysis as the standard method for ex-ante assessments of infrastructure projects is used [24].

The assessment is based on the cost-benefit ratio, which shows all monetarizable effects in monetary units and compares it with the investment costs. The practical determination of values is aggregated on the on hand by benefit sub-indicators (indicators such as declining transport time, increasing traffic safety, maintenance of traffic routes) and, on the other hand, by cost sub-indicators (indicators such as route, vehicle or material costs) and then summarized in an overall indicator. The result (cost-benefit ratio > 1) describes a meaningful project from a macroeconomic perspective.

Since the cost-benefit analysis alone is not sufficient to assess the impact of infrastructure concepts, it should be supplemented by a multi-criteria analysis. This analysis considers relevant but "missing" effects in a transparent and efficient way in the decision-making process [23, 25].

The composite modelling assessment (COSIMA) takes into account impact types such as regional, urban and environmental development, network accessibility or driver convenience, which have to be measured in appropriate type of quantitative units (e.g. regional development: number of new jobs created; network accessibility: gain in

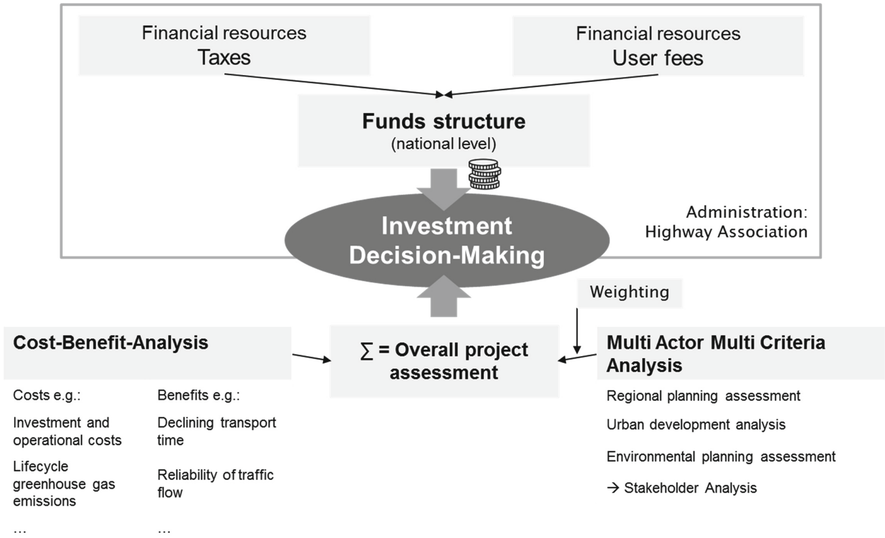


Fig. 6. Investment decision-making concept (developed by the authors basing on commission's fund structure [10] and COSIMA model [23]).

contact hours for specific trip types) [25]. Some impacts cannot be measured quantitatively, so that additional multi-actor analysis is used to include the different stakeholder's points of view [26]. The Analytical Hierarchy Process (AHP) makes it possible to determine a score for each alternative by setting up a target hierarchy, determining the priorities of the individual elements, calculating weighting factors, performing a consistency check and finally compiling an overall hierarchy [27]. The results from AHP scores and all quantifiable units are then translated through weighting factors (e.g. determined by simple multi attribute rating techniques such as SMARTER [28]) into an overall project assessment, which sum indicates the assumed total benefit of the individual planning variants.

This process is opposed by a funds structure at the national level combing both, money from the public and private sectors. In the case of Germany the focus here is on a reliable and adequate financing through traffic-specific taxes and fees, as well the expansion of the earmarked user financing [10]. At this level, the Investment Decision-Making Concept managed by an administration of Highway Association (Fig. 6) can be recommended for Latvia as a possible approach for the estimation of investment projects.

5 Conclusions

1. The role of the state in funding the transport sector remains the leading one, but it becomes obvious that the state cannot withstand the growing requirements for financing this sector. Public private partnerships are seen as an efficient way of financing the transport infrastructure.

2. The lack of an orderly investment mechanism was called by the European Commission among the most serious problems of implementing the investment policy in the development and maintenance of the transport industry.
3. The investment mechanism includes all levels of investment decision-making in the development and maintenance of transport infrastructure in accordance with the investment policy of the state and allows coordination of the investment policy of the state with the available financial resources.
4. The development of an investment mechanism and related tools can help Latvian politicians make a decision about the agents able to implement a particular project more efficiently – the state or the private sector.
5. Investments in transport infrastructure may be attractive for private investors, but only if the investment project developed on the basis of the adopted business model is attractive itself. The essence of the investment mechanism is to help find the most profitable investment solution and the most efficient way of financing an investment project.

The authors have refined their proposed conceptual approaches for developing investment decision-making mechanism into recommendations aimed at policy-makers, professionals, academics, and those with a broad interest in the field.

References

1. World Economic Forum: The Global Competitiveness Report 2016–2017, Geneva (2016). http://www3.weforum.org/docs/GCR2016-2017/05FullReport/TheGlobalCompetitivenessReport2016-2017_FINAL.pdf. Accessed 05 Nov 2017
2. Panayiotou, A., Medda, F.R.: Attracting private sector participation in transport investment. Procedia-Soc. Behav. Sci. **111**, 424–431 (2014)
3. Oxford Economics: Assessing the global transport infrastructure market: Outlook to 2025 (2015). www.pwc.com/outlook2025. Accessed 17 June 2017
4. Zhan, J.: Investment, Infrastructure and Financing the Sustainable Development Goals. WTO, 16 February 2015 (2015)
5. Ouaki, S.: Financial perspectives for EU's transport infrastructure, European Commission and DG mobility and Transport, Riga, 20 October 2016 (2016)
6. BR, Business Roundtable Road to Growth: The Road for Investing in America's Transport Infrastructure, Washington (2015)
7. Schenk, M., Behrendt, F., Trojahn, S., Müller, A.: Verkehrsinfrastruktur-Entwicklungschancen durch effiziente Logistik, pp. 95–100. Jahrbuch Logistik (2014)
8. Arvis, J.-F., Saslavsky, D., Ojala, L., et al.: Connecting to Compete 2016 – Trade Logistics in the Global Economy. World Bank; S.X, Washington (2016)
9. Schwab, K., Sala-i-Martín, X.: The Global Competitiveness Report 2016–2017. World Economic Forum, Geneva, p. 188f (2016)
10. Daehre, K.-H.: Zukunft der Verkehrsinfrastrukturfinanzierung. Final report, Federal and State Commission, Magdeburg, p. 37, 62 (2012)
11. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). Verkehr in Zahlen 2016/2017, Statistics, DVV Media Group GmbH, 45. Edition, Hamburg, p. 24f, 244f (2016)

12. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). Verkehrsinvestitionsbericht für das Berichtsjahr 2014. Drucksache des Deutschen Bundestages 18/8800, 18, Wahlperiode, p. 17 (2016)
13. Weitz, H.: Öffentliche Verkehrsinfrastruktur (2017). http://www.bauindustrie.de/themen/wirtschaft-und-recht/offentliche-verkehrsinfrastruktur/. Accessed 02 Aug 2017
14. LIAA: Transport and Storage, Investment and Development Agency of Latvia (2017). http://www.liaa.gov.lv/en/node/626. Accessed 12 June 2017
15. Ministry of Transport of Latvia (2016). http://www.sam.gov.lv/images/modules/items/PDF/item_6440_Dins_Merirands_20.10.2016.pdf. Accessed 02 July 2017
16. OECD: OECD Report "Strategic Transport Infrastructure Needs to 2030" (2015)
17. BCA: Infrastructure funding and financing, Securing investment in Australia's future, Business Council of Australia, BCA (2013)
18. Doll, C., Schade, W.: How using the Revenues of the German HGV motorway system efficiently and equitable under different regulatory frameworks and institutional settings. In: Proceedings of the 4th Infra-Train Conference, Berlin, 08 October 2005, p. 4 (2005)
19. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). ÖPP-Modelle (2017). http://www.bmvi.de/SharedDocs/DE/Artikel/StB/oepp-modelle.html. Accessed 02 Aug 2017
20. European Commission: European Strategies (2011). White Paper. http://ec.europa.eu/transport/themes/strategies/2011_wtite_paper_en
21. Investments in the development of transport infrastructure. Achieving high results with the least cost, Infra-2, PriceWaterHouse Coopers Int., Ltd. (2012). (in Russian)
22. HM Treasury: The Green Book. London, TSO (2013). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf. Accessed 07 Aug 2017
23. Barfod, M.B., Salling, K.B., Leleur, S.: Composite decision support by combining cost-benefit and multi-criteria decision analysis. Decis. Support Syst. **51**(1), 167–175 (2011)
24. Van Wee, B.: How suitable is CBA for the ex-ante evaluation of transport projects and policies? a discussion from the perspective of ethics. Transp. Policy **19**, 1–7 (2017)
25. Leleur, S., Petersen, N.B., Barfod, M.B.: The COSIMA approach to transport decision making: combining cost-benefit and multi-criteria analysis for comprehensive project Appraisal. In: Korean Development Institute and World Bank Conference, Seoul, pp. 2–5 (2007)
26. Macharis, C., Bernardini, A.: Reviewing the use of multi-criteria decision analysis for the evaluation of transport projects. Time for a multi-actor approach. Transp. Policy **37**, 177–186 (2015)
27. Saaty, T.L.: Decision making with the analytical hierarchy process. Int. J. Serv. **1**(1), 83–98 (2008). University of Pittsburgh
28. Goodwin, P., Wright, G.: Decision Analysis for Management Judgments, 3rd edn. Wiley, Chichester (2004)

Analysis of Riga International Airport Flight Delays

Iyad Alomar(✉), Juri Tolujew, and Aleksandrs Medvedevs

Transport and Telecommunication Institute, Lomonosova 1, Riga 1019, Latvia
{alomar.i,tolujevs.j,medvedevs.a}@tsi.lv

Abstract. During the preparation of this research, the negative impact of aircraft flight delays on airport and passengers were described and analysed. A detailed analysis of aircraft flight delays at Riga International Airport was performed. The result of the analysis demonstrates the season character of flight delay as well as the tight relation between ground handling services and aircraft flight delays. Special effect of flight delays was noted due to movement of ground handling vehicles at the aerodrome. Analysing the delays and their causes will be helpful to improve the prediction of future delays and reduce them as well as reduce of the waiting and downtime of the aircraft on the ground.

Keywords: Ground handling · Flight delay · On-time operation

1 Introduction

In recent years, the increase number of aircraft flight delays in the National Airspace System (NAS) has been the subject of several studies [1]. Flight delays became a fact of each air travel system user and sometimes it has dramatic consequences when the passengers have connection flight. Flight delays have negative consequences on airlines, airports and passengers. On-time operation of the airports and airlines schedules are the target of all airports and airlines stockholders in order to fulfil with passengers and customer requirements as well as getting more new customers [2]. As far as delays are considered as one of the most sensitive remembered performance indicators of any transportation system, it is quite important to give definition of this terminology.

In aviation sector delays are understood as the period of time by which a flight is late or postponed in other words a delay is the difference between the scheduled and the real time of departure or arrival of flight [3]. In other words, delays are defined as "the time lapse which occurs when a planned event does not happen at the planned time" [4].

Taking into consideration airports all over the world, we can note that approximately 20% of all flights have delays by several reasons as well as near to 4% flights are cancelled (see Fig. 1 and Table 1).

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_50

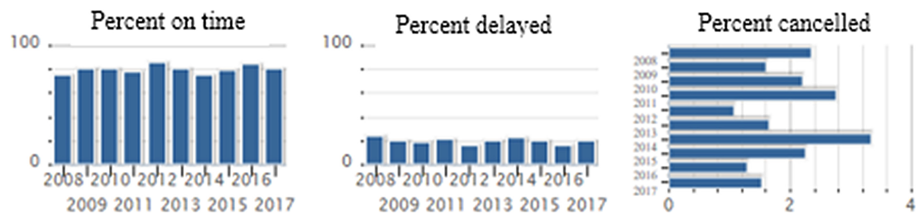


Fig. 1. Worldwide flight delay [7].

Table 1. Worldwide flight delay [7].

Year	Ontime arrivals	Ontime (%)	Arrival delays	Delayed (%)	Flights cancelled	Cancelled (%)	Diverted	Flight operations
2008	2,213,622	73.90	704,709	23.53	70,671	2.36	6,508	2,995,510
2009	2,116,172	79.47	497,520	18.68	42,970	1.61	6,134	2,662,796
2010	2,096,035	79.81	466,350	17.76	57,717	2.20	6,316	2,626,418
2011	1,918,094	76.59	511,002	20.41	68,901	2.75	6,242	2,504,239
2012	2,104,374	84.30	360,463	14.44	26,615	1.07	4,776	2,496,228
2013	2,077,611	79.40	490,102	18.73	43,520	1.66	5,379	2,616,612
2014	1,783,676	74.66	519,682	21.75	79,494	3.33	6,234	2,389,086
2015	1,867,232	78.27	458,474	19.22	53,715	2.25	6,194	2,385,615
2016	1,897,646	82.87	357,236	15.60	29,839	1.30	5,105	2,289,826
2017	1,824,076	79.17	439,424	19.07	35,162	1.53	5,281	2,303,943

In last decades, the expansion in aviation transport has not been smooth. It has expanded in spurts, where periods of growth and increased service were followed by industry recessions and cost cutting [5]. However, this expansion has seriously strained the air travel industry on the ground, where the infrastructures such airport gates, runways, and air traffic control systems, have begun to cause massive delays [6]. Delays have thus become a standard element of air travel.

When the airports commence the organisation of their schedules, they must commit their resources to satisfying their customers', this require all service providers involved in ground handling processes to ensure high efficiency of handling activities and avoiding delays [8]. To explain the level of complexity of flight delays, it is important to understand the network entirely starting from passengers' registration at origin airport and ending by baggage delivery at the destination airport. There is a known relationship between levels of delays and fares, aircraft sizes, flight frequency and complains about airline service [9–11].

From scheduling point of view, which is often built up months before the day of operation, the predictability of operation has a major impact to which extent the use of available resources (aircraft, crew, etc.) can be maximized [12].

Flight scheduling based on where and when the airline will fly. Schedules are set to make the profitability as maximum as possible. The revenue and cost associated with each schedule are based on very different views of the same information [13]. At the same time the schedules are subject to inconsistency due to several reasons. Flight delays could be seasonal or due to current conditions of operations, lack of equipment, bad management, etc. One of the main parts of airports and airlines expenses and losses is that losses are coming from flight delays. Delays of the aircrafts by reasons are tightly related to the ground handling services are quite often observed. Analysing of delays and their reasons will be helpful to improve the prediction of future delays and reduce them as well as reduce waiting and downtime of the aircraft on the ground.

The analysis at the European airport delays shows that:

– 80% of delays on flight arrival could be forecast from the delay on departure.
– The airline which flight destination has been studied allows a very small margin between the scheduled and the real flight and taxiing times, which does not allow absorption of any part of departure delays.
– 70% of delays on departure, could be forecast from the departure plane load factor, when the plane is not significantly delayed on arrival.
– There is some evidence that the scheduled stop time plays an important role in absorbing the arrival delays. For instance, a 45-min scheduled stop time does not allow a delay recovery, if the departure plane is 80% loaded. The shorter the station stop time, the greater the sensitivity of the delay on departure to the delay on arrival (see Fig. 2) [14].

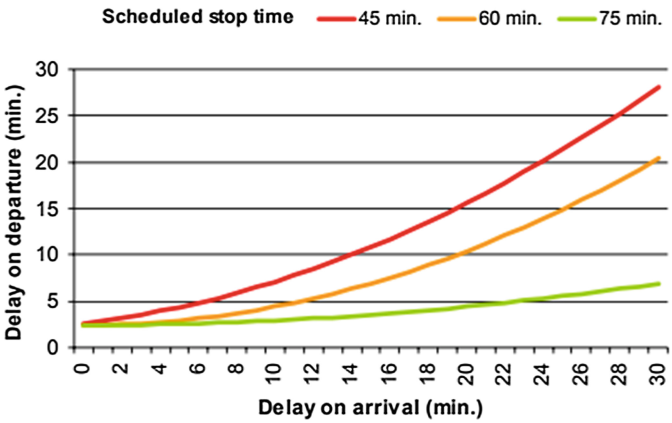


Fig. 2. Departure delay sensitivity to the delay on arrival [14].

From economic point of view, the direct costs originated by flight delays amounted in Europe to 1,250 million euros during 2010 according to the European airline delay cost reference values report from the Westminster University [15].

2 Aviation Sector in Baltic States

Reviewing the air transport passenger flow for Baltic states, we can see that in recent years the demand for air transport has been rising as demonstrated below (see Fig. 3). Based on the data presented on Fig. 3, we can make confirmation, that in each of the

Baltic States, Latvia, Lithuania, Estonia there is a significant increase in demand for passenger transport. In Lithuania it was 13%, in Latvia it was 4.6% and in Estonia 2.5%.

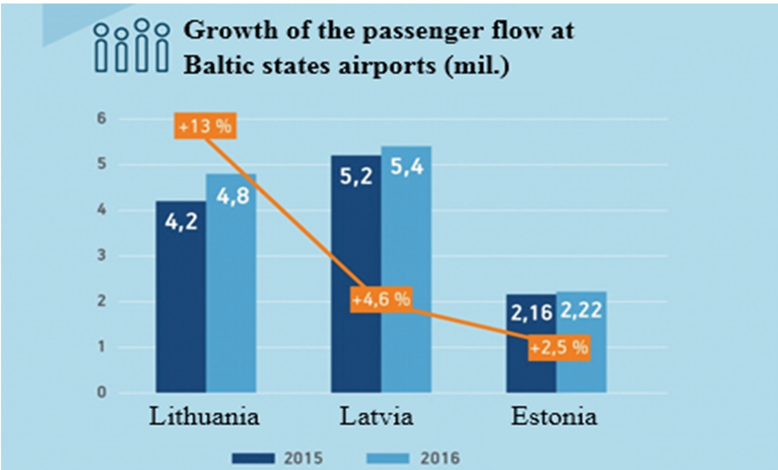


Fig. 3. Growth of the passenger flow at Baltic states airports [16].

Figure 4 demonstrates the growth in certain Baltic state airports in all three countries separately.

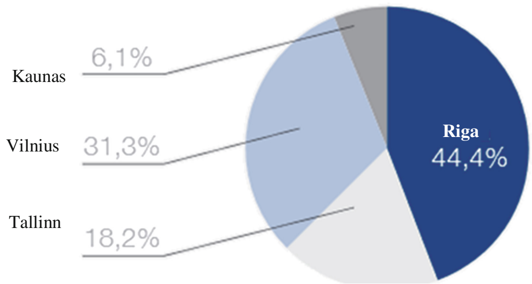


Fig. 4. Distribution of passengers at airports in the Baltic States in 2016 [17].

Related to passenger flow in Latvia, we can note the tendency of stable growth, not only for last 2 years, but for the period from 2004–2016, including the period of world-wide economic crises at 2008. Figure 5 demonstrates the growth of passenger flow for Riga international airport.

In Latvia, the aviation sector provides 0,7% GDP, at the same level as the light industry, at the same time the number of specialist involved in aviation sector for the year 2015 in Latvia increased by 100% comparing to year 2005 (see Fig. 6).

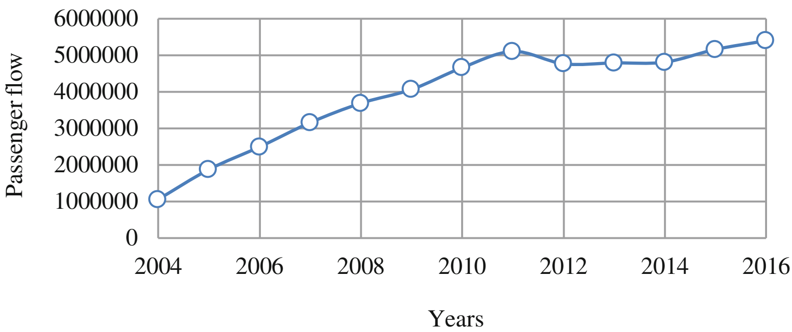


Fig. 5. Passenger flow growth in Latvia [18].

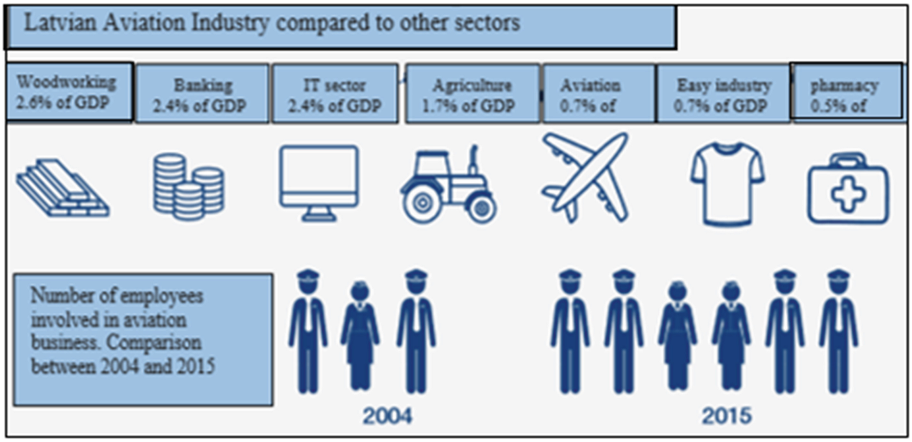


Fig. 6. Latvian Aviation Industry compared to other sectors [17].

During this research, we plan to perform statistical analysis of Riga International Airport flight delays, and then in future link and use the results of the analysis in our research.

3 Analysis of Aircraft Flight Delays at Riga International Airport

Using the data collected by the Riga International Airport statistical bureau, we note that more than 9000 (9251) flight delays were registered at Riga international airport in 2016. All flight delays were categorized by delays codes. Approximately 70 reasons of flight delays were registered. The total time of all flight delays in 2016 was 3200 h. For illustration, flight delays were grouped to months (see Fig. 7), from this figure, we can note that the influence of seasons on flight delays. The delays number increase by approximately 30% during high seasons, from 5–8% to 10–11%.

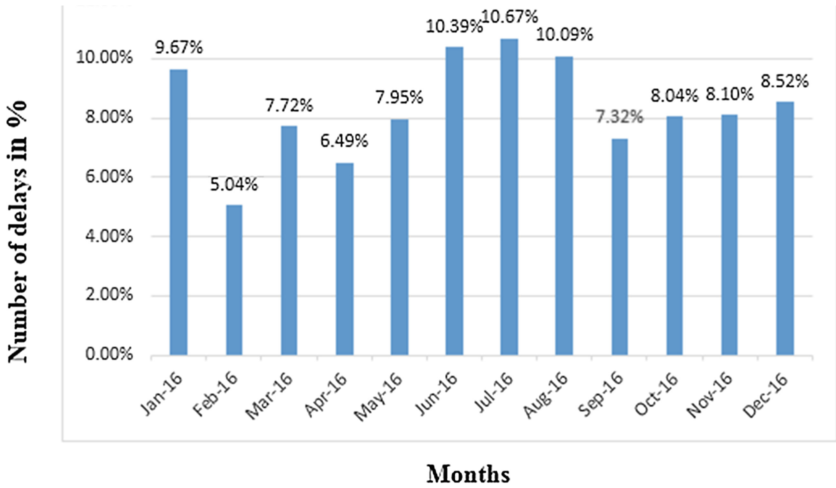


Fig. 7. Seasonality characteristic of flight delays.

Major part of delays was caused by reasons related to Fig. 8a:

- availability of aircrafts at Riga International Airport;
- restrictions at airport of destination or departure – airport and/or runway closed due to obstructions, etc., awaiting load from another flight, late arrival of aircraft from another flight or previous sector) – Total delays time for this category was more than 1 500 h. Average flight delay time was 19 min;
- technical maintenance of aircrafts including scheduled and non-scheduled maintenance and flight diverting due to maintenance reasons – total delays due to this reason was more than 600 h. Average flight delay time was 54 min;
- delays due to discrepancy with ATC data (Air Traffic Control) – total delays time for this category was more than 300 h. Average flight delay time was 19 min;
- delays due to ground handling process – total delays time for this category was more than 2500 h. Average flight delay time was 15 min.

For further and more detailed analysis, the flight delays by reasons not related to airport operation procedures (Air Traffic Control, weather, crew, etc.) were grouped separately from other flight delays reasons. This analysis mainly will deal with reasons related to airport operations procedures (group A of Table 2). The reasons in group B will be not discussed during this research as far as our research focuses on optimization of ground handling vehicles movement at the aerodrome. There are a lot of researches dealing with other reasons of flight delays such as [19] which deals with weather impact on flight delays and [20] which deals mainly with air traffic flow management.

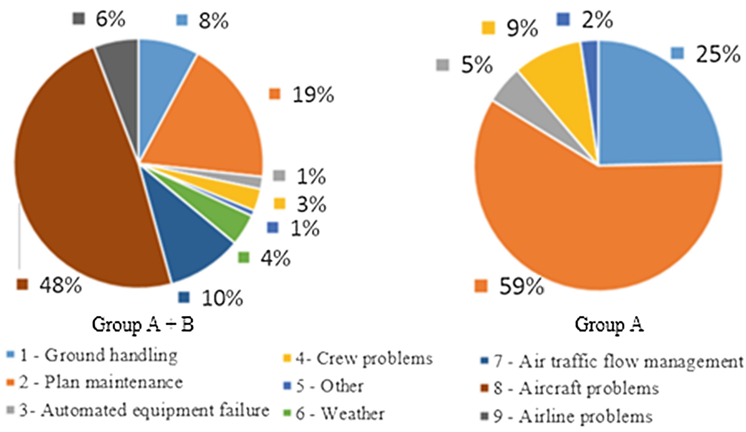


Fig. 8. Distribution of the flight delay times (h) for selected groups in accordance with Table 2.

Table 2. Characteristic of main flight delays groups.

No	Problems group	Numbers of delay	Time of delay total (h)	Time of delay total (min)	Average of delay (min)
Group A: characteristic of main flight delays related to airport procedure					
1	Ground handling	1 188	251.72	16 510	15
2	Plan maintenance	676	602.92	36 175	54
3	Automated equipment failure	209	50.92	3 055	15
4	Crew problems	472	91.42	5 485	12
5	Other	140	23.73	1 424	10
Total of the group A		2 685	1 020.70	62 649	23
Group B: characteristic of main flight delays related to other procedures					
6	Weather	468	131.43	7 886	17
7	Air traffic flow management	975	311.20	18 672	19
8	Aircraft problems	4 851	1 550.35	93 021	19
9	Airline problems	272	186.37	11 182	41
Total of the group B		6 566	2 179.35	130 761	20
TOTAL of groups A + B		9 251	3 200.05	193 410	21

Taking into account above mentioned correction, the flight delays related to ground handling services at Riga International Airport will be more than 25% of total delays time (see Fig. 8b).

From Fig. 8b and Table 2 it is seen that main cause of flight delays is technical maintenance of aircraft, the second by importance is ground handling services 25% of total delays. As far as ground handling process is quite wide, and there are a lot of reasons hidden behind, and not all of them related to ground handling vehicles movement at the aerodrome. Description of ground handling causes are illustrated in Table 3 and Fig. 9. At the meantime in Table 2, we can note that more than 600 flight delays or 49% of total delays related to ground handling causes, directly related to ground handling vehicles movement at the aerodrome (group B, D-I), as well as 44% from total delays were related to passenger's registration and errors in baggage registrations.

Table 3. Characteristic of flight delays related to ground handling services.

No	Problems group	Numbers of delay	Time of delay total (h)	Time of delay total (min)	Average of delay (min)
A	Check-in error	676	133.78	8 027	12
B	Baggage processing	2	0.23	14	7
C	Cargo	83	21.15	1 269	25
D	Loading/ unloading	507	129.27	7 756	15
E	Servicing equipment	36	4.83	290	8
F	Aircraft cleaning	18	2.62	157	9
G	Fuelling/ defueling	17	4.17	250	15
H	Catering	13	2.15	129	10
I	Operation requirements	74	8.25	495	7
TOTAL		1426	306.45	18 387	13

By performing analysis of flight delays related to the use of ground handling vehicles movement at the aerodrome, we can see that the main cause of the delays in this group is tightly related to aircraft loading/offloading procedure (group D, Table 3 and Fig. 9), more than 500 aircraft flight delay.

Maximum duration of flight delays is 15 min, total duration of flight delays by this reason is 130 h or 85% of delays in this certain group. Next causes are presented here in after (see Fig. 10).

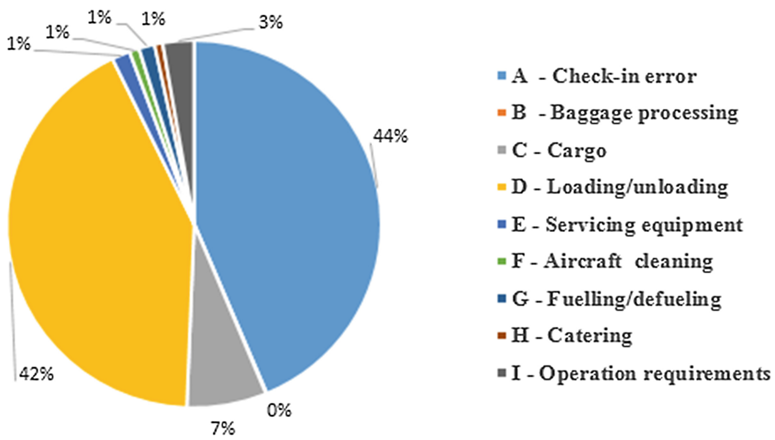


Fig. 9. Distribution of the flight delays times (h) due to ground handling services for selected groups in accordance with Table 3.

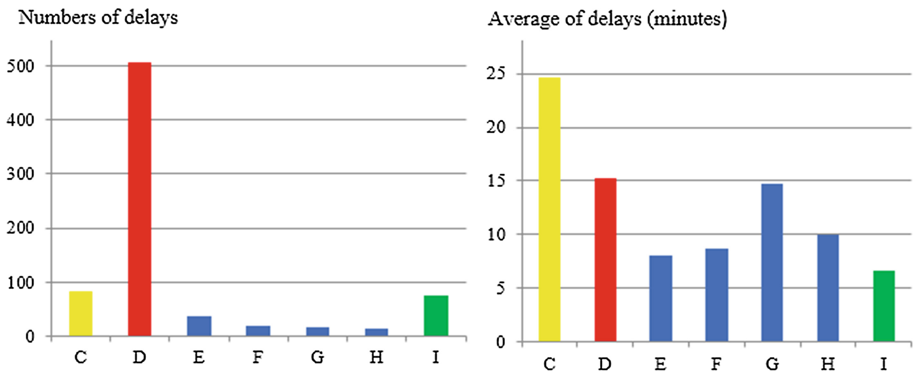


Fig. 10. Characteristic of flight delays by cause of ground handling movement at the aerodrome.

As we can see from above presented data, one of the main factors that play a key role in increasing the efficiency of airport operations is reducing the downtime of the aircrafts on the ground by reasons related to ground handling services.

The most important way to improve airport efficiency is control and optimization of ground handling vehicles movement at the aerodrome.

In order to deal with the delays related to ground handling movement at the aerodrome at Riga International Airport, a simulation model will be developed for fragment of Riga International Airport, where we plan to test and simulate the operation of above mentioned airport.

The authors of this research currently are working on the optimization of ground handling vehicles movement issues by using simulation modelling [21]. The researched methods of increasing of the effectiveness of ground handling vehicles movement are based on the implementation of prioritization rules of vehicles movement and changes in the geometry of sections of the road on which traffic congestion may occur. Data on flight delays at Riga International Airport are used by the authors to validate the simulation model of ground handling vehicles movement, as it becomes possible to compare the results of statistical modelling with real data. The models tested in this way can be used to optimize the ground handling vehicles movement in large airports with intensive traffic.

4 Conclusion

The influence of o ground handling vehicles movement at the aerodrome on aircraft flight delays was observed.

The analysis of flight delays data for Riga International Airport demonstrates that the number of flight delays increases by approximately 30% during high seasons. Unfortunately, we could not analyse the delays by day time and dropped it to companies as far as this data is confidential.

By analysing the reasons of flight delays, we note that these reasons are divided into two main groups, one group is related to services which are not directly related to airport operation procedure. This group was not studied during this research as far as it is not the subject of our research. The other group is tightly related to airport services. Analysing the second group the relation between flight delays and various aspects of airport services was observed. Analysis shows that there is a considerable impact of ground handling services on the flight delays. Taking into consideration that all ground handling services are performed by using ground handling vehicles especially loading/unloading, servicing equipment refuelling, etc., we can make conclusion that optimization of ground handling movement vehicles at the aerodrome will reduce the flight delays times.

As it follows from the analysis, which were performed above, the effectiveness of the organization of ground handling vehicles movement is not the most important factor of flights delays at Riga International Airport. However, at many airports, there is much more intensive traffic of ground handling vehicles observed and this intensive traffic increases the importance of this factor.

References

1. European Organization for the Safety of Air Navigation (EUROCONTROL): Delays to Air Transport in Europe, Brussels (2013)
2. Sternberg, A., Soares, J., Carvalho, D., Ogasawara, E.: A Review on Flight Delay Prediction. CEFET/RJ Rio de Janeiro, Brazil, 20 March 2017 (2017)
3. Wieland, F.: Limits to growth: results from the detailed policy assessment tool. In: AIAA/IEEE Digital Avionics Systems Conference – Proceedings, vol. 2, pp. 9.2-1–9.2-8 (1997)
4. Guest, T.: A matter of time. Air traffic delay in Europe. EUROCONTROL Trends in Air Traffic, vol. 2, Brussels, Belgium (2007)
5. U.S. Department of Transportation, Office of Airline Information. http://www.bts.gov/oai/indicators/top.html#PassengerService. Accessed 31 July 2017
6. Smith, T.: Airline Flight Delays and Flight Schedule Padding, Philadelphia (2000)
7. Bureau of transportation statistics. https://www.transtats.bts.gov/HomeDrillChart.asp. Accessed 22 July 2017

8. Mueller, E.R., Chatterji, G.B.: Analysis of aircraft arrival and departure delay characteristics. In: AIAA's Aircraft Technology, Integration, and Operations (ATIO), Los Angeles, California (2002)
9. Bhadra, D.: You (expect to) get what you pay for: a system approach to delay, fare, and complaints. Transp. Res. Part A Policy Pract. **43**(9–10), 829–843 (2009)
10. Pai, V.: On the factors that affect airline flight frequency and aircraft size. J. Air Transport Manag. **16**(4), 169–177 (2010)
11. Zou, B., Hansen, M.: Flight delay impact on airfare and flight frequency: a comprehensive assessment. Transp. Res. Part E Logistics Transp. Rev. **69**, 54–74 (2014)
12. Jetzki, M.: The propagation of air transport delays in Europe. Department of Airport and Air Transportation Research, RWTH Aachen University (2009)
13. Barnhart, C., Smith, B. (eds.): Quantitative Problem Solving Methods in the Airline Industry. A Modeling Methodology Handbook. Springer (2012). https://doi.org/10.1007/978-1-4614-1608-1
14. EUROCONTROL: Flight Delay Propagation Synthesis of the Study (2003)
15. Cook, A., Tanner, G.: European airline delay cost reference values, Performance Review Unit EUROCONTROL (2011)
16. Vilnius Airport (2016). http://www.vilnius-airport.lt/en/news/?id=1986099. Accessed 22 July 2017
17. Riga International Airport (2016). http://www.riga-airport.com/uploads/files/RU%20RIX%20Gadagramata%20makets%20RU(1).pdf. Accessed 22 July 2017
18. Riga International Airport (2017). http://www.riga-airport.com/ru/main/ob-aeroporte/statistika. Accessed 20 July 2017
19. Vane, R.: Flight delay analysis and possible enhancements with big data. Int. Res. J. Eng. Technol. **03**(06), 778–780 (2016)
20. Ivanov, N., et al.: Air traffic flow management slot allocation to minimize propagated delay and improve airport slot adherence. Transp. Res. Part A **95**, 183–197 (2017)
21. Alomar, I., Tolujevs, J., Medvedevs, A.: Simulation of ground vehicles movement on the aerodrome. Procedia Eng. **178**, 340–348 (2017)

Modelling and Simulation of the Riga International Airport to Reduce Turnaround Times of Crucial Clearance Processes

David Weigert[1(✉)], Alina Rettmann[1], Iyad Alomar[2], and Juri Tolujew[1]

[1] Fraunhofer Institute for Factory Operation and Automation IFF Magdeburg, Sandtorstraße 22, 39106 Magdeburg, Germany
david.weigert@iff.fraunhofer.de

[2] Transport and Telecommunication Institute, Lomonosova Street 1, Riga, Latvia
alomar.i@tsi.lv

Abstract. This paper focused on prioritisation of certain ground vehicles and improving the turnaround time of airplanes at the airport. This leads to acceleration of ground operations at the airport. Key indicators to measure delay and improvement within the system are the non-operation period of an airplane. Depending on the increasing number of passengers in the next few years, it is necessary to shorten this time and accelerate the processes of handling operations. The built conceptual model contains all necessary processes relating aircraft and ground vehicles and their movement. Based on the conceptual model a rough calculation was developed. The simulation model will be based on the rough calculation, therefore. Through the evaluation of gained data, new insights and new ideas for future simulation studies will be produced.

Keywords: Ground vehicle movement · Aircraft processes

1 Introduction and Problem Formulation

Introducing words. As low-cost carriers are rising within the airline industry and numbers of passengers are rising in general, airports and airlines alike try to optimise their processes in order to maximise their revenue. Worldwide the number of air passengers will nearly double between 2016 and 2035 [1]. This development can also be witnessed in the European union, as the number of passengers in air transport industry grew between 2014 and 2015 in the EU28 by 4,72% [2]. In fact, an airplane only creates sales revenue as long as it transports passengers. Therefore, airlines try to minimize ground times of the airplanes within their fleet, optimise all processes on the ground and minimise delay. Hence, the punctuality of departure for each flight is important as it keeps the flight schedule stable [3]. The efficient use of scarce resources at the airport can lead to minimal ground times and aide in maximising revenue. Thus, delay has to be avoided in every context, including late arrival of resources at the aircraft. Especially the evaluation for the airport RIX illustrates increasing passenger numbers of up to 660%

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_51

in contrast to 2004, seen in Fig. 1. According to own forecasts, in the year 2020 approximately 580 thousand passengers will be handled.

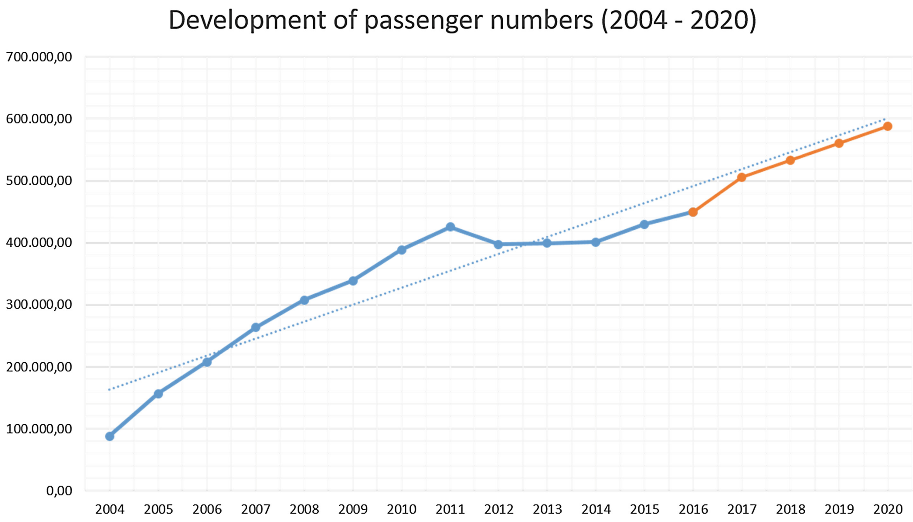


Fig. 1. Current figures for 2004 to 2016, identified forecasts 2016–2020.

Problem formulation. The keynote for this paper is the hypothesis that prioritisation of certain ground vehicles improves the turnaround time of airplanes at the airport. Key indicators to measure delay and improvement within the system are the non-operation period of an airplane, which can also be called the handling time of an airplane, the (estimated) time of travel for ground vehicles, the distance ground vehicles need to drive and utilisation of resources, in this case ground vehicles. Control factors, which can be used to control key indicators, are the parking position of the airplane and the routing algorithm of the ground vehicles as well as the way of prioritisation.

Proceeding. The conceptual model has to contain all necessary objects, processes and key indicators. Key indicators for measuring improvement were defined above; data for comparison, verification and validation is going to be created through a rough calculation. In addition, the rough calculation can validate the conceptual model, if the usual schedule within the rough calculation matches real time data and/or is based on it. In the next step, the formalisation of the conceptual model is going to be done. Afterwards, scenarios and experiments can be planned to gain wanted information. Projects following the evaluation of gained data will be discussed in short in the outlook and mark the end of this paper. The chosen method is the simulation of the movement of ground vehicles at apron three of the Riga International Airport (RIX). Because RIX offers more destinations in summer than in winter, the summertime has been chosen for the simulation study [4].

1.1 Airport

Riga Airport. RIGA International Airport, abbreviation used by IATA is RIX, abbreviation used by ICAO is EVRA, is the airport serving Riga city and is located 5.4 nautical miles (approximately ten kilometres) from Riga [5]. The airport was opened in 1974 [6]. The only holder of the airport is the republic of Latvia, with the Ministry of Transport of the republic of Latvia as its shareholder. The airport is the biggest international aviation company located within the Baltics, serving 89 destinations, and is serviced by full-service and low-cost airlines alike [4] In 2016 around 680 000 flights and about 5.4 Million passengers were served. The Riga Airport has four aprons and one terminal building, as seen in Fig. 2.

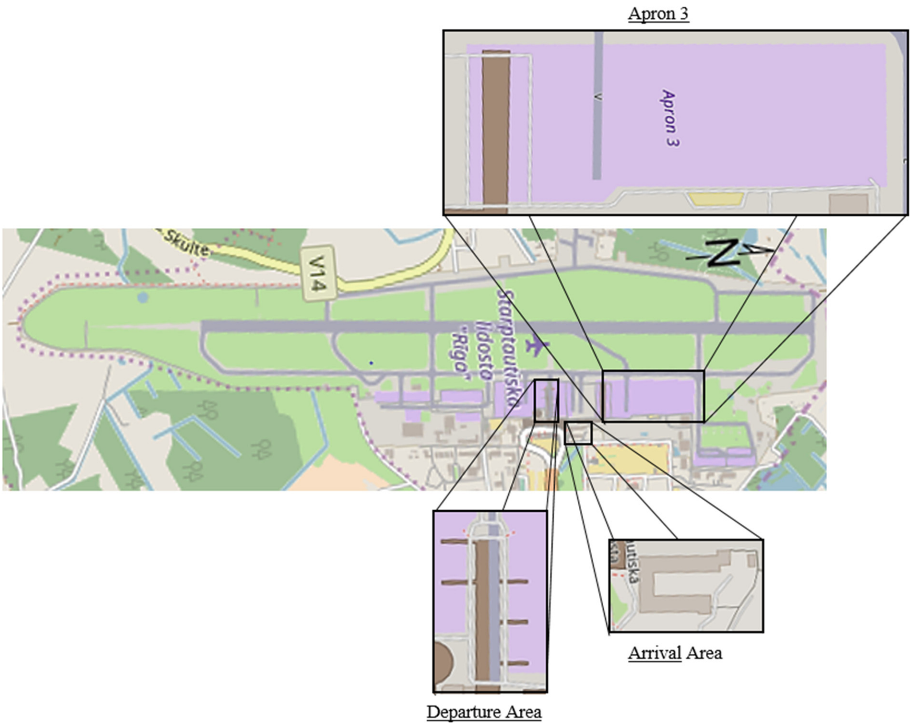


Fig. 2. Section of RIX [7] with enlarged Apron 3, Arrival Area and Departure area.

Airport Structure and usual processes. An airport is a complex system, which can be divided into three elements: landside spaces, airside spaces and the terminal building, connecting landside and airside [8]. Neither staff nor activities carried out are restricted to one of the functional parts of the airport [3, 9]. Within and through this system, passengers, cargo, aircrafts and surface or ground vehicles are being serviced. The airside consists of runways for landing and take-off, taxiways as connections between runway, terminal building, apron and gate. Within apron and gate areas, aircrafts are parked and serviced [8]. Beginning on the kerbside of the terminal building, the landside is not part of this contemplation. According to [3], the operations happening on the airport can be divided in either landside processes, including processes happening in the terminal building, or airside processes [3]. The processes needed for the handling of an aircraft on the apron are considered airside operations. This includes the supply of the aircraft with fresh water, fuel, electricity and in-flight catering as well as cleaning the aircraft. Furthermore, the transport of passengers, cargo and aircrew between the aircraft and the terminal building are a part of the airside processes [3]. This paper focuses on the airside processes. An Aircraft can be handled in three different ways; the turnaround, the withdrawal and the retrieving. The turnaround, means turning the aircraft around for the flight back, withdrawal (Abzug, cf. [3]), which means the transport of the airplane to a parking position or a dockyard, and provision, in the meaning of retrieving the airplane from its parking or dockyard position (Preparation, cf. [3]). In this work, only apron three is examined. For this reason, it is assumed, that only turnarounds take place.

Turnaround processes. The processes explained in the following can also happen simultaneously or overlap each other, in case of delay, to minimise ground time. In addition, only necessary and regular operations at the airport are described, as the processes happening differ among airlines, due to differences in resources and staff [10, 11]. After the airplane is parked and the jet engines have been turned off, the passengers are allowed to deplane [3, 10, 11]. Meanwhile, the unloading of baggage begins. If the airplane is at a terminal position, deplaning will happen with the help of a passenger boarding bridge. If the airplane has its parking position at the apron, the passengers will leave the airplane with the aid of stairs and an apron bus, which brings the passengers to the terminal. The process following are cleaning the passenger cabin, the toilets and – if necessary – the cockpit of the airplane, as well as catering the galley. If international hygiene standards handling or quality standards regarding food cannot be met by the turn around airport, the galley is filled at once for several future stops [3, 9]. Further service performance includes a toilet tanker to empty the airplanes toilets, a tanker for fresh water and visual inspections executed by mechanics. The fuelling of the airplane has to take place after the unloading and before the loading processes, as resources for both processes occupy the same sections around the airplane [12]. Fuel is delivered into the aircraft by a tanker truck or a fuelling pumper and a fuelling hydrant, which is connected through an underground system of pipelines to fuel tanks [3, 13]. The captain decides how much fuel is pumped into the airplane, which can be a reason for delay. After the fuelling, passengers begin to board the aircraft. The loading of the aircraft happens simultaneously or after the boarding process. Luggage is only allowed aboard an airplane, if it belongs to a passenger. Otherwise, the luggage must be unloaded again, resulting in delay. In addition to this, the centre of gravity and the balance of the airplane have to be considered during the loading process [3]. During wintertime, airplanes need to get de-iced and anti-iced. The first is the removal of ice from the airplane, as ice can influence control devices of the airplane, changes the aerodynamic way of acting of the airplane and adds up as deadweight unto the airplanes weight [14]. Anti-icing means, that a layer of protection to avoid formation of ice is applied. These processes happen immediately before departure. If the aircraft is ready to deplane, the pilot has got to ask for permission to start the jet engine and move towards the taxiway for take-off.

2 Method

2.1 Chosen Method: Simulation Study

As a method of solving problems, simulation-modelling means that a simulation model, built to delineate the original or existent system and/or its way of behaving is created using simulation software [15]. Therefore, experiments can be done without change to the existent system, as the study can be done on the simulation model. In case the simulation model is verified and validated, it is possible to transfer gained knowledge onto the existent system [15].

To be able to display the haphazardness on the apron of RIX, and to be able to work and experiment with the model built, the simulation study was chosen. Simulation models can be categorised. Whether time is considered (dynamic) or not (static). If the model changes as time goes by, it is considered a continuous model, while discrete models change at certain moments in time. A combination of both is possible, too. Furthermore, models can be categorised whether they are stochastic, therefore including haphazard distribution, or deterministic [12]. The model, which will be built for this project is a dynamic, discrete and continuous, as well as stochastic model.

2.2 Conceptual Model

The conceptual model is a logical model a part of reality that is to be built with the aid of simulation software. With this method, important aspects can be determined and simulated and less important aspects can be left out.

Given Layout. In Fig. 3, all essential buildings are shown in the layout. The numbered rectangles are parking spaces for airplanes. The colourful lines (blue, green, red) indicate driving paths for ground vehicles. The blue lines symbolise both-way paths for ground vehicles, which allow a driving speed up to 15 kilometres per hour, while green lines allow a maximum speed of 30 kilometres per hour. Red lines are one-way paths, which limit the driving speed to 30 kilometres per hour. In Fig. 2 the lower red line, is only used to drive into northern directions, therefore to Depot 2. The upper red pathway can only be driving in the southern direction, leading towards the Arrival and Departure area. Grey pathways are for airplanes. It is presumed, that airplanes drive with a speed of 30 kilometres per hour on the apron and all connecting roads. The building labelled with “Arrival” is the destination of the passenger bus coming from the aircraft and marks the Arrival area. “Departure” marks the Departure Area, where passengers enter the bus to be transported to the aircraft. In Depot 2 tanker trucks, the busses for deplaning passengers and tug and dollies for unloading the airplane are stored. Busses for enplaning as well as tugs and dollies for loading baggage into airplanes.

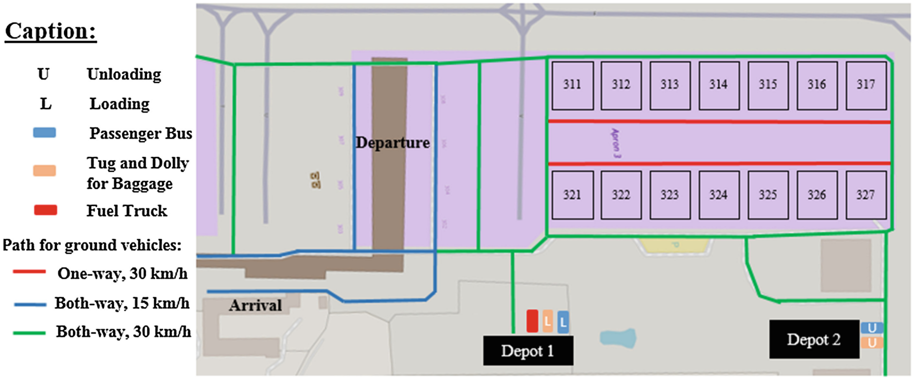


Fig. 3. Layout of Apron three of RIX, following [7].

Process Chain model. Due to the fact, that only Apron 3 is simulated (cf. Fig. 1), certain adjustments to the processes described in Sect. 1.1 were made. Within our model, only turnarounds are simulated; all passenger seats are being occupied, because overbooking is a common practice used by airlines [16]. Depending on the airline, each passenger carries a certain amount of baggage, ranging from 17 to 96 kg in total. For deplaning, a constant rate of 18 passengers per minute was chosen [10, 11]. The passengers reach the arrival area with the aid of an apron bus, which can carry up to 99 passengers. During the unloading process, 1.8 m^3/min can be unloaded. During the fuelling process, 1 500 litres of fuel are pumped by a tank truck into the airplane. It is presumed, that fuel tanks have to be filled 100%, as the amount of fuel needed for a flight with an airplane depends on many variables. These variables range from expected airplane weight, which depends on the number of passengers, weather conditions and possible or expected diversion or holding patterns during the flight to fuel prices at the destination airport [3]. Simultaneously to the fuelling process, the passengers for the next flight are boarding the apron bus and their luggage is transported towards the airplane from the Departure Area. After their arrival at the airplane, the passengers enplane the airplane and their luggage is loaded into the airplane. After this, the airplane is ready for departure and its captain receives the allowance to start jet engines. De- and anti-icing process are excluded, because the simulation study is situated in summertime (compare Sect. 1 Introduction).

2.3 Data Preparation and Rough Calculation

Data Preparation. Through the website of Riga International Airport, real departure and arrival times were taken. Furthermore, data concerning the aircraft type and the number of seats were selected. It was decided, that the amount of luggage per passenger depends on airline specific regulations. Thus, the number of passengers per airplane was multiplied with the maximum allowed luggage per airline. In general, cargo flights were omitted. In order to gain data, which is close to reality and values per minute, turnaround time charts from aircraft manufacturers and aircraft specific data were taken. Values per minute, which were subdivided into values per second, are less error-prone in the conversion from hundredths of an hour to sixtieth of an hour. Furthermore, turnaround times for any aircrafts can be calculated with these values. For the road map, a distance matrix was compiled. Maximum allowed speed is taken for driving speed. Acceleration and deceleration are excluded.

The frequencies of short, medium and long haul flights in Fig. 4 were determined by the analysis of 458 data sets in the winter schedule 2017 (00:00–23:59, Mon–Sun, August–October). The data sets each describe a fixed destination, time and airline. There are still no repetitions of the flight in the 458 sets determined. A further evaluation would yield even more data. However, these are not important in the current concept phase. As can be seen in Fig. 3, the rush hours on the RIX are between 6:30 am and 8:00 am, as well as 11:30 am and 1:30 pm. Here, the planning and process flow of the handling processes are particularly required.

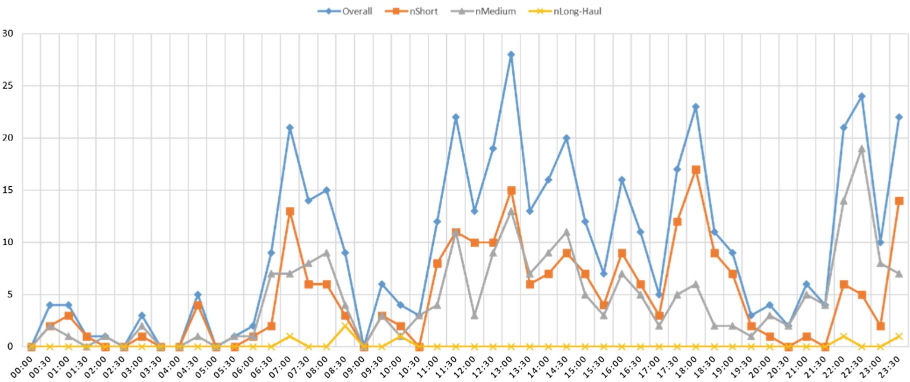


Fig. 4. Frequencies of Start/Landing for short, medium and long haul flights.

Rough Calculation. The next step was to unite all data in one calculation with the help of Microsoft excel. By means of random numbers, a possible time schedule was created. With the values per second, specific turnaround times per airplane could be calculated. This time schedule will be used as a basis for simulation to verify the simulation model, as the verified model should have about the same time values as the rough calculation.

3 Result of the Conceptual Model

3.1 Formalised Model and Experiments

Formal model. For the formal model, the simulation software Anylogic was chosen, due to the possibility to combine elements of different simulation libraries in combination with visualisation and animation of the simulation model [15]. For the layout, three independent road networks, one for airplanes, one for ground vehicles and one for prioritised ground vehicles, will be built. In this layout, all airplanes will be prioritised before all ground vehicles and all prioritised vehicles will be prioritised before the remaining ground vehicles. In Fig. 3, black lines indicate the road network for airplanes, while the colourised lines are the road network for ground vehicles. The allowed driving directions and maximum driving speed can be extracted from Sect. 2.2. As possible locations to cause traffic jams, junctions are important and crucial for a smooth flow of ground vehicles. As a consequence, junctions will be simulated in segments, as seen in Fig. 5. Through this, the approach to junctions can be atomised. The motorized vehicle (ground vehicles or airplanes) approaches the junction and reaches a prior defined point, at which the driver receives the order to either halt the vehicle, or to continue driving. If the vehicle stopped, the driver needs to wait for permission to start driving again. As ground vehicles are subordinate to aeroplanes, it is unlikely that aeroplanes have to wait, unless in case of an emergency.

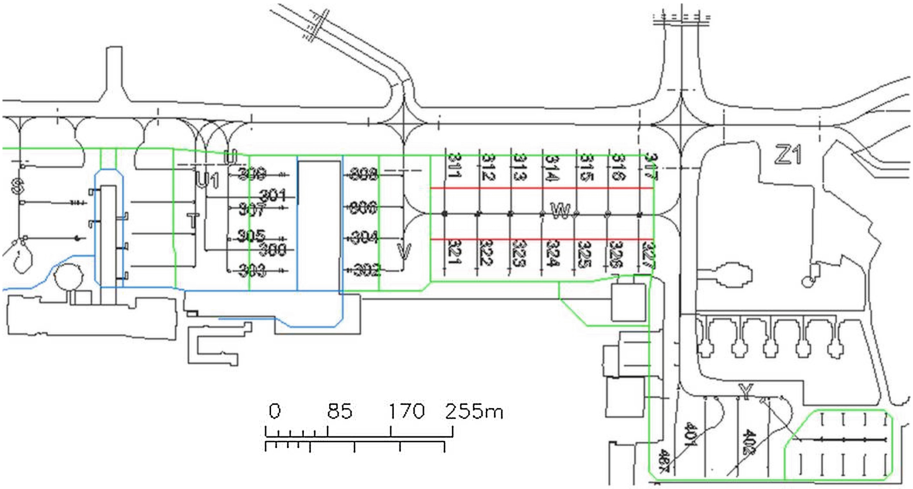


Fig. 5. Basic layout for the formal model.

Planned experiments and scenarios. The simulation model offers several opportunities to work with control factors through experiments to find new insights.

The type of vehicles driving on the third, prioritised network is a control factor, which will be explored with the aid of the simulation model. Questions arising are whether just one kind of vehicles, for example only tank trucks, or only vehicles belonging to certain stands, for all regarding the one way drives, or even only ground vehicles for certain aircrafts are allowed to use the third lane. Through the type of vehicle allowed to drive on the prioritised lane the number of vehicles driving on this lane can be controlled. Nevertheless, the possibility of a traffic jam at conjunctions remains the same or even increases. Thus, the capability to drive quicker on the prioritised lane is the only way, to take an advantage from driving on this lane. Through experiments, a concrete number, below which this capability is kept, could be found. In addition, a fluent change between the prioritised lane and the normal lanes should be able, to keep the number of vehicles on the prioritised lane as small as possible at all times.

The assignment algorithm for the ground vehicles is also a control factor for experiments. The released ground vehicles could drive either back to their depot, or be assigned a new task. Possible constraints for choosing the next task could be based on the driving way, driving time the next task or the timeline, therefore the next timely free task is assigned, without regarding driving way. Furthermore, a combination of two or all three constraints could be done.

4 Conclusion and Outlook

Conclusion. All in all the built conceptual model contains all necessary processes relating aircraft and ground vehicles and their movement. Based on the conceptual model a rough calculation was developed. The simulation model will be based on the rough calculation, therefore the rough calculation will be used for validating the simulation model. Furthermore, the simulation model will be validated, if it is orientated on the rough calculation. Several promising ideas for experiments, were presented, which means, that the best one should be found through the simulation study. It might be possible to find indicators, which scenario is best to use in what situation.

Outlook. Through the evaluation of gained data, new insights and new ideas for future simulation studies will be produced. From the algorithms regarding the assignment of new tasks to ground vehicles, conclusions could be drawn for the automated assignment of work for ground vehicles. If the airport transport network management will equip every ground vehicle with a device to tracks its current location, speed and the approximate time until the ongoing task is finished, an automated assignment of tasks could be implemented on the airport to minimise human error. In addition, during this project, several ideas concerning future simulation studies were suggested, but not realised as it would have gone beyond this simulation study. Approaches for further simulations studies could also be to work with the ideas given in this paper or to expand the given ideas. For example, only two road networks could be built, one for airplanes and one for ground vehicles, but with several turning lanes, one for turning right and one for turning left and for driving straight ahead. Another idea could be to expand the cellular structure on to all roads. If all vehicles of the airport are equipped with a location-tracking device, a combination of both methods could increase security on the apron and allow better presumptions, which vehicle will be where at what time, ensuring a smoother vehicle flow and a better scheduling.

References

1. Weltweite Anzahl der Flugpassagiere bis 2035 | Prognose. https://de.statista.com/statistik/daten/studie/374860/umfrage/flugverkehr-entwicklung-passagiere-weltweit/. Accessed 15 Sept 2017
2. Growth in total passenger air transport by Member State. http://ec.europa.eu/eurostat/statistics-explained/index.php/File:2014_2015_growth_in_total_passenger_air_transport_by_Member_State.png. Accessed 15 Sept 2017

3. Schlegel, A.: Bodenabfertigungsprozesse im Luftverkehr. Gabler Verlag/GWV Fachverlage GmbH Wiesbaden, Wiesbaden (2010)
4. Yearbook 2016 (2016)
5. Technical information. http://www.riga-airport.com/en/main/about-company/facts-about-rix/technical-information. Accessed 27 Sept 2017
6. 2014. http://www.riga-airport.com/en/main/about-company/facts-about-rix/history. Accessed 27 Sept 2017
7. OpenStreetMap. http://www.openstreetmap.org/search?query=Riga%20International%20Airport#map=16/56.9260/23.9855. Accessed 27 Sept 2017
8. Jenney, L.L.: Airport System Development. Science, Transportation, Innovation Program, Washington D.C. (1984)
9. Ashford, N.J., Stanton, H.P.M., Moore, C.A., Coutu, P., Beasley, J.R.: Airport Operations, 3rd edn. The McGraw-Hill Companies, United States of America (1997–2015)
10. 737 Airplane Characteristics for Airport Planning. http://www.boeing.com/assets/pdf/commercial/airports/acaps/737.pdf. Accessed 15 Sept 2017
11. Airbus A320 Aircraft Characteristics. http://www.aircraft.airbus.com/fileadmin/media_gallery/files/tech_data/AC/Airbus_AC_A320_May17.pdf. Accessed 15 Sept 2017
12. Norin, A.: Airport Logistics, Department of Science and Technology. Linköping University, Norrköping (2008)
13. Fuelling Hydrants Operating at Riga Airport. http://www.riga-airport.com/en/main/newsroom/news/fuelling-hydrants-operating-at-riga-airport. Accessed 26 Sept 2017
14. Mensen, H.: Handbuch der Luftfahrt, 2., neu bearb. Aufl. 2013. Springer, Heidelberg, s.l. (2013)
15. Use of Simulation. https://www.anylogic.com/use-of-simulation/. Accessed 02 Oct 2017

A Thorough Review of Big Data Sources and Sets Used in Transportation Research

Maria Karatsoli(✉) and Eftihia Nathanail

Department of Civil Engineering, University of Thessaly,
Pedion Areos, 38334 Volos, Greece
{makarats, enath}@uth.gr

Abstract. The development of Information and Communications Technology (ICT) and the Internet provide Intelligent Transport Systems (ITS) with a huge amount of real-time data. These data are the so-called "Big Data" which can be collected, interpreted, managed and analyzed in a proper way in order to improve the knowledge around the transport system. The use of these technologies has greatly enhanced the efficiency and user friendliness of ITS, providing significant economic and social impacts, contributing positively to the management of sustainable mobility.

In this paper, different sources of big data that have been used in ITS are presented, while their advantages and limitations are further discussed. Analytically, big data sources that have been used within the last 10 years are identified. Then, a review of current applications is done, in order to disclose the most used and proper data source per case.

Aim of the present study is to improve the knowledge around the usage of big data in transport planning and to contribute to the better support of ITS, by providing a roadmap to decision makers for big data collection methods.

Keywords: Data collection · Intelligent Transport Systems
Information and Communications Technology · Big data classification
Traffic information · Real-time data

1 Introduction

ITS have been developed in order to support decision making based on continuously collected traffic data. With the collection of these data, complex sets of massive volumes are created and processed in order to produce useful information. These sets, which are called big data, provide a comprehensive and clear description of the situation and can contribute in traffic, freight and mass transport management [83].

Big data refers to all those data whose scale, diversity and complexity require new analysis techniques and algorithms [14]. De Mauro [19] defines big data as an information asset with high Volume, Velocity and Variety that require specific methods in order to be transformed into value. Except for the three V-characteristics, big data demand also Value and Veracity in order to be reliable and accurate [24, 35]. The detailed definition of the five Vs can be found below:

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_52

- *Volume*: is a basic dimension of big data and measures the amount of data that are available. Big data are huge datasets characterized by the large volume of data generated by different sources, following currently an exponential increase [20, 63, 75];
- *Velocity:* measures the speed in which the data are generated, retrieved, aggregated and stored [4];
- *Variety*: is an important characteristic of big data and refers to the structure of them. Data can be structured, semi structured, unstructured or mixed [61];
- *Value:* is related to the extraction of useful data from a large dataset, generated from different sources [37];
- *Veracity:* refers to the uncertainty of the data which should be eliminated from any abnormalities and noises [4] in order to assure the reliability and accuracy of the dataset [13].

The value of big data in transport is undoubted since their use reduces the costs for infrastructure and service operators, and provides new means of solving complex transport problems [56]. The analysis and visualization of real-time conditions of transport networks, the optimization of routes and schedules, the analysis of accidents, and the management of public transport data, improve transportation systems in terms of minimizing congestion, improve traveler information and traveler assistance service, "understand" the commuters' needs, increase road safety and enhance overall efficiency [36, 68]. In this study, an overview of big data sources matched with the field they serve is conducted.

2 Types of Big Data Sources in Transport

The growth of technology increased the amount of available data in the transport sector. In this paper, potential sources of transport big data that can be used in traffic management, public transport operation and traveler information are identified. Mobile phones, Global Positioning System (GPS), social media, smart card systems and other systems generate big data and can shed light on transport systems operation. Each type of data has specific advantages and limitations for transportation. The advantages and limitations of the aforementioned data sources are shown in Table 1.

GPS Devices. GPS devices are used for the collection of location trajectories for the extraction of mobility patterns, through the periodical report of their position and time. GPS devices can be found on vehicles [65], bikes [31] or commuters [84]. The GPS data consist of the device ID, a time stamp based on the recording, the location of the GPS device at the time of recording, as well as the speed and direction of the vehicle. This technology is often used for traffic flow analysis, route planning, as well as to monitor public transport operations [2].

Mobile Phones. Network carriers collect mobile phone data for billing and other operational reasons. Those data include the time and date of each mobile phone activity, the phone number as well as the coordinates of the tower which routes the communication. The wide coverage and the long observational period of these data

Table 1. Advantages and limitations of big data sources

BD source	Advantages	Disadvantages
GPS [84]	– Passive collection (no participation of user required) – Accurate location information – High sampling frequency – Large sample size of public vehicles (i.e. taxis)	– Lack of social-economic, demographic, or social network information about users – Small sample size of private vehicles or individuals – High transmission costs
Mobile phone data [10]	– Low cost – Large sample – Broad spatial and temporal coverage – Long observation periods	– Missed travel activity between two phone activities, spatial uncertainties – Lack of social-economic or demographic information – Not suitable for estimating travel time on roads – Sparse and noisy measurements
Smart Card Data [5]	– Identification of transport mode – Large volumes of personal travel data – Access to continuous trip data covering longer periods of time – Credible reflection of transit demand	– Constrained to specific traffic mode – Sometimes unknown destination stop because passengers are required only to check in – Informal nature of textual content – Scarce location-related information
Social Media [18]	– Exact location of user – Large sample of qualitative data i.e. social-economic, demographic, and social network information – Highly active data – Useful data for evaluation stages – Low cost	– Lack of travel information between two active posts – Unknown transport mode – Increase of fake and bot-accounts
Points of Interest [2]	– Used as a supplement to other sources – Low cost – Easily accessible	– Limited information – Cannot be used autonomously

lead to a more accurate estimation of the location (home, work, leisure activities location) for a rich sample of residents [29].

Smart Card Data. The main reason for the introduction of a smart card system for public transport is the secure fare collection. However, each transaction offers valuable information regarding the mobility patterns of a city's public transport, that is useful in city planning. The collected data include the card ID, the transportation mode (bus, metro), the route, boarding/alighting times as well as boarding/alighting stations [2].

Point Detectors. Point Detectors are permanently installed and monitor the vehicle speed, the distance between them, CO2 emissions as well as occupancy [2].

Social Media. Social media has become a source of valuable information. Their wide spread encourages the users to share their location more often, leading to an exponential increase of their volume day by day. The social media users share publicly information (tweets, videos, photos, check-ins) on platforms such as Twitter, Facebook, Google+, Foursquare rendering them powerful tools, suitable for transport data collection [43].

Points of Interest. Points of Interest are businesses and important places in a city. Yellow pages and Google Places or applications such as Trip Advisor are the main sources of this kind of data. Information such as the opening hours, reviews and crowdedness per hour of a place are some examples of the type of data that can be retrieved [2].

3 Review of Data Sources and Analysis

The goal of this review is to identify the most frequent big data source used in transportation research and disclose in which application field these sources have contributed the most. The analyzed transport studies applied on the fields of traffic management (TM), operation of public transport (PT) and travel behavior/ accessibility (TB/A). Table 2 summarizes the big data sources, sorted alphabetically based on author's name, that have been used in 63 previous studies which were published within the last decade. The review was performed on conference proceedings and peer-reviewed journals based on the following list of keywords and their combination: "big data", "transport studies", "mobility", "social media", "smart cart data", "big data sources", "GPS data", and "mobile phone data". The data sources considered are Point Detectors (PD), Mobile Phones (MP), Smart Card data (SC), Social Media (SM), Points of Interest (PoI) and Other methods (O), e.g. logins to public WiFi, financial transactions etc. Although this is a non-exhaustive list of studies, the authors believe that this particular sample represents adequately the current trends. The selected 63 studies describe thoroughly the use of the selected big data source, have a clear scope and focus on one or more of the three mentioned fields.

According to the analysis, the most frequent big data source is GPS in the traffic management field, smart card data in public transport operation and mobility phone data in accessibility/travel behavior studies. The lower percentages are demonstrated in sources of qualitative data such as social media and points of interest. The 81% of the studies used one big data source, 14% used two sources and the rest used three sources. In 42% of the studies, where data from more than one source are derived, big data used along with data from point detectors, while half of them used qualitative data as a supplement to other big data sources. The majority of the research studies focused on the analysis of travel behavior and the enhancement of accessibility.

Table 2. Classification of studies regarding big data in transport.

Study	Big data source							Application		
	GPS	PD	MP	SC	SM	PoI	Other	TM	PT	TB/A
1	2	3	4	5	6	7	8	9	10	11
Amin et al. [1]	X							X		
Artikis et al. [3]		X			X			X		
Barrow [6]				X					X	
Bekhor et al. [7]			X							X
Bertrand et al. [8]					X			X		X
Bien et al. [9]	X	X						X	X	
Calabreze et al. [11]			X			X				X
Castro et al. [12]	X							X		
Chao et al. [15]	X									X
Cheng et al. [16]					X					X
Christian et al. [17]			X							X
Dewulf et al. [21]	X							X		
Digital Bonanza [22]					X					X
Eggermond et al. [23]				X	X					X
Eom et al. [25]				X					X	
Furletti et al. [26]			X							X
Ge et al. [27]	X							X		
Gokasar et al. [28]				X					X	X
Gonzales et al. [29]			X							X
He et al. [30]				X					X	
Hood et al. [31]	X									X
IMDA [32]							X			X
Iqbal et al. [33]			X							X
ITS China [34]			X					X	X	X
Jagadish et al. [35]	X	X				X		X		
Lin et al. [38]					X			X		
Long et al. [39]				X						X
Long et al. [40]					X					X
Ma et al. [41]				X						X
Moller-Jensen et al. [42]	X							X		
Munizaga et al. [45]				X						X
Munizaga et al. [44]	X			X					X	
Network Rail [46]	X		X					X		
Noulas et al. [47]			X		X	X				X
Owen et al. [48]	X	X								X
Pan et al. [49]	X							X		
Pang et al. [50]	X							X		
Papacharalampous [51]			X							X
Pelletier et al. [52]				X					X	
Phithakkitnukoon et al. [53]						X				X

(*continued*)

Table 2. (*continued*)

Study	Big data source							Application		
	GPS	PD	MP	SC	SM	PoI	Other	TM	PT	TB/A
Romph [54]			X							X
Roth et al. [55]				X						X
Santi et al. [57]	X							X		
Schmoecker et al. [58]				X						X
Schulz et al. [59]					X			X		
Seaborn et al. [60]				X					X	
Song et al. [62]			X							X
Tabbitt [64]	X	X			X			X	X	
Toole et al. [66]			X							X
Trepanier et al. [67]				X						X
van Oort et al. [69]	X			X					X	
Wang et al. [70]			X							X
Wang et al. [72]	X							X		
Wang et al. [71]					X			X		
Wanichayapong et al. [73]					X			X		
Wanq et al. [74]	X							X		
Weinstein [76]	X			X					X	
Widhalm et al. [77]			X							X
Wood et al. [78]					X				X	
Yeung et al. [79]				X						X
Yu et al. [80]				X				X	X	X
Yuan et al. [81]	X							X		
Zhang et al. [82]	X									X
Total number	22	5	15	18	13	4	1	23	14	35
Percentage	35%	8%	24%	29%	21%	6%	2%	37%	22%	56%

4 Conclusion

Big data in transport ensures good knowledge of the system and manages to achieve key objectives, such as increasing travel security and travel comfort and reducing travel time and costs. Big data sources' development goes in hand with new technologies' development. This paper analyzed the current big data sources and their characteristics in transport and revealed the most frequent data source which is GPS. In most of the cases, the extracted big data were oriented to enhance accessibility/ travel behavior studies and support traffic management. As studies for big data sources and applications in the transport sector have tremendously increased, authors believe that the incorporation of additional studies to this review in the future will bring new information for better understanding and new possibilities for more efficient management of transport systems.

Acknowledgements. This work has been supported by the ALLIANCE project (http://alliance-project.eu/) and has been funded within the European Commission's H2020 Programme under contract number 692426. This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

References

1. Amin, S., Andrews, S., Apte, S., Arnold, J., Ban, J., Benko, M., Bayen, R.M., Chiou, B., Claudel, C., et al.: Mobile century using GPS mobile phones as traffic sensors: a field experiment, pp. 16–20 (2008)
2. Anda, C., Fourie, P., Erath, A.: Transport modelling in the age of big data. In: Future Cities Laboratory (2016)
3. Artikis, A., et al.: Self-Adaptive Event Recognition for Intelligent Transport Management, pp. 319–325 (2013)
4. Arun, K., Jabasheela, L.: Big data: review, classification and analysis survey. Int. J. Innovative Res. Inf. Secur. (IJIRIS) **1**(3), 17–23 (2014)
5. Bagchi, M., White, P.: The potential of public transport smart card data. Transp. Policy **12** (5), 464–474 (2005)
6. Barrow, K.: Big Data predicts train delays before they occur. http://www.railjournal.com/index.php/commuter-rail/big-data-predicts-train-delays-before-they-occur.html. Accessed 11 Aug 2017
7. Bekhor, S., Cohen, Y., Solomon, S.: Evaluating long-distance travel patterns in Israel by tracking cellular phone positions. J. Adv. Transp. **47**, 435–446 (2013)
8. Bertrand, K.Z., Bialik, M., Virdee, K., Gros, A, Bar-Yam, Y.: Sentiment in New York City: a high resolution spatial and temporal view, New England complex systems institute, Cambridge, United States (2013). http://www.necsi.edu/research/social/newyork/
9. Biem, A., Bouillet, E., Feng, H., Ranganathan, A., Riabov, A., Verscheure, O., Koutsopoulos, H., Moran, C.: IBM InfoSphere streams for scalable, real-time, intelligent transportation services. In: SIGMOD 2010, 6–11 June, Indianapolis, Indiana, USA (2010)
10. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr., J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transp. Res. Part C Emerg. Technol. **26**, 301–313 (2013)
11. Calabrese, F., Lorenzo, G.D., Liu, L., Ratti, C.: Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Comput. **10**(4), 36–44 (2011). ISSN 1536-1268
12. Castro, P., Zhang, D., Li, S.: Urban traffic modelling and prediction using large scale taxi GPS traces. In: Kay, J., Lukowicz, P., Tokuda, H., Olivier, P., Krüger, A. (eds.) Pervasive Computing, pp. 57–72, Berlin, Heidelberg (2012)
13. Chandio, A.A., Tziritas, N., Xu, C.-Z.: Big-data processing techniques and their challenges in transport domain (2015)
14. Chandrasekar, P.: Big data and transport modeling: opportunities and challenges. Int. J. Appl. Eng. Res. **10**(17), 38038–38044 (2015). ISSN 0973-4562
15. Chao, C., Daqing, Z., Zhi-Hua, Z., Nan, L., Atmaca, T., Shijian, L.: B-Planner: night bus route planning using large-scale taxi GPS traces. In: 2013 IEEE International Conference on Pervasive Computing and Communications (PerCom) (2013)
16. Cheng, Z., Caverlee, J., Lee, K., Sui, D.: Exploring millions of footprints in location sharing services. In: Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media, Barcelona, Spain (2011)

17. Christian, M., Schneider, V.B., Couronne, T., Smoreda, Z., Gonzalez, M.C.: Unraveling daily human mobility motifs. J. R. Soc. Interface **10**, 20130246 (2013)
18. Cici, B., Markopoulou, A., Frias-Martinez, E., Laoutaris, N.: Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, Washington, 2632055, pp. 201–211. ACM (2014)
19. De Mauro, A., Greco, M., Grimaldi, M.: A formal definition of big data based on its essential features. Libr. Rev. **65**(3), 122–135 (2016)
20. Demchenko, Y., Laat, C.D., Membrey, P.: Defining architecture components of the big data ecosystem. In: Proceedings of International Conference Collaboration Technologies and Systems (CTS 2014), pp. 104–112 (2014)
21. Dewulf, B., Neutens, T., Vanlommel, M., Logghe, S., De Maeyer, P., Witlox, F.: Examining commuting patterns using floating car data and circular statistics: exploring the use of new methods and visualizations to study travel times. J. Transp. Geogr. **48**, 41–51 (2015)
22. Digital Bonanza – Cover Story: Binghamton Research Magazine, Winter Issue, pp. 12–19 (2014)
23. Eggermond, M., Chen, H., Erath, A., Cebrian, M.: Investigating the potential of social network data for transport demand models. In: Transportation Research Board 95th Annual Meeting, United States (2015)
24. Emani, C.K., Cullot, N., Nicolle, C.: Understandable big data: a survey. Comput. Sci. Rev. **17**, 70–81 (2015)
25. Eom, J., Song, J., Moon, D.-S.: Analysis of public transit service performance using transit smart card data in Seoul. KSCE J. Civ. Eng. **19**, 1–8 (2015)
26. Furletti, B., Gabrielli, L., Renso, C., Rinzivillo, S.: Analysis of GSM calls data for understanding user mobility behavior. In: IEEE International Conference on Big Data, United States (2013)
27. Ge, Y., et al.: An energy-efficient mobile recommender system. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010), p. 899. ACM Press, New York (2010)
28. Gokasar, I., Simsek, K., Ozbay, K.: Using big data of automated fare collection system for analysis and improvement of BRT-Bus rapid transit line in Istanbul. In: Transportation Research Board 94th Annual Meeting, United States (2014)
29. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
30. He, K., Wang, J., Deng, L., Wang, P.: Congestion avoidance routing in urban rail transit networks. In: 2014. IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), pp. 200–205. IEEE (2014)
31. Hood, J., Sall, E., Charlton, B.: A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. **3**(1), 63–75 (2011)
32. IMDA Infocom Media Development Authority: Smart Nation big on Big Data 14. https://www.imda.gov.sg/infocomm-and-media-news/buzz-central/2016/6/smart-nation-big-on-big-data. Accessed 23 July 2017
33. Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C.: Development of origin–destination matrices using mobile phone call data. Transp. Res. Part C Emerg. Technol. **40**, 63–74 (2014)
34. ITS China. http://www.itschina.org/article.asp?articleid=2259. Accessed 15 May 2015
35. Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. Commun. ACM **7**(57), 86–94 (2014)

36. Ju, G., Cheng, M., Xiao, M., Xu, J., Pan, K., Wang, X., Shi, F.: Smart transportation between three phases through a stimulus-responsive functionally cooperating device. Adv. Mater. **25**(21), 2915–2919 (2015)
37. Kemp, G., Vargas-Solar, G., Da Silva, C.F., Ghodous, P., Collet, C.: Aggregating and managing big real-time data in the cloud: application to intelligent transport for smart cities. In: Proceedings of the 1st International Conference on Vehicle Technology and Intelligent Transport Systems, pp. 107–112, Lisbon, Portugal (2015)
38. Lin, J., Ryaboy, D.: Scaling big data mining infrastructure: the Twitter experience. ACM SIGKDD Explor. Newslett. **14**(2), 6 (2013)
39. Long, Y., Zhang, Y., Cui, C.: Identifying commuting pattern of beijing using bus smart card data. J. Geogr. Sci. **67**, 1339–1352 (2012)
40. Long, Y., Han, H., Tu, Y., Shu, X.: Evaluating the effectiveness of urban growth boundaries using human mobility and activity records. Cities **46**, 76–84 (2015)
41. Ma, X., Wu, Y.J., Wang, Y., Chen, F., Liu, J.: Mining smart card data for transit riders' travel patterns. Transp. Res. Part C Emerg. Technol. **36**, 1–12 (2013)
42. Møller-Jensen, L., Kofie, R.Y., Allotey, A.N.: Measuring accessibility and congestion in Accra. Norsk Geografisk Tidsskrift-Norwegian. Geogr. **66**(1), 52–60 (2012)
43. Morstatter, F., Pfeffer, J., Liu, H., Carley, K.M.: Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In: ICWSM 2013, June 21; cs. SI (2013)
44. Munizaga, A.N.: Using smart card and GPS data for policy and planning: the case of Transantiago. Res. Transp. Econ. **59**, 242–249 (2016)
45. Munizaga, M., Palma, C.: Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smart card data from Santiago, Chile. Transp. Res. Part C Emerg. Technol. **24**, 9–18 (2012)
46. Network Rail: Asset Management Services (2013)
47. Noulas, A., Mascolo, C.: Exploiting foursquare and cellular data to infer user activity in urban environments. In: 2013 IEEE 14th International Conference on Paper presented at the Mobile Data Management (MDM), vol. 1, pp. 167–176 (2013)
48. Owen, A., Levinson, D.M.: Modeling the commute mode share of transit using continuous accessibility to jobs. Transp. Res. Part A Policy Pract. **74**, 110–122 (2015)
49. Pan, B., Zheng, Y., Wilkie, D., Shahabi, C.: Crowd sensing of traffic anomalies based on human mobility and social media. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 344–353. ACM (2013)
50. Pang, L.X., Chawla, S., Liu, W., Zheng, Y.: On detection of emerging anomalous traffic patterns using GPS data. Data Knowl. Eng. **87**, 357–373 (2013)
51. Papacharalampous, A.E.: Aggregated GSM data in origin destination studies. Masters' thesis. Technical University of Delft, Netherlands (2014)
52. Pelletier, M., Trépanier, M., Morency, C.: Smart card data use in public transit: a literature review. Transp. Res. Part C Emerg. Technol. **19**(4), 557–568 (2011)
53. Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C.: Activity aware map: identifying human daily activity pattern using mobile phone data. In: Human Behavior Understanding, pp. 14–25. Springer (2010)
54. Romph, E.: Using big data in transport modelling. Data Model. Magaz. **10**, Summer Issue (2013)
55. Roth, C., Kang, S.M., Batty, M., Barthelemy, M.: Structure of urban movements: polycentric activity and entangled hierarchical flows. PLoS One **6**(1), 1–8 (2011)
56. Rusitschka, S., Curry, E.: Big data in the energy and transport sectors. In: New Horizons for a Data-Driven Economy, pp. 225–244 (2015). Chapter 13

57. Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C.: Quantifying the benefits of vehicle pooling with shareability networks. Proc. Natl. Acad. Sci. **111**(37), 13290–13294 (2014)
58. Schmöcker, J.D., Shimamoto, H., Kurauchi, F.: Generation and calibration of transit hyperpaths. Transp. Res. Part C Emerg. Technol. **36**, 406–418 (2013)
59. Schulz, A., Ristoski, P., Paulheim, H.: I see a car crash: real-time detection of small scale incidents in microblogs. In: The Semantic Web: ESWC 2013 Satellite Events, pp. 22–33. Springer (2013)
60. Seaborn, C., Attanucci, J., Wilson, N.H.M.: Analyzing multimodal public transport journeys in London with smart card fare payment data. Transp. Res. Rec. J Transp. Res. Board **2121**, 55–62 (2009)
61. Sharma, S.: Expanded cloud plumes hiding big data ecosystem. Future Gener. Comput. Syst. **59**, 63–92 (2016)
62. Song, C., Qu, Z., Blumm, N., Barabasi, A.-L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)
63. Swan, M.: Philosophy of big data: expanding the human-data relation with big data science services. In: Proceedings of First International IEEE Conference of Big Data Computing Service and Applications, pp. 468–477 (2015)
64. Tabbitt, S.: Big data analytics keeps Dublin moving. http://www.telegraph.co.uk/sponsored/sport/rugby-trytracker/10630406/ibm-big-dataanalytics-dublin.html. Accessed 6 May 2015
65. Tamor, M.A., Gearhart, C., Soto, C.: A statistical approach to estimating acceptance of electric vehicles and electrification of personal transportation. Transp. Res. Part C Emerg. Technol. **26**, 125–134 (2013)
66. Toole, J.L., Ulm, M., Gonz, M.C., Bauer, D.I.: Inferring land use from mobile phone activity. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing, pp. 1–8, Beijing, China (2012)
67. Trépanier, M., Tranchant, N., Chapleau, R.: Individual trip destination estimation in a transit smart card automated fare collection system. J. Intell. Transp. Syst. **11**, 1–14 (2007)
68. van Oort, N., Brands, T., de Romph, E.: Short-term prediction of ridership on public transport with smart card data Transp. Res. Rec. J. Transp. Res. Board **2535**, 105–111 (2015)
69. van Oort, N., Cats, O.: Improving public transport decision making, planning and operations by using big data cases from Sweden and the Netherlands. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC) (2015)
70. Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., Gonzalez, M.C.: Understanding road usage patterns in urban areas. Sci. Rep. **2**, 1001 (2012)
71. Wang, X., Zeng, K., Zhao, X.L., Wang, F.Y.: Using web data to enhance traffic situation awareness. In: IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), pp. 195–199. IEEE (2014b)
72. Wang, Y., Zheng, Y., Xue, Y.: Travel time estimation of a path using sparse trajectories. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 25–34. ACM (2014)
73. Wanichayapong, N., Pruthipunyaskul, W., Pattara-Atikom, W., Chaovalit, P.: Social-based traffic information extraction and classification. In: 11th International Conference on ITS Telecommunications (ITST), pp. 107–112. IEEE (2011)
74. Wanq, Q., Taylor, J.E.: Quantifying Human Mobility Perturbation and Resilience in Hurricane Sandy (2014)
75. Watson, H.J.: Tutorial: big data analytics: concepts, technology, and applications. Assoc. Inf. Syst. **34**, 5–16 (2014)

76. Weinstein, S.L.: Innovations in London's transport: big data for a better customer experience. http://2015.data-forum.eu/sites/default/files/1600-1640%20Weinstein_SEC.pdf. Accessed 20 Aug 2017
77. Widhalm, P., Yang, Y., Ulm, M., Athavale, S., González, M.: Discovering urban activity patterns in cell phone data. Transportation **42**, 1–27 (2015)
78. Wood, S.A., Guerry, A.D., Silver, J.M., Lacayo, M.: Using social media to quantify nature-based tourism and recreation. Sci. Rep. **3**, 2976 (2013)
79. Yeung, C.H., Saad, D., Wong, K.M.: From the physics of interacting polymers to optimizing routes on the London underground. Proc. Natl. Acad. Sci. **110**(34), 13717–13722 (2013)
80. Yu, W., Mao, M., Wang, B., Liu, X.: Implementation evaluation of Beijing urban master plan based on subway transit smart card data. In: 22nd International Conference on Geoinformatics, Kaohsiung, Taiwan (2014)
81. Yuan, N.J., et al.: T-finder: a recommender system for finding passengers and vacant taxis. IEEE Trans. Knowl. Data Eng. **25**, 2390–2403 (2013)
82. Zhang, W., Qi, G., Pan, G., Lu, H., Li, S., Wu, Z.: City-scale social event detection and evaluation with taxi traces. ACM Trans. Intell. Syst. Technol. **6**(3), 1–20 (2015)
83. Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., Zhang, Q., Yang, L.: Big data for social transportation. IEEE Trans. Intell. Transp. Syst. **17**(3), 620–630 (2016)
84. Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y.: Mining interesting locations and travel sequences from GPS trajectorie. In: Proceedings of International Conference on World Wide Web (WWW 2009), Madrid, Spain, pp. 791–800. ACM Press (2009)

Evaluating Smart Urban Freight Solutions Using Microsimulation

Ioannis Karakikes[1(✉)], Lambros Mitropoulos[1], and Mihails Savrasovs[2]

[1] Department of Civil Engineering, University of Thessaly, Pedion Areos, 38334 Volos, Greece
iokaraki@uth.gr

[2] Transport and Telecommunication Institute, Lomonosova street 1, Riga 1019, Latvia

Abstract. Last mile distribution remains a difficult-to-solve variable in urban congestion's equation, especially in Europe, due to increased population, economic growth and limited space. Over the last decades, several European projects have contributed significantly into that direction, by developing innovative concepts (e.g., electric solutions, ITS adoption, effective policy-based strategies). A great number of measures has been deployed and considered as possible solutions to the last mile distribution problem of European cities, however, only a few of them have actually been implemented and tested over a long period of time and their impacts have been quantified.

This study focuses on the evaluation of three smart urban freight transport measures on an urban interchange – Commercial port – by using a microscopic traffic simulation tool in order to decide which is the most effective in environmental and transportation terms. Each measure is being evaluated as if it was to be implemented now (2017) and in 2030 in order to assess measures' effectiveness in the short as well in the long term. The analysis is completed by using a multi-criteria multi-stakeholder decision making tool to generate the Logistics Sustainability Index (LSI) for each measure, to summarize results and provide a sustainability based rating to support local decision-making.

Keywords: City logistics · Evaluation · Micro simulation
Logistics Sustainability Index

1 Introduction

Urban freight transport constitutes one of the biggest problems that modern cities have to deal with today, as it can produce many adverse impacts (economic, environmental, societal and transport) and deteriorate the quality of life for citizens of urban areas. Today, there is a wide-range toolkit with measures for transport engineers [1] that stems from various European projects (ECCENTRIC, TRENDSLETTER, TELLUS, ARCHIMEDES, 2MOVE2, MIMOSA, DESTINATIONS, SUCCESS, VIVALDI, MOBILIS, MODERN, NOVELOG) that restricts the negative impacts of urban freight logistics and provide sustainable solutions. However, not every measure can be implemented in every case and not every measure can be to the same extent successful. Every city, district or neighbourhood has its own unique characteristics and in order to achieve the optimum

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_53

result and successfully be implemented, all proposed measures should be carefully studied before implementation.

Simulation provide a safe way to decision makers to test and explore a proposed measure under different criteria/indicators and conditions, taking into account specific attributes such as the transport infrastructure, local driving regulations, modal split, traffic volumes etc. This study is aiming at evaluating three logistic measures through a case study in a medium sized city in Greece, in order to reveal the “best” measure, based on environmental and transport related indicators. The evaluation will be performed for years 2017 and 2030 so as to assess measures’ effectiveness in the short as well in the long term. The selection of the measures and indicators is limited to the restrictions of this particular example, however, the performed analysis and the methodology that is followed can be considered as a guidebook for future practitioners or stakeholder categories for evaluating smart logistic solutions by using microsimulation tools.

2 Methodology

The evaluation of the measures was performed by modeling the traffic network around the commercial port of the city of Volos with the highest accuracy possible in order to reproduce the transport system characteristics. In this study, measures for evaluation were selected based on two conditions:

- The ability to be simulated by using a traffic simulation tool,
- To be classified as a soft measure (i.e. non-structural interventions are required).

The evaluation of measures was performed by using indicators that are quantified by using the proposed microsimulation tools. For the purpose of this study, indicators are grouped into two impact areas, namely the Environment and Transport Performance to highlight their significance when planning for transportation solutions. Due to multiple indicators and the presence of different stakeholders in planning agencies, evaluation of measures becomes challenging, thus the usage of a single index that incorporate the information of all indicators per measure is required. A multi-criteria multi-stakeholder decision-making tool – Evalog – aiming at assessing the performance of smart urban logistics measures is used in this study in order to aggregate indicators per measure into a single index. Evalog incorporates a multiple weighting scheme, and elimination and ranking techniques and models, for the facilitation of decision-making, by considering the participation, viewpoint, and contribution of all involved stakeholders to the confirmation of the final decision. Additionally, the evaluation process considers life cycle (creation, through operation and maintenance to closure) impacts of the assessed measures [2], however only operation stage is considered in this study. The output of Evalog is a Logistics Sustainability Index (LSI) that is used for comparing the sustainability performance of the three proposed measures.

3 Application in Smart Urban Freight Measures

3.1 Study Area

The study area, which is composed of a commercial port and a road segment, is located in the city of Volos in Greece. The commercial port is vital for the economic life of the city since it serves the needs of the whole region of Thessaly. Figure 1 illustrates the cargo shipment loads of the port during the last decade.

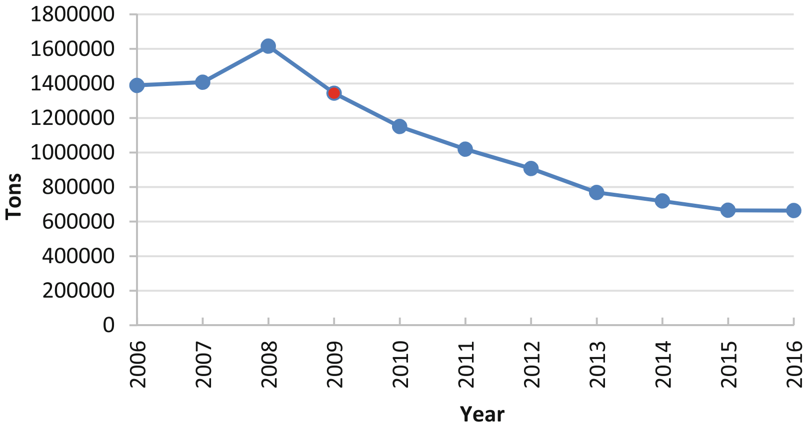


Fig. 1. Cargo shipment loads at Volos commercial port [3].

The commercial port is connected with an interurban road that serves the industrial area of the city with a corridor that runs through a mixed-use area and has a length of 1.495 km. The traffic volumes on this corridor during the peak hour (09:45 am–10:45 am) are on average 450 veh/h per lane per direction (2 lanes per direction) and the percentage of Heavy Good Vehicles (HGVs) is 4% based on on-site measurements which were realized on 07/06/2017. The study area is of high interest since a further development of the port is anticipated by 2030 [4], that will lead to increased traffic flows and emissions. To ensure the sustainable transport of goods in 2030, as well as to quantify the impacts of the suggested measures and make safe proposals, a traffic volume forecast has been carried out for further investigation.

The forecast model considers the induced traffic due to the growth of the city of Volos. The city of Volos attracts approximately 300 new residents per year, which corresponds to 0.2% increase of the population [5]. Assuming that the increase of traffic is affected by the increase of residents and economic status, a 0.5% increase of traffic is considered. The forecasted volume for the year 2030 is estimated as follows (Eq. 1):

$$F = (1 + r), \tag{1}$$

where F is the future volume, P is the present volume, n is the number of years and r the growth factor per year.

Except for the residents and economic status, an additional increase should be considered due to the increase of cargo volumes transferred to/from the commercial port of Volos by the year 2030. This increase in cargo volumes will be reflected in our model by a higher number of HGVs entering/exiting the commercial port. Given that the current volumes (current volumes are considered equal with the volumes of the year 2016, since no data are available for the year 2017 by the Port Authority) that are served by port's authorities are equal to 663,490 tons and that in 2030 the cargo volumes are expected to be equal to the volumes of 2009 (1,343,908 tons) (see Fig. 1) [6], the number of HGVs is expected to increase by the same share, meaning 103%, which corresponds to 164 additional HGVs/h entering/exiting the commercial port.

Finally, vehicle share by type for 2017 and 2030 was estimated based on vehicle sales data from the [5, 7] and [8]. It was assumed that changes in land use and other factors did not influence the trip lengths in 2017 and 2030. The vehicle mix of HGV and passenger vehicles for the city of Volos is currently composed of 4 different engine technologies: petrol, diesel, Compassed Natural Gas (CNG) and electric (see Table 2).

3.2 Traffic Model Development

To evaluate the impacts of the smart urban freight measures, a microsimulation model was designed in Vissim, replicating the traffic on the corridor that connects Volos' commercial port with the interurban road. For the development of the model counted traffic volumes were derived from on-site observation measurements at specific locations. All the rest operational elements of the model have been determined either from google maps or on-site observation. The base model was then calibrated to ensure a validated representation of traffic patterns. The microsimulation model parameters that were used as calibration parameters are:

- Speed distribution,
- Reduced speed areas,
- Car-following model parameters.

The traffic volumes counted by on-site observation were the reference and basis for the calibration process. Also, the comparison of the on-site and modeled queue lengths (based on observation) is a secondary indicator of the calibration process although no quantitative analysis is possible in this case. The calibration parameters listed above were altered through an iterative process until the GEH (Geoffrey E. Havers) statistic for the sum of all link flows reaches a value less than 4 [9, 10]. The formula for the computation of the "GEH statistic" is the following:

$$\mathrm{GEH} = \sqrt{\frac{2(M - C)^2}{M + C}}, \tag{2}$$

where M is the hourly traffic volume exported from the model and C the hourly counted volumes.

Table 1 illustrates the results of the GEH value based on the counted volumes and model's estimated volumes for each trial:

Table 1. The iterative process for calibrating the model.

Number of trials	GEH value
Trial 1 (Model Default)	5.1
Trial 2 (Calibration 1)	5.6
Trial 3 (Calibration 2)	4.2
Trial 4 (Calibration 3)	4.3
Trial 5 (Calibration 4)	4.1
Trial 6 (Calibrated Model)	3.6

3.3 Description of Logistic Measures

Three urban freight measures are evaluated as possible solutions in this study. The performance of the measures was evaluated by comparing the results of each measure with the results of the base model (without having implemented any new measures) for the two base years, 2017 and 2030.

Measure 1 – Increasing HGVs Load Factor. The first measure ("Measure 1") aims at increasing the load factor of the HGVs at the commercial port of Volos. According to estimations given by Volos port, the average load factor for HGVs approaches 85%. Although, most HGVs servicing the commercial port are container trucks, there is room for improvement for HGVs that transport bulk shipments. Such HGVs could achieve higher load factors if there was a real-time online system that would inform carriers about the exact volume of the remaining product which ultimately would indicate the optimum selection of truck type/size to carry it. Other factors that affect the load factor of HGVs can be poor programming and other unexpected events. The adoption of information systems in conjunction with better programming could end up to an average 95% load factor, which indicates a 5.5% reduction of HGVs entering/exiting the commercial port. For the analysis

Measure 2 – Alternative Fuel Vehicles. The second measure ("Measure 2") focuses on increasing the share of HGVs that are powered by alternative fuels (e.g. electric and compressed natural gas). Even though electric vehicles are used more often in last mile distribution and are small sized-vehicles, this measure explores the potential of alternative fuel for HGVs. Specifically, Measure 2 is used to evaluate emissions improvement in a "what if" scenario within which the share of HGVs that use green technologies increases by 5%. The share of HGV technologies for testing the measure is:

- Diesel HGV: 95%,
- Compressed Natural Gas HGV: 0.5%,
- Electric HGV: 4.5%.

Measure 2 is evaluated only in year 2017; HGVs mix for 2030 changes relative to 2017 based on local sales of alternative fuel vehicles and expected market penetration of alternative fuel vehicles, following EU projections. Therefore, base scenario for 2030 incorporates these assumptions and the vehicle mixes for both heavy-duty and light-duty vehicles are presented in Table 2.

Table 2. Parameters for EnViVer.

Parameters	Values	
Road type	Urban	
Vehicles newer than 1 year: %	5%	
Average vehicle age: (years)	13.5	
Average exit age (years)	27	
Euro legislation introduction:	Euro 1–1993 Euro 2–1996 Euro 3–2000 Euro 4–2005 Euro 5–2009 Euro 6–2014	
Vehicle type: Heavy-Duty	2017	2030
	Petrol: 0% Diesel: 100% CNG: 0% Electric: 0%	Petrol: 0% Diesel: 78% CNG: 2% Electric: 20%
Vehicle type: Light-duty	2017	2030
	Petrol: 92.53% Diesel: 6.41% CNG: 1.05% Electric: 0.01%	Petrol: 79.55% Diesel: 6.41% CNG: 1.05% Electric: 1.82%

Measure 3 – Enforcement and ITS Adoption for Control and Traffic Management. "Measure 3" focuses on enforcement and ITS adoption for control and traffic management. Specifically, Measure 3 improves the "green wave" for the three successive intersections on the tested corridor and suggests the adoption of Cooperative Intelligent Transport Systems (C-ITS) Services for the users of the road network around the port. Based on the calibrated simulation model, a better coordination is achieved by slightly offsetting earlier the signals of the last two intersections, by 4 and 5 s for 2017, and 8 and 6 s for 2030. This offset arose from the minimization of the percentage of vehicles that drive through the intersection without stopping, allowing loaded HGVs to avoid unnecessary decelerations and accelerations and achieve a smoother rolling with fewer emissions and lower fuel consumption. For the green wave all preconditions are covered [11]:

- Distance between intersections is 650 m < 800 m,
- More than one lane for through traffic,
- No parking spaces along the main direction,
- Coordination is possible in both directions,
- Cycle time of 72 s for all intersections.

C-ITS have the potential to improve the overall traffic situation without resulting in the construction of new transport infrastructure. Specifically, the introduction of Green

Light Optimal Speed Advice (GLOSA) can improve further the quality of traffic flow, through the reduction of the number of stopped vehicles at intersections equipped with a vehicle to infrastructure (V2I) communication. Vehicles that are equipped with GLOSA when approaching the intersection, they receive information regarding the switching times of the next traffic light. According to several relevant research projects like TRAVOLUTION, DRIVE C2X, EcoMove and Compass4D, if the driver follows this advice several advantages can be realized:

- Fuel consumption, emissions, and noise at the intersection area are reduced, since the vehicle may avoid stopping;
- Fuel consumption caused by inefficient, sudden deceleration due to a red traffic light is reduced;
- The efficiency of the intersection is increased and delays are reduced since freely flowing traffic has higher throughput than a stopped queue which is accelerating;
- Security at the intersection is increased for all users by eliminating the dilemma zone.

Although systems like GLOSA are difficult to be modeled, previous research projects showed 4.4% total reduction in carbon dioxide (CO_2) emissions, 5.3% in nitrogen oxides NO_x, 12.3% in fuel consumption 12.3% and up to 2.0% in travel times [12].

3.4 Environmental Model

For the evaluation of the measures certain environmental and traffic related indicators have been identified. The generated indicators' values exported by simulation have been compared with the ones generated for the base model in 2017 and 2030, in order to test which measure is the most efficient. The indicators that are used are:

- Environmental: CO_2 (g/km), NO_x (g/km), particle matter – PM_{10} (mg/km),
- Transport: Delays (seconds).

To estimate the environmental impact of the measures, the EnViVer software was used. EnViVer is based on VERSIT + exhaust emission model and has been designed by TNO. There is a significant number of publications demonstrating EnViVer software's successful application in practice [13–15].

EnViVer takes vehicles' trajectories data generated by Vissim for different vehicle classes (Light-duty, Bus and Heavy-duty), road types (Urban, Highway) and fuel types (petrol, diesel, LPG, CNG, electric) and provides accurate pollutant emissions measurements considering individual driving patterns (speed and acceleration data) and vehicles' properties (weight and power). In order to export reliable results, EnViVer takes into account additional data about the age distribution of the vehicles by setting up parameters, such as % of vehicles newer than 1 year, average vehicle age and average exit age. The determination of the aforementioned parameters together with emission legislation data produces the distribution of the vehicles by the euro norm. Table 3 below, demonstrates the input data used in EnViVer.

Table 3. Generated indicators' values from simulation per case.

Case	CO_2 (g/km)	NO_x (g/km)	PM_{10} (mg/km)	Delays (s)
2017				
Base model	338 543	1 212	67 918	39
Measure 1	336 512	1 210	67 714	38.5
Measure 2	334 112	1 176	66 575	39
Measure 3	332 409	1 198	67 377	36.1
2030				
Base model	456 453	2 113	104 317	75.8
Measure 1	436 538	1 985	98 267	62.7
Measure 3	428 604	1 987	98 678	46.9

4 Analysis and Results

The evaluation results of the model correspond to all vehicle classes, during the peak hours. The environmental indicators were measured for the whole network, while the delays represent the average delay of a section of the corridor that has a length of 1410 m. Table 3 shows the generated values of the selected indicators per case.

Table 4 shows the change of indicators per measure and ranks the measures based on the LSI. The LSI depicts the % change between the LSI of the base model in 2017 and 2030, and the LSI after having implementing each measure. For the computation of the LSI, all three environmental indicators were equally weighted (0.166), while for the traffic delay indicator a weight of 0.5 was attributed in order to evaluate equally the transport and environmental impact areas.

Table 4. Change of indicators per measure and ranking of measures based on LSI change

Ranking	Measure	CO_2	NO_x	PM_{10}	Delays	LSI change
Year 2017						
1st	Measure 3	–1.8%	–1.2%	–0.8%	–8.6%	4.7%
2nd	Measure 2	–1.3%	–3.0%	–2.0%	0.0%	1.0%
3rd	Measure 1	–0.6%	–0.2%	–0.3%	–2.5%	0.8%
Year 2030						
1st	Measure 3	–6.1%	–6.0%	–5.4%	–38.1%	28.2%
2nd	Measure 1	–4.4%	–6.1%	–5.8%	–17.3%	12.7%

According to Table 4, all measures had non negative impacts. The most effective measure in terms of CO2 emissions and traffic delay is Measure 3 both for 2017 and 2030. In terms of NOx and PM10 the most effective measure for the year 2017 is Measure 2 and for 2030 Measure 3. All measures in both years present the highest improvement in traffic delay and the least in CO2 emissions. Measure 1 seems to perform better in the year 2030 where higher volumes load the network. In overall, the highest ranked measure is Measure 3 with a remarkable 38.1% improvement in delays for the year 2030, while for 2017 the lowest ranked is Measure 1. It is important to note here that traffic delays for Measure 2 in 2017 have remained unchanged, as the only change is that the share of HGVs using green technologies has been increased.

5 Conclusions

The European Union has funded numerous programs and projects that focus on urban logistics and urban interchanges due to their impacts on transport networks, economy, and environment. In this direction, several logistics and traffic measures have been proposed and tested in urban areas to evaluate their potential towards improving the sustainability of a transport system. This study focuses on the evaluation of three such measures by using a microsimulation tool. Results showed that the measure that achieves the highest improvement in CO_2 emissions is the “Enforcement and ITS adoption for control and traffic management”.

In order to improve the analysis of this case study a future research should include the extension of the current network up to the industrial area in order to reflect the total changes in the area due to measures’ implementation. A larger network could also be used as a test bed to evaluate additional solutions. As regards the evaluation methodology, future research could consider a combination of simulation software that deal with internal processes as usual traffic simulation software and optimization techniques are difficult to simulate freight systems. AnyLogic software could be proved an excellent tool as it could simulate the whole arrival, unloading, storing and loading to HGVs processes of the cargo. In this case, optimization and measure testing could be achieved not only during the distribution but during the whole course of the transfer of the cargo from the port to the warehouse. Finally, AnyLogic as a multimethod software could extend the evaluation in other impact areas i.e. economy and energy or society, contributing in a holistic assessment of each considered solution.

Acknowledgements. This work has been supported by the ALLIANCE project (http://alliance-project.eu/) and has been funded within the European Commission’s H2020 Programme under contract number 692426. This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

References

1. http://civitas.eu/mobility-measures. Accessed 9 Oct 2017
2. Nathanail, E., Gogas, M., Adamos, G.: Smart interconnections and urban freight transport towards achieving sustainable city logistics. Transp. Res. Procedia **14**, 983–992 (2016). 6th Transport Research Arena, Warsaw, Poland, 18–21 April. https://doi.org/10.1016/j.trpro.2016.05.078
3. Port Authority of Volos. http://www.port-volos.gr/. Accessed 30 May 2017
4. http://www.ellinikiaktoploia.net. Accessed 9 Oct 2017

5. Hellenic Statistical Authority. http://www.statistics.gr/en/home. Accessed 16 June 2017
6. Naftemporiki.gr. http://www.naftemporiki.gr/. Accessed 22 Apr 2017
7. International Council on Clean Transportation, European Market Statistics. http://eupocketbook.org/. Accessed 16 June 2017
8. European Automobile Manufacturers Association (2017). http://www.acea.be/. Accessed 1 July 2017
9. Wisconsin Department of Transportation, Microsimulion Guidelines. http://www.wisdot.info/microsimulation/index.php?title=Model_Calibration#The_GEH_Formula. Accessed 30 May 2017
10. Karakikes, I., Spangler, M., Margreiter, M.: Motorway simulation using bluetooth data. Transp. Telecommun. J. **17**(3), 242–251 (2016). https://doi.org/10.1515/ttj-2016-0022
11. Wu, X., Deng, S., Du, X., Ma, J.: Green-wave traffic theory optimization and analysis. World J. Eng. Technol. **2**, 14–19 (2014)
12. Barth, M., Mandava, S., Boriboonsomsin, K., Xia, H.: Dynamic ECO-driving for arterial corridors. In: 2011 IEEE Forum on Integrated and Sustainable Transportation Systems, Vienna (2011)
13. Linton, C., Grant-Muller, S., Gale, W.: Approaches and techniques for modelling CO_2 emissions from road transport. Transp. Rev. **35**, 533–553 (2015)
14. Borge, R., Lumbreras, J., Pérez, J., de la Paz, D., Vedrenne, M., de Andrés, J., Rodríguez, M.: Emission inventories and modeling requirements for the development of air quality plans: application to Madrid (Spain). Sci. Total Environ. **466–467**, 809–819 (2014)
15. Quaassdorff, C., Borge, R., Pérez, J., Lumbreras, J., de la Paz, D., de Andrés, J.: Microscale traffic simulation and emission estimation in a heavily trafficked roundabout in Madrid (Spain). Sci. Total Environ. **566–567**, 416–427 (2016)

Innovative Economics

Industry 4.0 – For Sustainable Development of Lean Manufacturing Companies in the Shipbuilding Sector

Anatoli Beifert[1(✉)], Laima Gerlitz[1(✉)], and Gunnar Prause[1,2(✉)]

[1] Wismar University of Applied Sciences, Philipp-Müller-Str. 14, Wismar, Germany
{anatoli.beifert,laima.gerlitz}@hs-wismar.de
[2] Department of Business Administration, Tallinn University of Technology, Akadeemia tee 3, 12618 Tallinn, Estonia
gunnar.prause@ttu.ee

Abstract. The world we live today has been considerably influenced by industrial revolutions, continuous integration of new technologies and innovations in the production process. Identifying the key role of the industrial sector for job and innovation growth, as well as improvement of productivity and research infrastructure, a number of political, business and scientific initiates have boosted industrial renaissance on the top agenda in the innovation driven economies. Although, industry generates 80% of the EU private innovations and 75% of its exports, the global share of European manufacturing value added dropped from 36% in 1991 to 25% in 2011. EU industry accounts for about 15% of the total gross value added and in the shipbuilding sector the situation is even more challenging: European shipbuilding produces' share worldwide decreased from 66% in 1960 to 2.4% in 2015 (Stopford, 2016). Many manufacturing initiatives have been started all over the world, re-establishing an industrial share in the economy. With an industrial gross value added of about 30% in Germany this approach has been named "Industry 4.0,". A number of similar initiatives, e.g. "Advanced Manufacturing Partnerships" or "Made in China 2025" are being implemented in USA and China respectively. Innovation driven markets expect manufacturing companies, also in the shipbuilding sector to react appropriately to the rapidly changing network environments. One of the promising approaches matches virtual and real worlds by linking virtual and manufacturing environment. The paper investigates current shortcomings in the lean manufacturing of the shipbuilding sector or shipbuilding suppliers, utilizing modular manufacturing in a lean production process and the opportunities of utilizing Industry 4.0 approaches are discussed. The presented research is empirically validated by using the primary data samples from the shipbuilding sector companies from the Northern Germany.

Keywords: Digital revolution · Lean manufacturing · Shipbuilding

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_54

1 Introduction

After two decades of decline, manufacturing and re-industrialization have started enjoying the so-called industrial renaissance on the Western agenda [17]. This is due to growing incentives of politicians, business leaders and scientists to recall the role of the industrial sector as a key driver for research, productivity and job creation. Industry generates 80% of the EU private innovations and 75% of its exports. A closer look reveals that the global share of European manufacturing value added dropped from 36% in 1991 to 25% in 2011. Currently, the EU manufacturing industry accounts for only about 19% of the total gross value added. In the shipbuilding sector the situation is more challenging: European produces' share worldwide decreased from 66% in 1960 to 2.4% in 2015 [46]. Moreover, after the financial crisis, the manufacturing recovery has fallen behind, and the EU generated lower manufacturing outputs than Asia, US and South Korea. In aggregate, faster recovery in Asian manufacturing implied multiplication of manufacturing in China, which now shares higher manufacturing outputs value than that of the EU or US [18]. Furthermore, this change was associated with the emergence of new smart technologies and missing capabilities to quickly respond to such changes. Implications of undertaken outsourcing to lower labor cost regions and service activities on a long-term basis such practices have lead in Europe to reduction of living standards in industry nations, such as Germany, UK, Italy, etc. Increased competition in Europe from overseas, especially in the Far East through stronger manufacturing of high-tech products meant a drain of know-how and competences from Europe to the East and induced a shift in European competitiveness. This negative trend has to be overcome by tracing back towards the cornerstone of Europe's performance – smart manufacturing, especially, high technology and medium-high technologies, smart services and smart society [25]. A number of manufacturing initiatives have been launched all over the world, calling for regaining an industrial share in the economy (Eurostat, 2015). A promising approach in this context seems to be matching of virtual and real worlds through strengthening the linkage and interplay between IT, internet and manufacturing. In Germany, with an industrial gross value added of about 30%, this approach has been named "Industry 4.0". There is a number of similar initiatives, e.g. "Advanced Manufacturing Partnerships", "Made in China 2025" that are being implemented in USA and China. Industry 4.0 aims at developing cyber-physical systems and dynamic production networks in order to provide with flexible and open value chains for the manufacturing companies [8, 29]. Industry 4.0 business approach targets also at energy and resource efficiency, shortening of innovation and time-to-market cycles as well as productivity and envisaged value-added increase [5, 26, 35, 49, 52]. Thus, Industry 4.0 may improve the competitiveness of European manufacturing and high-tech industrial countries. It may be also that small and medium sized enterprises (SMEs) represent the backbone of the European manufacturing, whereas some of those SMEs are world market leaders, utilize lean manufacturing along with innovation technologies to secure sustainable growth [48]. However, in spite of the evident benefits of the lean method the manufacturing SMEs still experience challenges that hinder efficient lean utilization [42]. Moreover, the role in international supply chains have been mainly restricted to second- or third-tier suppliers or to highly specialized service providers [9, 12–14, 23, 27, 32]. This may change with the spread of Industry 4.0, and increasing interactions offline and online environment. The smart production and logistics solutions are expected to activate the entire supply chain from product design and development, operations management and logistics to distribution by connecting physical and virtual systems. The number of players is increasing, the manufacturing and services systems become more dynamic and complex. The demand for interoperability, virtualization, decentralization, real-time capability, service orientation and modularity [24]. In this perspective, Industry 4.0 requires innovative business models and structures along with new concepts for managing information and taking into account the needs of internationally operating manufacturing companies [6, 10, 11, 19, 30]. Currently, companies have started to gain first experience with concepts like production in networks or smart logistics. They also begin developing new organizational structures and models to benefit more from opportunities that the new technology offers. It may be stated that there is research gap on the opportunities and integration requirements for the lean shipbuilding in the context of Industry 4.0 and the question of efficient interplay between lean manufacturing and Industry 4.0 is significant research field that is to be further explored [42]. Along with the challenges of the innovation shortcoming in the lean manufacturing the shipbuilding sector or shipbuilding suppliers based on a case study approach, the paper investigates here possible future developments and implementation of Industry 4.0 that may open new business opportunities for manufacturing companies in the shipbuilding sector and how they could be benefit from new technologies of other innovation driven industries.

In this context, the presented work investigates if Industry 4.0 approaches may be efficiently applied in the lean manufacturing of the shipbuilding sector that is utilizing modular manufacturing in a lean production process. Furthermore, the current shortcomings and problems as well as the opportunities of utilizing Industry 4.0 approaches are discussed. In order to address the research question, the authors reflect lean manufacturing that applies to the shipbuilding sector, including dimensions of lean manufacturing and current challenges mentioned in the literature. Furthermore, Industry 4.0 phenomena and its effects on the lean manufacturing in general and on the shipbuilding companies in particular will be discussed. Finally, based on primary research data gained from case studies the possible implementation models of lean production in the shipbuilding through Industry 4.0 will be presented. The given research is mostly based on a qualitative approach, case studies employed here have been gained within study research projects at Wismar University of Applied Sciences, Department of Maritime Studies. The action research as a feasible methodological framework contributing to looping of the empirical observations to a bigger picture has been applied. The body of empirical data is comprised of case studies from shipbuilding companies and suppliers. The examination and implications of the selected case studies is supported by expert interviews, exploratory reports, reflection memos, synopsis and storytelling. The presented research is empirically validated by using the primary data samples from the shipbuilding sector companies from the Northern Germany, here: Hamburg, Rostock and Wismar areas. The empiric analysis is based on semi-structured expert interviews data and also secondary data comprising the synergies between lean shipbuilding and Industry 4.0 approaches.

2 Lean Manufacturing in the Shipbuilding

Nowadays, the growing deployment of the automation systems, computer-based and IT technologies in manufacturing lead to a more customer-value focused method of manufacturing called Lean Manufacturing, successfully introduced and implemented in automobile industry by Toyota Motor Corporation that caused considerably decrease of resources' wastes and increased productivity efficiency [42]. Following these success developments other industries, including shipbuilding sector worldwide and also in Europe are implementing or intending to integrate lean manufacturing approach to increase manufacturing efficiency [34]. Therefore, the lean manufacturing may be considered as an appropriate methodology to improve productivity and decrease costs in manufacturing companies in various industries [2, 42] and can be described as a multi-faceted manufacturing approach that identifies and creates value aiming to secure smooth flow of production processes through the organization that comprises a set of practices focused on reducing of so called production wastes, i.e. non-value added activities from the company's manufacturing activities and consists of such lean principles or tools: Just-In-Time production (JIT), Total Quality Management (TQM), Kaizen (continuous improvement), Hansei (reflexive approaches), lean thinking and lean training [28, 39, 43, 53]. Although lean manufacturing may be considered as a common framework for modern operations in a manufacturing organization, still there is a lack in understanding of a clear mechanism for a successful lean process implementation [22]. Lean concept in the shipbuilding may be therefore described as a relatively new production process that comprises the methods of lean manufacturing and lean construction or assembling.

Womack and Jones (1996) identify major lean elements that are in particular applicable to the shipbuilding: specific identification of every product's value and "value stream" for each product, whereby value flow shall be uninterrupted and the customer are initiating every transaction (here: pull) simultaneously reducing wastes (e.g. time, space, costs, and mistakes). In addition, lean thinking and lean training tools mostly aim at introducing and motivation to implement lean concepts into practice in manufacturing companies, whereby the objectives need to be fully understood, cross-functional teams are expected to be active in the value stream creation and design process is shifted along the value stream [53, 55].

"Hansei" approach in lean shipbuilding may be described as reflections and considerations that indicate a major impact that production process induces on the environment, thus underlining corporate responsibility for the environmental issues the importance of green and innovative technologies along the supply chain aiming at mutual beneficial learning between partners or customer-suppliers [50].

In lean manufacturing Kaizen approach may be described as a continuous improvement process that involves employee from different levels, motivating employees to make improvement suggestions on a regular basis for the further analysis and implementation of the most efficient proposals. Therefore, Kaizen model may be defined rather as a tactical instrument like safety and efficiency improvements. The Kaizen ideas are not specifically linked to production, marketing or management, on the contrarily can be deployed at any place where enhancements can be achieved (Atari and Prause, 2017).

Further it can be stated that the modern manufacturing models including shipbuilding, embrace mostly modular and fractal approaches as well as network-orientation, flexibility and responsiveness. The manufacturing companies attempt to follow smart specialization and smart production trends in order to gain networked production experiences and develop new business models for efficient adaptation to constantly changing network environments [37, 38].

However, shipbuilding might be a challenging sector for the lean approach. Parveen and Rao (2009) outlined that major aspect of lean like reduced inventories, reduced wastes and lead times, higher quality reduced scrap, an ability to keep to schedules, increased flexibility, efficient automation and better infrastructure and equipment utilization do not completely apply in shipbuilding, since the main objective to keep the schedule may be prior to all the other operational objectives. The representative from the interview shipbuilding companies also confirmed the strategic importance of keeping the delivery schedule in construction and manufacturing. Paying attention to the importance of fixed prices and time contracts in the shipbuilding in regard to the possible requirements changes or order's variation that might distort the schedule and production flow for a current and also for other projects Dugnas and Oterhals (2008) also underlined the challenging character of lean philosophy in the shipbuilding industry. They also pointed out at risks of control deficits of phases in their transition that might cause a big increase of project costs, interruption of workflow overburdening of workforce, thus resulting in late delivery. However, at the same time they also recommend simplification, visualization and improved information flow as a possible opportunity to improve manufacturing and construction efficiency that might be also applied by shipbuilding companies or suppliers.

3 Industry 4.0 and Its Implications in Shipbuilding

Three widely acknowledged industrial revolutions may be described through the following general characteristics: (1) deploying mechanical production facilities driven mostly by steam or water power; (2) mass production through the machines' deployment that were driven by electrical energy; (3) wider deployment of IT and automation systems in industrial manufacturing. The terminology "Industry 4.0" was first publicly presented during Hannover Fair in 2011. Since that time, being widely discussed by political, business and research community, Industry 4.0 is seen as the fourth industrial revolution that deploys the principles of cyber-physical systems (CPS), innovative IT technologies and smart systems with an efficient human-machine interaction, which enables every entity enhancement in the value stream that leads to digitalized mass customization in lean manufacturing [36, 47].

Since the potential of lean manufacturing calls for sustainable and durable actions that aim to overcome identified hindrances, Industry 4.0 approach offers a huge opportunity to transfer a manufacturing company to a smart producer by utilizing advanced information and communication (IT) systems and future-oriented technologies. A number of researchers underline the importance nowadays of integration of lean manufacturing with Industry 4.0 phenomena, pointing out that the modern technologies and concepts of Industry 4.0 may act as powerful enablers for a sustainable business development [38, 42].

Industry 4.0 may considerably influence the manufacturing environment completely though: modifying traditional and optimizing relatively new production operations; enabling real-time interactions and dynamic self-optimization of manufacturing processes. The perspectives of intelligent and self-learning and self-optimizing production facilities in the production line that can synchronize themselves along the entire value chain from order of materials from suppliers to delivery of final products to customers seemed to be fiction recently, now shape our reality [45]. Applications of visualizations of manufacturing processes, computer based simulations of prototype developments, deployment of smart supply chains, utilizing interactive IT technologies for employee training and risks [51].

Industry 4.0 also aims at dynamic production networks in order to provide with flexible and open value chains in the manufacturing of complex mass customization products in a small series up to lot size 1. Industry 4.0 business models target at energy and resource efficiency, shortening of innovation and time-to-market cycles as well as an increase in productivity and envisaged value-added. Industry 4.0 is expected to enhance the competitiveness in the manufacturing and high-tech sectors. This bears also a promising perspective on the macro-regional level, e.g. for the Baltic Sea Region (BSR), since manufacturing in general and shipbuilding sector in particular play a significant role for the regional economies [2]. The BSR region demonstrates the highest innovation level among the EU macro-regions, sophisticated ICT infrastructure and has highly qualified workforce. Also the role of small and medium-sized enterprises (SMEs) in international shipbuilding supply chains have been mainly restricted so far to second – tier suppliers – or to highly specialized service providers. Industry 4.0 may change the situation, because the smart production and logistics solutions are expected to touch the entire supply chain from product design and development, operations management and logistics to distribution. Consequently, Industry 4.0 requires new business models and structures together with new concepts for managing information and business administration taking into account the perspective of internationally operating entrepreneurs and SMEs. Currently, companies have started to gain first experience with concepts like production in networks or smart logistics. They also begin developing new organizational structures and models to benefit more from opportunities that the new technology offers.

4 Integrating Lean Shipbuilding and Industry 4.0

Case study based qualitative data from two shipbuilding suppliers and one shipbuilding yard has been collected and analyzed in the framework of a study research project executed by the Department of Maritime Studies of Wismar University of Applied Sciences.

Although the lean shipbuilding is clearly recognized as the building process of ships and offshore units, ruled and performed by lean principles that are being integrated and applied in the shipbuilding sector, the interview results showed that the current implementation of lean approaches by the shipbuilder may be regarded as insufficient. According to Parveen and Rao (2009), the communication and transaction costs, high market volatility and demanding customized products may be the identified as the possible reason for this situation. Entire and sophisticated production planning and deficits in optimization have been identified as current problems in the lean shipbuilding. It has been also stated that the level of non-value adding activities or wastes, which increase costs, delay the delivery schedule, compromise quality and rather low contribute to the Client's expectations in most shipbuilding yards and supplying companies is rather high.

The following lean principles have been identified by the interviewed companies as the most relevant for the shipbuilding sector: (1) defining value with a strong customer involvement in specification, design and pricing process; (2) mapping value stream, including product development from functional aspects to ship drawings, procurement, construction and manufacturing, systems integration and certification (i.e. from installation through hook-up, mechanical completion, commissioning and final delivery to the customer); (3) creating flow by eliminating "batch and queue" and rather focusing on step-to-step operations or rather "piece by piece" construction flow in very small batches; (4) pull principles through reducing overproduction, keeping low inventories, reducing of wastes, minimizing working capital and improved communication. The following lean characteristics have been named as the most advantageous for the shipbuilding: improved quality of final products, reduced construction time; enhanced image among final customers and also suppliers; improved materials management system. High costs, long term introduction or restructuring and success dependency on the top-down approach have been named as the main barriers for the realization of the lean shipbuilding. On this background, a stronger focus on Industry 4.0 opportunities and tools have been mentioned as perspective method for the smooth implementation of the lean concepts in the shipbuilding. The interview companies identified following benefits of Industry 4.0 for the shipbuilding sector: better supplier development through technological networks that will e.g. allow sharing of intangible resources like competences, know-how, tacit knowledge [4]; efficient computer based technologies and interactivity utilized for optimization of customer needs, involving customers in the production process in all stages of production; manufacturing planning and lean optimization modelling; error handling and error detection deployment that aims at reducing failures and risk costs; deploying layout modelling and modifications in order to secure products compliance, improve overall equipment effectiveness and enhance warehouse space utilization; using animation systems for maintenance; and integrating virtual reality for staff training.

In the framework of the study, the interviewees were asked to identify maximum six most appropriate Industry 4.0 features for the shipbuilding sector, distribute the weight of the identified features and evaluate their current application using grading scale from 1 (poor) to 10 (excellent) (Table 1).

Table 1. Applicable Industry 4.0 instruments for the shipbuilding sector.

Instruments or methods	Weight (%)	Scoring
Customer(s) interactive inclusion	30	8
Animation to reduce design and development costs	18	7
Lean modeling and lean optimization tools	16	6
Visualization for suppliers development	14	4
Training tools	12	3
Process design	10	4

At the same time, relatively high costs, integration complexity, potential cyber-crimes, and low IT knowledge of the employees have been named as the main barriers for efficient utilization of Industry 4.0 in the current shipbuilding sector.

Furthermore, it has been mentioned that Industry 4.0 opportunities might be employed as a methodological and strategic tools to accelerate engagement of shipbuilding suppliers and yards in regional and global networks. This is especially crucial in the emerging domain of Industry 4.0 applications and increasing related business and social interactions.

5 Conclusions

Although the lean principles are also relevant, being integrated and applied in the shipbuilding sector, the interview results showed that the current implementation of lean approaches by the shipbuilder may be regarded as insufficient. The communication and transaction costs, high market volatility and demanding customized products may be the identified as the answer for this [33].

Since Industry 4.0 related supply a value chains take place in complex and intertwined manufacturing networks, which can be characterized by a high degree of fragmentation, lower entry barriers for SMEs and new business models spur the integration of SME into Industry 4.0 supply and value chains. Consequently, regional SMEs need and, much more important, should learn from their bigger counterparts in employment of Industry 4.0 in their business interactions and to find their own digital industrialization strategies that fit the market, Industry 4.0 standards and regulative framework. Being or becoming a part of Industry 4.0 supply and value chains, a regional SME depends no longer only on performance parameters like cost levels or specific knowledge and competencies, but is rather to a large extent conditioned by its performance in transferring from Industry 3.* to Industry 4.0. Handling successfully an internal digitalization process becomes a key factor for SMEs to benefit, integrate, compete and thus develop and grow in a smart and sustainable way. In this, the research provides entrepreneurs, SMEs and other practitioners with practical tools and recommendations. Theoretical contributions are also proposed and discussed, which enhance the emerging literature on business models within the Industry 4.0 context as well as organizational change and potential responses of organizations to the changes.

References

1. Ang, J.H., Goh, C., Li, Y.: Smart Design for Ships in a Smart Product Through-Life and Industry 4.0 Environment. University of Glasgow (2016). http://eprints.gla.ac.uk/118544/1/188544.pdf. Accessed 14 July 2017
2. Atari, S., Prause, G.: Performance aspects of lean manufacturing in the context of Intrapreneurship. In: Mitra, J. (ed.) Proceedings of the 15th International Entrepreneurship Forum, Venice, Italy 14–16 December, 2016, pp. 11−19. Essex Business School, Essex (2017)
3. Argyris, C.: On Organisational Learning. Blackwell Publishers, Cambridge (1992)
4. Beifert, A., Gerlitz, L., Prause, G.: Sustainable business development models for regional airports. In: Kersten, W., Blecker, T., Ringle, C. (eds.) Innovations and Strategies for Logistics and Supply Chains, Proceedings of the Hamburg International Conference of Logistics (HICL), pp. 256−284. epubli GmbH, Berlin (2015). www.epubli.de
5. Benedetto, M.F., Berrone, S., Pieraccini, S., Scialò, S.: The virtual element method for discrete fracture network simulations. Comput. Methods Appl. Mech. Eng. **280**, 135–156 (2014)
6. Blythe, C.: Business models for value generation in the Internet of Things. In: Data- and Value-Driven Software Engineering with Deep Customer Insight. Proceedings of the Seminar no. 58314308, pp. 8–15 (2014)
7. Borgatti, S.P., Halgin, D.S.: On network theory. Organ. Sci. **22**(5), 1168–1181 (2011)
8. Brettel, M., Friendrichsen, N., Keller, M., Rosenberg, M.: How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 perspective. World Acad. Sci. Eng. Technol. Int. J. Mech. Aerospace Ind. Mechatron. Eng. **8**(1), 37 (2014)
9. Briscoe, G., Dainty, A.: Construction supply chain integration: an elusive goal? Supply Chain Manage. Int. J. **10**(4), 319–326 (2005)
10. Bucherer, E., Uckelmann, D.: Business models for the internet of things. In: Architecting the Internet of Things, pp. 253–277. Springer, Heidelberg (2011)
11. Chan, H.C.Y.: Internet of things business models. J. Serv. Sci. Manage. **8**, 552–568 (2015)
12. Ciliberti, F., Pontrandolfo, P., Scozzi, B.: Investigating corporate social responsibility in supply chains: a SME perspective. J. Clean. Prod. **16**(15), 1579–1588 (2008)
13. Croom, S.R.: The impact of e-business on supply chain management: an empirical study of key developments. Int. J. Oper. Prod. Manage. **25**(1), 55–73 (2005)
14. Danese, P., Romano, P.: Supply chain integration and efficiency performance: a study on the interactions between customer and supplier integration. Supply Chain Manage. Int. J. **16**(4), 220–230 (2011)
15. Denscombe, M.: The Good Research Guide: For Small-Scale Social Research Projects. McGraw-Hill Education, England (2014)
16. Dugnas, K., Oterhals, O.: State-of-the-art shipbuilding: towards unique and integrated lean production systems. In: The International Group for Lean Construction World Conference (IGLC16), Manchester (2008)
17. European Commission: For a European Industrial Renaissance, Communication from the Commission SWD (2914) 14 Final, Brussels (2014a)
18. European Commission: Advancing Manufacturing – Advancing Europe: Report of the Task Force on Advanced Manufacturing for Clean Production, Commission Staff Working Document, SWD (2014) 120 Final, Brussels (2014b)

19. Fan, P.F., Zhou, G.Z.: Analysis of the business model innovation of the technology of internet of things in postal logistics. In: 2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management (IE&EM), pp. 532–536. IEEE, September 2011
20. Freeman, L.: The Development of Social Network Analysis: A Study in the Sociology of Science (2004)
21. Freeman, R.E., Phillips, R.A.: Stakeholder theory: a libertarian defense. Bus. Ethics Q. **12**(03), 331–349 (2002)
22. Giunipero, L., Pillai, K., Chapman, S., Clark, R.: A longitudinal examination of JIT purchasing practices. Int. J. Logist. Manage. **16**(1), 51–70 (2005)
23. Gunasekaran, A., Ngai, E.W.T.: The successful management of a small logistics company. Int. J. Phys. Distrib. Logist. Manage. **33**(9), 825–842 (2003)
24. Hermann, M., Pentek, T., Otto, B.: Design Principles for Industrie 4.0 Scenarios: A Literature Review, Dortmund. Technische Universität Dortmund (2015)
25. Jaegers, T., Lipp-Lingua, C., Amil, D.: Industry, trade and services. Eurostat Stat. Focus **1**(2013), 1–7 (2013)
26. Kagermann, H., Wahlster, W., Helbig, J.: Securing the future of German manufacturing industry: recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group, acatech – National Academy of Science and Engineering (2013)
27. Kaplinsky, R.: Spreading the gains from globalization: what can be learned from value-chain analysis? Probl. Econ. Trans. **47**(2), 74–115 (2004)
28. Lanigan, J.: EMS feature-5S provides competitive lean foundation. SMT Surf. Mount Technol. Tulsa **18**(5), 70–72 (2004)
29. Lee, J., Hung-An, K., Yang, S.: Service innovation and smart analytics for Industry 4.0 and big data environment. Procedia CIRP **16**, 3–8 (2014)
30. Leminen, S., Westerlund, M., Rajahonka, M., Siuruainen, R.: Towards IoT ecosystems and business models. In: Internet of Things, Smart Spaces, and Next Generation Networking, pp. 15–26. Springer, Heidelberg (2012)
31. Levitt, B., March, J.G.: Organisational learning. Ann. Rev. Sociol. **14**(1), 319–338 (1988)
32. Lunati, M.: Enhancing the Role of SMEs in Global Value Chains. Staying Competitive in the Global Economy, p. 65 (2008)
33. Parveen, M., Rao, T.V.: An integrated approach to design and analysis of lean manufacturing system: a perspective of lean supply chain. Int. J. Serv. Oper. Manage. **5**, 175–208 (2009)
34. Piercy, N., Rich, N.: Lean transformation in the pure service environment: the case of the call service centre. Int. J. Oper. Prod. Manage. **29**(1), 54–76 (2009)
35. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming competition. Harvard Bus. Rev. **11**, 1–23 (2014)
36. Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., De Amicis, R., et al.: Visual computing as a key enabling technology for industrie 4.0 and industrial internet. IEEE Comput. Graph. Appl. **35**(2), 26–40 (2015). http://dx.doi.org/10.1109/MCG.2015.45
37. Prause, G.: Sustainable business models and structures for Industry 4.0. J. Secur. Sustain. Issues **5**(2), 159–169 (2015). http://dx.doi.org/10.9770/jssi. Accessed 2 May 2015)
38. Prause, G.: E-residency: a business platform for Industry 4.0? Entrepreneurship Sustain. Issues **3**(3), 216–227 (2016). https://doi.org/10.9770/jesi.2016.3.3(1)
39. Ramakrishnan, S., Testani, M.V.: The role of leadership in sustaining a lean transformation. In: IIE Annual Conference, Proceedings, p. 1. Institute of Industrial and Systems Engineers (IISE) (2010)
40. Reason, P., Bradbury, H.: The Sage Handbook of Action Research: Participative Inquiry and Practice (2008)

41. Rowley, T.J.: Moving beyond dyadic ties: a network theory of stakeholder influences. Acad. Manage. Rev. **22**(4), 887–910 (1997)
42. Sanders, A., Elangeswaran, C., Wulfsberg, J.: Industry 4.0 implies lean manufacturing: research activities in Industry 4.0 function as enablers for lean manufacturing. J. Ind. Eng. Manage. JIEM **9**(3), 811–833 (2016)
43. Shah, R., Ward, P.T.: Defining and developing measures of lean production. J. Oper. Manage. **25**(4), 785–805 (2007). http://dx.doi.org/10.1016/j.jom. Accessed 19 Jan 2007
44. Simonite, T.: The internet of things. MIT Technol. Rev. **117**(4), 68–78 (2014)
45. Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, M., Schlund, S.: Produktionsarbeit der Zukunft – Industrie 4.0. Fraunhofer Verlag, Stuttgart (2013)
46. Stopford, M.: Challenges in the world maritime industry. OECD Homepage. https://www.oecd.org/sti/ind/Session%202_%20a%20-%20Martin%20Stopford%20-%20Web.pdf. Accessed 14 July 2017 (2016)
47. Valdez, A.C., Brauner, P., Schaar, A.K., Holzinger, A., Zieflea, M.: Reducing complexity with simplicity-usability methods for Industry 4.0. In: Proceedings of 19th Triennial Congress of the IEA, Melbourne, Australia, pp. 9–14. RWTH Publications, Germany (2015)
48. Venohr, B., Meyer, K.E.: The German miracle keeps running: how Germany's hidden champions stay ahead in the global economy. In: Working Paper. Institute of Management, Berlin School of Economics, Berlin, SSRN 991964. http://dx.doi.org/10.2139/ssrn.991964. Accessed 14 July 2017 (2007)
49. Vermessen, O., Friess, P.: Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems. River Publishers, Aalborg (2013)
50. Walker, D.H.T.: Innovation and value delivery through supply chain management. In: Construction Innovation and Process Improvement, pp. 125–153. Wiley-Blackwell (2012)
51. Wan, J., Cai, H., Zhou, K.: Industrie 4.0: enabling technologies. In: International Conference on Intelligent Computing and Internet of Things (ICIT), pp. 135–140. IEEE, Harbin (2015). http://dx.doi.org/10.1109/ICAIOT.2015.7111555
52. Westerlund, M., Leminen, S., Rajahonka, M.: Designing business models for the internet of things. Technol. Innov. Manage. Rev. **4**(7), 5 (2014)
53. Womack, J.P., Jones, D.T., Roos, D.: The Machine that Changed the World. Simon & Schuster, New York (1990)
54. Womack, J.P., Jones, D.T.: Beyond Toyota: how to root out waste and pursue perfection. Harvard Bus. Rev. **74**, 140 (1996)
55. Zayko, M.J., Broughman, D.J., Hancock, W.M.: Lean manufacturing yields world-class improvements for small manufacturer. IIE Solutions **29**(4), 36–40 (1997)

Corporate Governance Disclosures: The Case of Latvian Companies Listed on Baltic Stock Exchange

Ieva Kozlovska(✉)

Transport and Telecommunication Institute, Lomonosova street 1, Riga 1019, Latvia
Kozlovska.I@tsi.lv

Abstract. There are studies showing inconclusive results toward mutual influence between corporate governance and company's performance. There are studies with a strong positive relation, others with a negative one or even the third category of studies concluding that there are no any relationship at all [30]. Nevertheless most of studies show the positive link between corporate governance and company's performance, there are also ones showing mixed results or even no any mutual relationship between corporate governance and company's performance. Therefore, mechanisms of corporate governance do have influence on company's financial performance – some positively, others in a negative way. This research paper studies whether there is any mutual relation between corporate governance disclosures and performance of Latvian listed companies on Baltic stock exchange. The research is based as on the analysis of theoretical literature and research papers within the area of corporate governance disclosures, its influence on companies' performance as on analysis of corporate governance codes and annual reports of Latvian companies listed on Baltic stock exchange. Results of this research might be of interest for academic researchers as well as educators and practitioners of companies' annual reports analysis. Conducted research enabled to define new recommendations relating corporate governance disclosures of Latvian listed companies on Baltic stock exchange.

Keywords: Corporate governance disclosure · Performance
Latvian listed companies on Baltic stock exchange

1 Introduction

This research paper presents the outcome of the exploration of possible mutual relation between corporate governance disclosures and performance of Latvian listed companies on Baltic stock exchange. More specifically, the author analyzed existing approaches in all 25 Latvian companies listed on Baltic stock exchange related to "voluntary" corporate governance disclosures and companies' performance by use of secondary sources of information available at the Nasdaq Riga per time period from year 2007 to year 2016. Further, the author developed "voluntary" corporate governance disclosure index for each of the company for every year in time period from year 2007 to year 2016 based on calculation by use of the unweighted disclosure index approach. Financial performance of companies were evaluated by calculating return on assets (ROA), return on

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_55

equity (ROE), profit per share, net profit, and average volume weighted price of company' share in the particular period. The results of this research support idea that there is mutual, but not consistent relation between voluntary corporate governance disclosures and performance of Latvian companies listed on Baltic stock exchange in time period from year 2007 to year 2016. This research paper is the first study examining theory and practice of mutual influence between corporate governance disclosures and financial performance of Latvian listed companies on Baltic stock exchange. Result of this research might be of interest for academic researchers as well as educators, practitioners, also investors and other stakeholders analyzing companies' annual reports and companies' corporate governance reports. Also, results obtained might be of interest for Baltic stock exchange to turn to develop new additional rules how information should be reflected in companies' corporate governance reports thus contributing to develop of Latvian capital market and economy.

Nowadays, corporate governance and its reporting becomes more and more pressing issue for the listed companies in areas relating going-concern reporting, risk management, internal controls, board balance, and directors' remuneration [2]. Strong and efficient corporate governance is essential and critical for capital markets and economy of the particular country. Due to globalization to develop local capital market every country should take into account issues concerning investors' protection followed by the transparency and disclosure at the highest level. Financial information is one used to take decisions in long-term perspective. "Corporate governance and transparency have been a major concern in financial markets. Better governance practices and more transparency have been encouraged in many countries for several reasons; for example, better practices are associated with efficient price discovery, more efficient allocation of resources, financial development and economic growth. Moreover, transparency in markets reduces uncertainty, which is often measured by stock price volatility" [25].

The paper is structured as follows: the literature review provides an overall overview of the literature in the area of this research in order to develop conceptual foundation for the current study. The research methodology and findings are presented and discussed, followed by the conclusions.

2 Theoretical Framework of the Research

2.1 Corporate Governance

"The word "corporate governance" is liberally used by many in the industry and has become a subject of intense debate and discussion all over the world" [37]. There is quite substantial amount of literature and definitions about the term "corporate governance" defined by academics and by industry. Key difference among these definitions is that they are focusing on various factors relating corporate governance – finance, accounting, law, management. "Corporate governance is the process and structure that is used to direct and manage the business and affairs of the company toward enhancing business prosperity and corporate accountability… A good corporate governance practice and ethical business practices is believed to reflect image and value of a company and become one way to attract and restore investors' confidence" [31]. Corporate governance – set of relationship between company's management, board, shareholders and all other stakeholders – is one of key elements used to improve efficiency of economy, growth and investors' confidence [28]. Corporate governance is the document building mutual relationship between companies' owners and managers, and even stakeholders in a clear, transparent manner with a long-term perspective. "In the last period it has been observed a growing concern for developing and implementing good corporate governance practices due to their major impact on companies' image, market share, companies organic growth, customers and long-term strategy" [10]. Also globalization has fostered situation the quality of company's corporate governance system plays crucial role for the company to survive in a global market place. Corporate governance system is the system by which companies are managed and controlled. It is an overall framework by which company's board set company's financial policy and control its implementation by regular reports of these activities to company's shareholders [5]. Corporate governance is not a structure, process or a set of policies. It is company-level mechanism managed by boards and from which better business performance outcomes can be reached [11].

Nowadays in order to ensure competitiveness and long-term sustainability for well-run companies significant role plays effective corporate governance framework. Therefore, there is the corporate governance code defining key principles to ensure corporate governance in listed companies on the particular stock exchange in the most effective way. Corporate governance practices differ not only among countries, but also among companies within the particular country [12, 17]. Main differences in corporate governance among countries are due to factors as history of the particular country, rich families, influence of various factors [e.g. wars, catastrophes, upheavals, etc.], various business groups (e.g. very common in Japan), law (e.g. stock market has to be positively correlated with shareholder legal protection), trust, origins (each of country has its own culture, traditions, legal systems, etc.), evolution, transplants, large outside shareholders, financial development, politics, entrenchment [24]. Nevertheless every country does have authority to develop its own model and there are no any single and good model applicable to every country all over the world, there are key principles to be included in this governing document. These principles are following: equality, transparency, accountability and responsibility.

2.2 Practices of Corporate Governance in Latvian Listed Companies on Baltic Stock Exchange

Latvian companies listed on Baltic stock exchange should to follow defines key principles in order to ensure corporate governance in these companies are lead in the most effective way. It is based on transparency, accountability, long-term perspective.

According with rules of Baltic stock exchange Latvian listed companies on Baltic stock exchange should develop and report the Corporate Governance Report since year 2006. During the time period from year 2006 to year 2016 there are in total following three documents "Principles of Corporate Governance and Recommendations on Their implementation" issued defining principles of corporate governance and the recommendations on their implementation for Latvian listed companies on Baltic stock exchange:

- From year 2006 to year 2008;
- Year 2009;
- From year 2010 to year 2016.

During time period this period key goal of these recommendations was to increase overall quality of corporate governance in Latvian companies listed on Baltic stock exchange. It has been done by fostering business ethics and promoting idea why qualitative corporate governance is so important for the pubic, business and the company. Qualitative corporate governance of the company allows this company to reach its financial and strategic goals in the most efficient way also increasing the value of this company.

Still as it is from the very beginning the corporate governance principles defined for Latvian listed companies on Baltic stock exchange are developed in form of recommendations for their implementation. Therefore, the rule "comply or explain" has been implemented allowing investors to understand various specifics and nuances of corporate governance practiced in the particular company.

2.3 Corporate Governance and Company's Performance

Corporate governance does have significant influence on companies' financial and operating performance, it mutually correlates with stronger impact in countries characterized with weak legal environments. This has been proved by such ratios as return on assets (ROA), return on equity (ROE), economic value added (EVA), market value added (MVA), Tobin-Q ratios, net profit margin (income/sales), return on equity (income/book equity), one-year sales growth. Corporate governance is strongly correlated with stock returns [4, 12, 15, 22, 32, 35]. Governance factor does have strong positive impact on stock returns that is even stronger than such factors as firm size and book-to-market ratio [6]. There is mutual relationship between company's corporate governance index and company's share price. Companies with the highest corporate governance scores significantly outperform the market [3, 21]. Board diversity shareholder engagement with employees and other external stakeholders do have significant influence on company's profitability [1]. Companies with stronger shareholder rights do have higher company's value, profits, sales growth and lower capital expenditure plus fewer corporate acquisitions [12]. Quality of corporate governance significantly determines company's stock price [20, 35]. "The presence of a corporate governance policy mitigates crash risk" [29]. Also, the company's corporate governance system does have impact on company's managers' behaviors in the decision-making processes concerning investments in research and development (R&D). Impact of R&D expenditures varies parallel with the level of managerial ownership, chief executive officer (CEO) duality and the independence level of the company's board. R&D investments and outcomes from this process have positive impact from managerial share ownership and the separation of the CEO and the chairman of the company's board of directors [13]. Corporate governance has been also influenced by the audit in relation with company's financial performance. Based on recent researches on the field the monitoring and incentive plans are the most important ingredients for corporate governance mechanisms including the board structure with size, implications and responsibilities of board, shareholders rights and specialized committees. Mechanisms of corporate governance do have influence on company's financial performance, some positively, others in a negative way. Further, board independence, CEO duality, the number of meetings of the board members, compensation, indemnities, the existence of advisory committees, institutional investors correlate with company's performance indicators as well [9, 34]. To evaluate corporate governance influence on company's performance such elements as duality of the CEO (the CEO holds the position of the Chairman of the Board), independence of the audit committee members, stock ownership of institutional investors, the use of stock options for executive compensation can be used [36]. Companies with improved internal corporate governance do have higher company's value [33].

Nevertheless most of studies show the positive link between corporate governance and company's performance, there are also ones showing mixed results or even not any mutual relationship between corporate governance and company's performance. There is no meaningful relationship between corporate governance level and ROE, ROA, return on sales ratio (ROS), net profit (NP) [14]. Furthermore, there is no any consistent relationship between company's financial performance and structure of the particular corporate governance [1].

Further evaluating and measuring company's performance from the aspect of corporate governance various issues arise. One of such an issue is relating rules and regulations of countries [18, 19]. Different structures and behavior of corporate governance system is another issue. German/Japan system rely more on legal protection of creditors while UK and US ones more on protection of shareholders [16].

Based on the research performed the author concludes that there are different studies toward mutual relationship between company's corporate governance and performance. There are also numerous studies showing inconclusive results toward mutual influence between corporate governance and company's performance. There are studies with a strong positive relation, others with a negative one or even the third category of studies concluding that there are no any relationship at all [30].

3 Research Methodology

The review of theory conducted by the author in the previous section of this research paper allowed the author to form the general idea of the interest in the topic relating "voluntary" corporate governance disclosures of Latvian companies listed on Baltic stock exchange and its influence on companies' performance as well as the extent of its development in the scientific literature plus formulate the basic research question (RQ): Is there any relation between "voluntary" corporate governance disclosures and performance of Latvian companies listed on Baltic stock exchange?

In order to answer RQ the author analyzed existing approaches in all 25 Latvian companies listed on Baltic stock exchange related to "voluntary" corporate governance disclosures and companies' performance. These companies represent following industries – basic materials (code: 1000), industrials (code: 2000), consumer goods (code: 3000), health care (code: 4000), utilities (code: 7000), financials (code: 8000), technology (code: 9000), telecommunications (code: 6000). Company representing telecommunication industry was not included in this research because according with rules of Baltic stock exchange Latvian companies listed on Firth North (Baltic MTF) list of Baltic stock exchange should not develop and report the Corporate Governance Report [26].

The author investigated whether is there any relation between the developed "voluntary" corporate governance disclosure index and performance of Latvian companies listed on Baltic stock exchange. Therefore, following secondary sources of information available at the Nasdaq Riga website were used:

- Information from the Corporate Governance reports of Latvian companies listed on Baltic stock exchange per time period from year 2007 to year 2016;
- Information from annual reports of Latvian companies listed on Baltic stock exchange per time period from year 2007 to year 2016;
- Historical information available about security trading history of Latvian companies listed on Baltic stock exchange per time period from year 2007 to year 2016.

The Corporate Governance recommendations defined by the Nasdaq Riga are with a recommendatory nature rather than mandatory rules [27]. Also, there is no any indexes available calculated by the Nasdaq or any other institution. Therefore, the author developed the "voluntary" corporate governance disclosure index. This index describes practices of "voluntary" corporate governance of all Latvian companies listed on Baltic stock exchange (in total 24 companies) based on the principles of corporate governance and the recommendations on their implementation defined by the stock exchange Nasdaq Riga. In the time period from year 2007 to year 2016 Nasdaq Riga issued in total following 3 different documents defining principles of corporate governance and the recommendations on their implementation:

- From year 2006 to year 2008;
- Year 2009;
- From year 2010 to year 2016.

Therefore, the author developing the "voluntary" corporate governance disclosure index [VCGDI] investigated each of these periods separately. This index has been constructed in the following way – for each of this period maximum possible point of this index were calculated by assigning 1 point to each of the section in each of the company's Corporate Governance report per time period from year 2007 to year 2016:

- From year 2006 to year 2008 maximum possible points were 79 points;
- Year 2009 maximum possible points were 82 points;
- From year 2010 to year 2016 maximum possible points were 86 points.

Information provided in the Corporate Governance reports of Latvian companies listed on Baltic stock exchange per time period from year 2007 to year 2016 was analyzed by use of the disclosure index approach described above. The level of voluntary disclosure of Latvian listed companies on Baltic stock exchange has been measured based on “voluntary” corporate governance disclosure index. Past studies have used two types of disclosure score indexes – weighted index and unweighted index. Statistically weighted disclosure index and unweighted disclosure index are interchangeable because of their equivalent effects [7, 8]. Weighted disclosure index for different items of information can be assigned by investigating the views of users [8]. Other study concludes that unweighted disclosure index approach ensures avoidance of situations with subjective weights of a multiple users [23]. Therefore the author’s developed “voluntary” corporate governance disclosure index for each of the company for every year in time period from year 2007 to year 2016 based on calculation by use of the unweighted disclosure index approach – each of the disclosure item has been evaluated of the same importance scored with the same amount of points. This index was calculated in the following way:

- 1 point was added in case the particular company confirms that it applies the particular corporate governance principle;
- 0,5 point was added in case the particular company confirms that it applies the particular corporate governance principle partly or with specific exception;
- 0 points in case the particular company confirms that does not apply the particular corporate governance principle;
- “Voluntary” corporate governance disclosure index (VCGDI) has been calculated dividing the total score of the particular company by maximum possible score obtained by the company.

Further, in order to evaluate overall performance of all 24 Latvian companies listed on Baltic stock exchange the author has analyzed annual reports of these companies per time period from year 2007 to year 2016. Based on research performed above the author for this research calculated the mean of return on assets (ROA), return on equity (ROE), profit per share, net profit, and average volume weighted price of company’ share in the particular period.

4 Analysis of Research Results

The sample for this study is drawn from 25 companies listed on Baltic stock exchange in 2016 representing eight industrial sectors as detailed out in Table 1 below. The sample excludes company “Baltic Telecom” representing telecommunication industry was not included in this research because according with rules of Baltic stock exchange Latvian companies listed on Firth North (Baltic MTF) list of Baltic stock exchange should not develop and report the Corporate Governance report. This brings the final sample to 24 Latvian listed companies on Baltic stock exchange, representing seven industrial sectors as detailed out in Table 1 below.

Table 1. Industries of Latvian companies listed on Baltic stock exchange, 2016.

Industry	Frequency	Percentage
Basic materials	2	8
Industrials	6	25
Consumer goods	9	38
Health care	3	13
Utilities	1	4
Financials	1	4
Technology	2	8
Telecommunications	NA	NA
	24	100

Source: Author's own study.

According to the information provided in the Table 2 below Latvian companies listed on Baltic stock exchange are partly following rules defining which/how items should be disclosed in the Corporate Governance Report (Tables 3 and 4).

Table 2. Industries of Latvian companies listed on Baltic stock exchange, 2016.

Industry	AVG. points CG per industry	%	CG index
Health care	632	75	0.75
Industrials	598	71	0.71
Consumer goods	552	66	0.65
Financials	537	64	0.63
Basic materials	527	63	0.62
Utilities	422	50	0.50
Technology	72	9	0.08
Telecommunications	NA	NA	NA

Source: Author's own study.

Table 3. Average "Voluntary" CG disclosure index and average ROA (mean), average ROE (mean) of Latvian companies listed on Baltic stock exchange per industries, 2007–2016.

Industry	Average "Voluntary" CG disclosure index	ROA Mean	ROE Mean
Health care	0.75	8%	6%
Industrials	0.71	17%	2%
Consumer goods	0.65	5%	1%
Financials	0.63	2%	1%
Basic materials	0.62	3%	2%
Utilities	0.50	6%	5%
Technology	0.08	1%	2%
Telecommunications	NA	NA	NA

Source: Author's own study.

Table 4. Average "Voluntary" CG disclosure index and performance of Latvian companies listed on Baltic stock exchange per industries, 2007–2016.

Industry	Average "Voluntary" CG disclosure index	Share market price (MEAN)	Profit per share Mean	Net profit mean
Health care	0.75	4.83	0.06	5,515,217
Industrials	0.71	0.60	0.02	629,624
Consumer goods	0.65	1.63	0.01	574,179
Financials	0.63	0.80	0.01	59,620
Basic materials	0.62	6.10	0.02	1,503,072
Utilities	0.50	9.13	0.05	33,131,200
Technology	0.08	0.71	0.02	274,762
Telecommunications	NA	NA	NA	NA

Source: Author's own study.

Author explored that there is not consistent relation between corporate governance and companies' performance. Not consistency could be explained by the fact during the time period explored the country was experiencing significant financial crisis. Further there are companies which have been listed just a few years or companies which are not reporting the Corporate Governance Report. This structural feature of the Latvian companies listed on Baltic stock exchange provides evidence these companies are forced to become more transparent due to such pressures as globalization and internationalization. This issue has to be taken into account by Nasdaq Riga – defining rules for Latvian companies listed on Baltic stock exchange and local authorities – setting accounting standards for companies operating in the territory of Republic of Latvia.

5 Conclusions

The results of this research support idea that there is mutual relation between voluntary corporate governance disclosures and performance of Latvian companies listed on Baltic stock exchange in time period from year 2007 to year 2016. This mutual relation is not consistent that could be explained by the fact during the time period explored the country was experiencing significant financial crisis. Therefore, there is significant need for further research.

Based on results of this research paper Latvian company listed on Baltic stock exchange are only partly following rules defined which/how items should be disclosed in companies' Corporate Governance Report. This structural feature of these companies provides evidence the ones are forced to become more transparent due to such pressures as globalization and internationalization. This issue has to be taken into account also by Nasdaq Riga defining rules for Latvian companies listed on Baltic stock exchange.

References

1. Ayuso, S., Rodriguez, M.A., Garcia-Castro, R., Ariño, M.A.: Maximizing stakeholders' interests: an empirical analysis of stakeholder approach to corporate governance. Bus. Soc. **53**(3), 414–439 (2014)
2. Bhasin, M.: Voluntary corporate governance disclosures made in the annual reports: an empirical study. Int. J. Manag. Innov. **4**(1), 46–67 (2012)
3. Black, B.S., Hasung, J., Woochan, K.: Does corporate governance predict firms' market value? Evidence from Korea. J. Law. Econ. Organ. **22**(2), 366–413 (2006)
4. Boyacioglu, M.A., Akdogan, Y.E.: Corporate governance in financial institutions: an application on the Istambul stock exchange. J. Acad. Res. Econ. **2**(2), 180–196 (2010)
5. Cadbury, A.: The Report on the Financial Aspects of Corporate Governance. The Committee on the Financial Aspcets of Corporate Governance and Gee and Co. ltd., December, pp. 1–90 (1992)
6. Carvalhal, A., Nobili, C.: Does corporate governance matter for stock returns? Estimating a four-factor asset pricing model including a governance index. Quant. Financ. **11**(2), 247–259 (2011)
7. Chau, G.K., Gray, S.J.: Ownership structure and corporate voluntary disclosure in Hong Kong and Singapore. Int. J. Account. **37**, 247–265 (2002)
8. Chow, C.W., Wong-Boren, A.: Voluntary financial disclosure by Mexican corporations. Account. Rev. **I.XII**(3), 533–541 (1987)
9. Cioban, L., Narcisa, A., Hlaciuc, E.: The challenges of corporate governance and audit in relation to financial performance – trends, limitations, and directions to follow. USV Ann. Econ. Public Adm. **16**(2), 126–135 (2016)
10. Crişan-Mitra, C.: Good corporate governance: priorities and principles. Contemp. Legal Econ. Issues **5**, 465–483 (2015)
11. Crow, P., Lockhart, J.: Is corporate governance a structure, a process, a group of policies, or something else? In: Proceedings of the European Conference on Management, Leadership & Governances, pp. 89–95 (2016)
12. Gompers, P.A., Ishii, J.L., Metrick, A.: Corporate governance and equity prices. Q. J. Econ. Forthcoming **9**(2), 107–155 (2003)
13. Jermias, J.: The effect of corporate governance on the relationship between innovative effors and performance. Eur. Account. Rev. **16**(4), 827–854 (2007)
14. Kara, E., Acar Erdur, D., Karabiyik, L.: Effects of corporate governance level on the financial performance of companies: a research on BIST corporate governance index (XKURY). Ege Acad. Rev. **15**(2), 265–274 (2015)
15. Klapper, L.F., Love, I.: Corporate governance, investor protection, and performance in emerging market. J. Corp. Financ. **10**(5), 703–728 (2004)
16. Kvaal, E.: Topics in accounting for impairment of fixed assets. Ph.D. thesis. BI Norwegian School of Management, pp. 1–201 (2005)
17. La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.: Law and finance. J. Polit. Econ. **106**(6), 1113–1155 (1998)
18. Lantto, A.-M., Sahlström, P.: Impact of international financial reporting standards adoption on key financial ratios. Account. Financ. **49**(2), 341–361 (2009)
19. Lubyanaya, A.V., Izmailov, A.M., Nikulina, E.Y., Shaposhnikov, V.A.: Evaluation of the effect of non-current fixed assets on profitability and asset management efficiency. Int. J. Environ. Sci. Educ. **11**(15), 7745–7753 (2016)

20. Malik, M.F., Shah, S.M.A.: Value relevance of firm specific corporate governance and macroeconomic variables: evidence from Karachi stock exchange. Pak. J. Commer. Soc. Sci. **7**(2), 276–297 (2013)
21. Mans-Kemp, N., Viviers, S.: Does investing in sound corporate governance pay? A South African study. South Afr. J. Bus. Manag. **48**(2), 33–43 (2017)
22. Michelon, G., Parbonetti, A.: The effect of corporate governance on sustainability disclosure. J. Manag. Gov. **16**(3), 477–509 (2012)
23. Mohamed, A., Monirul Alam, H., Mahmud, H., Yao, L.: Corporate governance and voluntary disclosure in corporate annual reports of Malaysian listed firms. J. Appl. Manag. Account. Res. **7**(1), 1–19 (2009)
24. Morck, R.: A History of Corporate Governance Around The World, Family Business Groups to Professional Managers, 1st edn. The University of Chicago Press, Chicago, London (2005)
25. Mugaloglu, Y.I., Erdag, E.: Corporate governance transparency and stock return volatility: empirical evidence from the Istambul stock exchange. J. Appl. Econ. Bus. Res. **3**(4), 207–221 (2013)
26. Nasdaq Baltic Homepage, Baltic Equity list. http://www.nasdaqbaltic.com/market/?pg=mainlist&lang=en. Last Accessed 19 Aug 2017
27. Nasdaq Baltic Homepage, Review Of Corporate Governance Practices In The Baltic States (2015). http://www.nasdaqbaltic.com/files/baltic/CG%20review/Review%20of%20CG%20practices%20in%20the%20Baltics%20in%202015.pdf. Last Accessed 01 July 2017
28. OECD: OECD Principles of Corporate Governance. https://www.oecd.org/corporate/ca/corporategovernanceprinciples/31557724.pdf. Last Accessed 12 Aug 2017
29. Panayiptis, A., Constantinos, A., Joanne, H., Christodoulos, L.: Corporate governance and firm-specific stock price crashes. Eur. Financ. Manag. **22**(5), 916–956 (2016)
30. Pintea, M.-O., Fulop, M.-T.: Literature review on corporate governance – form performance relationship. Ann. Univ. Oradea Econ. Sci. Ser. **24**, 94 (2015)
31. Said, R., Crowther, D., Amran, A.: Ethics, Governance and Corporate Crime: Challenges and Consequences. Development in Corporate Governance and Responsiblity (e-book). Emerald Group Publishing Limited, Bingley (2014). ISBN 9781783506736. 9781783506743
32. Sreenu, N.: Correlations among the corporate finance decision, governance and financial performance in emerging market value: the study in indian cotext. J. Commer. Manag. Thought **8**(3), 616–638 (2017)
33. Sungmin, K., Hee Sub, B., Eun Jung, L.: Does "Vote No" change corporate governance and firm value? Evidence from the shareholder activism of the Korean national pension service. Emerg. Markets Financ. Trade **50**(5), 42–59 (2014)
34. Tan, I.Y., Tze San Ong, T.S., Babatunji, S.A., Chong, L.L.: Auditors switching in the relationship between corporate governance and financial performance – evidence from Malaysian Public Listed Companies (PLSc). Int. J. Econ. Manag. **10**(1), 53–68 (2016)
35. Tehrani, N.G.: Investing the effects of corporate governance system quality on company performance. Int. J. Manag. Account. Econ. **3**(7), 444–452 (2016)
36. Pham, T.T.K.: The relationship between corporate governance and the performance of the firm: a literature review with a focus on the vietnamses enterprises. In: Proceedings of the European Conference on Management, Leadership & Governance, pp. 460–465 (2016)
37. Venugopal, K.: Corporate governance: a cybernetic view. IUP J. Corp. Gov. **15**(2), 59–78 (2016)

Customer Satisfaction with Banking Services and Its Estimation

Ishgaley Ishmuhametov(✉)

Transport and Telecommunication Institute, 1 Lomonosova str., Riga 1019, Latvia
Ishmuhametovs.I@tsi.lv

Abstract. The increased competition in the banking sphere influences not only price but also non-price competition methods implementation. Marketing strategies focus on maximum customer satisfaction. Special attention is paid to corporate clients. The goal of this research is to determine the way for assessing the level of corporate customer satisfaction with banking services, and the intra-bank procedure for such assessment implementation. The author of the paper considers the methods employed to estimate customer satisfaction based on the analysis of advantages and disadvantages of these methods. In the paper, the results of the study performed using Lambin method are presented; this method was modified to overcome the drawbacks of selecting the aspects and deviations of the assessment of their variability. The practical applicability of the method was proven, and the possibility of its employment in the intra-bank procedure of assessing the level of corporate customer satisfaction was shown. The obtained results can be useful not only for banks in general, but they also can be applied by managers of different branches and operational offices for monitoring, timely detecting and correcting problem areas related to cooperation with corporate clients.

Keywords: Satisfaction with banking services · Corporate clients
Assessing the level of satisfaction

1 Introduction

The increased competition in the banking market led to the fact that most of the banks began to pay special attention not only to price, but also to non-price competition methods.

Marketing strategies are more focused on the maximum satisfaction of customers and improving the quality of services to adapt the bank branches to the constant changes in the external environment, to keep and maintain cooperative relationship with those who already use banking services, and to attract new ones.

Armstrong and Kotler [1] have repeatedly emphasized that the key to the market success for any company is to win and retain the customer due to the effective satisfaction of his needs. Modern managers are obliged to understand the important role played by satisfaction in the formation of a stable client base and long-term profit. According to Hill et al. [2], enterprises lose 10–30% of their customers each year because of the lack

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_56

of effective methods to meet the needs of consumers. In the opinion of these authors, the degree of customer satisfaction determines their future behaviour: whether they become regular customers of the organisation and whether they will recommend the company and its products to others or not. Customer loyalty is increasing due to various factors such as increasing customer service, improving network quality and value added, increasing loyalty, ensuring the growth of the company, so the marketing strategy should be planned, so that it is able to support affordable consumers [3]. Different scholars have tried to explain service quality and customer satisfaction with the help of financial inclusion, customer retention, customer awareness on modern banking services, mobile cash transfer and macro-financial variables [4].

The customer satisfaction and provision of the services corresponding to the expectation are capable of forming satisfaction via building long-term partnerships between the Bank and its clients. The special attention is paid to the corporate clients of the banks, when banks act as partners of their customers in the system of equal contractual relationship in accordance with the principle of the modern economy. In this regard, the topic is relevant in view of the need of the banking business to assess and analyse the satisfaction of corporate clients, bringing the largest revenue to the bank, as well as the relevance and practical applicability of the proposed method and the internal procedure of corporate satisfaction assessment based on this method. The bank as an organisation can make decisions aimed at preserving existing clients and attracting new customers only having objective information obtained on the basis of comprehensive analysis of the degree of clients' satisfaction with the services received.

Therefore, the goal of this research is to determine the method for assessing the level of satisfaction of corporate clients with bank services and the intra-bank procedure for its implementation. The author of the paper examines the existing methods used to assess customer satisfaction, basing on an analysis of the methods merits and shortcomings. There presented the results of the study with employment of Lambin (Jean-Jacques Lambin) technique, which was modified to eliminate the disadvantages of selecting the aspects and deviations in assessing their variability. Main research questions are as follows:

RQ1: Is there possible a single concept of satisfaction?
RQ2: Which of the existing methods of assessing the organisation customer satisfaction is the most suitable for assessing the satisfaction of corporate clients with the services of the bank?
RQ3: Is it possible to implement the most appropriate method for assessing the satisfaction of corporate clients of an organisation through an intra-bank procedure?

2 Literature Review

Many researchers point out that the services sector including the commercial banks, need to focus on non-financial perspectives in terms of service performance, consumer behavioural responses [5]. The commercial banks regard the customer satisfaction as the primary criterion used to assess the relationships of banks with the market [6].

There are a lot of definitions of satisfaction. Therefore, the lack of a common concept makes it impossible to give a universal definition of satisfaction, since its formulation depends on the position from which it is considered. It is important not to replace the concept of "satisfaction" with the concept of "gratification". Gratification is a short-term emotional experience, which is formed as a reaction to certain events. Satisfaction manifests itself as a long-term phenomenon and is formed as a result of a multiple experience of gratification. Blackwell et al. [7] explains gratification as a positive assessment of the chosen alternative; the judgment of the consumer that the product purchased by him, at least meets expectations or even surpasses them. In the context of economic sciences (marketing, theory of consumer behaviour), gratification is treated as the result of consumption of a product or service [8].

Satisfaction affects the need ambiguously: as the satisfaction increases, the need in the corresponding good can both weaken and intensify. Satisfaction depends on both the strength of the need, and on the magnitude of the desired good. The materials of marketing research conducted by F. Reicheld, contain the statement that companies can continue to lose market positions even while demonstrating high rates of satisfaction of their customers [9]. The author explains this also by the lack of customer loyalty. Loyalty can be seen as a sense of customer's affection towards the goods, services, personnel, environment and traditions of the company as a result of his satisfaction.

Customer satisfaction should be purposefully formed by organisations throughout the entire period of cooperation with the client. The various aspects of the concept "satisfaction" can be distinguished for its more accurate definition: satisfaction as an attitude of personality; satisfaction as a state; satisfaction as an appraisal; satisfaction as a purpose; satisfaction as a motive.

In the process of examining the satisfaction of corporate clients of the bank, the authors drew attention to the fact that the concept of "bank customer" is not always clearly specified in the legislation. More often, a client of a bank is understood as a natural or legal person who is on service with a credit institution. All legal entities and individual entrepreneurs, for which the definition "client of the bank" above is justified are taken as the corporate clients of the bank. According to Zhukov [10] the needs of corporate clients are so diverse that their effective satisfaction requires structuring the market of bank services. Satisfaction of a legal entity with the services of the bank is also based on satisfaction and subjective evaluation of specific people who have the direct contact with the bank and who sign the payment documents [11]. According to Danchenok and Denisova [12], the client begins to form his subjective evaluation after the first experience of consumption, using the following criteria: qualitative primary (reputation of the Bank, reliability, etc.); quantified regular (monthly, operational, and other types of tariffs, fees for services, etc.); quantitative irregular (one-time payment for withdrawal, for early termination of the contract, etc.); qualitative secondary (personal liking for the staff, availability of parking, etc.).

It is also important to note that the client's expectations are often formed under the influence of the bank employees, and it means that the qualification of the personnel has a very significant role for the formation of customer satisfaction, starting with the formation of their expectations.

In summary, it is possible to conclude that satisfaction as a subjective evaluation represents the difference between consumer's expectations and his perception of the real experience of using a product or service of the company. The higher the expectations of the client, the harder it can be to achieve his complete satisfaction. Therefore, organisations need to focus their efforts not only on improving the quality of the goods and services and servicing the customers, but also on the situation when the statements about the company's product does not exceed its actual possibilities.

3 Specifying the Methods of Customer Satisfaction Estimation

There are a number of techniques implemented for examination and assessment of the level of customer satisfaction. The most well-known are SERVQUAL, SERVPERF, BANKSERV, CSI, EPSI, Method of Lambin, etc. [13, 14]. The choice depends on many factors: what kind of information the company needs; the scope of the research; possibilities of the study implementation; budget, etc. The situation for banks is more complicated by the fact that initially all methods were calculated for implementing within the consumer sector of the economy, but not the financial one. The peculiarity of banks is the presence of two completely different customers, namely so-called "ordinary, individual customers" and "corporate clients". These customers are consumers of various banking services. This makes it impossible to simultaneously assess the satisfaction of both types of customers employing the same method. The most attractive clients for the bank are only corporate clients, because they bring more revenue to the bank. Therefore, this work is devoted to the study of the satisfaction of corporate clients.

The literature mentions that for assessing the quality of services there most often used the method SERVQUAL (from "service quality") offered by Parasuraman et al. [15]. This method is employed in order to provide the companies with opportunity to make specific managerial decisions basing on information about the quality of their product, that is, goods or services. Measurements are carried out in five directions – the so called service quality dimensions: reliability, responsiveness, assurance, empathy, tangibles [16]. According to Lee et al. [17], a consumer's assessment of his expectations during a survey can be greatly overstated, leading to an unjustifiably large gap between his expectations and perceived service. An analysis of the criticisms of the method is as follows: the method does not provide the variables weighing if they have different importance for consumers; separate measurements of the method overlap, especially regarding such factors as "empathy" and "responsiveness"; it is not applicable to all services, the composition of each measurement is strongly influenced by the specifics of the industry, etc.

In particular, SERVQUAL contributed to the emergence of such techniques as SERVPERF [18] and BANKSERV [19], the latter being designed for the Australian banking sector to assess customer satisfaction. The BANKSERV method is based on measuring the quality of service from the point of view of customer perception. The scale of the BANKSERV model also avoids the negatively built phrases found in the SERVQUAL tool. Nevertheless, this technique, as emphasized by the author, was originally developed for the assessment of satisfaction of various customers.

The method CSI (Customer Satisfaction Index) is offered by Dařena et al. to determine the satisfaction of each individual client, basing on his satisfaction with the separate services of the company [20]. At the same time, this method has a number of peculiarities, which generally limit the method application for assessing the satisfaction of corporate clients with bank services.

The ACSI (American Customer Satisfaction Index) method is an economic indicator measuring customer satisfaction throughout the US economy [21].

EPSI (Extended Performance Satisfaction Index) is a system designed to collect and analyse information about an organisation image, customer preferences and perceived quality, as well as loyalty of both customers and employees [22]. EPSI has developed a set of questionnaires for various industries. Due to its closedness and payment basis, it is not possible to assess the satisfaction of corporate clients with services using the EPSI method by the individual bank.

Lambin method for assessing the quality of services assumes the analysis as follows: in the beginning, the degree of general satisfaction with the service/services is assessed, then the satisfaction for each attribute and its importance are estimated [23]. This allows measuring the conformity of the satisfaction with the attribute of the service or product and the expectations of the consumer. To measure the degree of conformity of the goods, the ratio of satisfaction/importance (SI), expressed as a percentage, is used. Lambin points out that the results can be classified into four zones [14]:

- SI is above 100%: satisfaction exceeds importance; perhaps there is irrational direction of the company efforts;
- SI is between 90% and 100%: high satisfaction with important attributes;
- SI is between 80% and 90%: insufficient satisfaction with important attributes;
- SI is less than 80%: satisfaction is significantly below the importance of the attribute.

The method of Lambin is not only a method for studying consumer satisfaction, but also an effective tool for strategic planning. Graphic representation of the data allows quick respond to changes in consumer preferences and employment of corrective strategic decisions. The model proposed by Lambin provides more room for adaptation to a specific type of business, since the choice of attributes (in this case, aspects of provision of the services) is made by the researcher himself. Due to its peculiarities, this method has a high degree of applicability for assessing the satisfaction of the bank corporate clients.

Thus, the certain shortcomings are typical for most methods; the most significant are as follows: the impossibility or complexity of adaptation, paid basis, closedness, low information, lack of interpretation of results, lack of recommendations. All considered techniques are based on obtaining information from customers through questionnaires. The most applicable methods for analysing the satisfaction of corporate clients by the individual bank with specific characteristics and shortcomings were the CSI and Lambin methods.

4 Research Results

The study, which was conducted jointly with A. Simakov, involved 48 corporate clients of the bank. In order to determine the aspects of interaction that affect the overall level of satisfaction of corporate clients with the bank services, a questionnaire was conducted. The questionnaire included 28 questions divided into 6 groups:

1. Convenience of reception of services (*Location of bank branches; Availability of parking spaces for customers; The condition of the interiors of bank branches; Navigation within bank branches; Availability and relevance of information materials (booklets, necessary information);Office hours*);
2. Quality of services (*Consultations on the services offered; Accuracy and correctness of registration of documents; Flexibility of the conditions for the services provided (loans, deposits, cash services); Provision of services within the agreed period; Quality of technical support*);
3. Variety of services (*Range of services; Development of the system of remote banking services; Timing and number of lending programmes; Terms and amount of deposit programmes*);
4. Cost of services (*The cost of cash services; The cost of remote banking services; The cost of credit; Deposit interest*);
5. Bank staff (*The appearance of the bank staff; Competence and discipline of the bank staff; The rate of work of the bank staff; Orientation of personnel on customer problems*);
6. Attitude to the client (*Individual approach and personal participation in solving the problems of the customer; Establishment of confidential relations with bank staff; Participation of the bank management in solving the problems of clients; Politeness of the bank staff; The desire of the bank staff to correct inaccuracies and errors*).

It was necessary to assess separately the satisfaction with the services of the bank as a whole.

It is obvious any initial choice of aspects included in the questionnaire considers those aspects of interaction which might either have a weak relationship or even have no direct relationship with the satisfaction with the bank services as a whole.

Development of an intra-bank procedure for assessing customer satisfaction with corporate clients. The Lambin method has the following advantages: universality of application, freedom of choosing the aspects, assessment of satisfaction with specific aspects, clear criteria for interpreting results, visual representation of results. The disadvantages are the lack of recommendations on the selection of aspects and the deviations in assessing the variability of the aspects.

In order to eliminate the shortcomings, the following changes were made:

A. A pair correlation analysis is proposed as a possible way of selecting the interaction aspects which have strong direct relationship with the general level of satisfaction of corporate clients with the services of the bank as a whole;

B. In order to eliminate the shift in estimating the variability of the answers for each aspect, it was suggested to use the root-mean-square deviation instead of the unbiased standard deviation.

A. Reducing the number of analysed aspects of interaction is necessary in order to focus on the key aspects, which initially allows avoiding the cost of excess resources used for work with those that are weaker than others associated with the satisfaction of corporate customers with the bank services in general, and then focusing efforts on the aspects with the strongest interconnection; allows avoiding the inefficient financial costs for adjusting aspects of interaction having weak relationship.

First, it is necessary to determine whether the correlation coefficient is significantly different from zero, that is, whether there is a relationship in general; it requires selecting a significance level. In this case, it was chosen equal to 5%.

Thus, there were chosen aspects having a significant direct relationship with the overall level of satisfaction of corporate clients with the bank services.

Since the correlation coefficient indicating a direct relationship can take values from 0 to 1 then, for further analysis there have been selected aspects with value of coefficient above 0,5. This fact indicates their strong direct relationship with the general satisfaction of corporate clients with the bank services.

B. According to the Lambin method, it is necessary to calculate the mean and standard deviation for the selected aspects.

The root-mean-square deviation is defined as the square root of the variance of the random variable.

$$\delta = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \tag{1}$$

The sample variance (δ^2) is calculated from the data sample. Dispersion is the mean square deviation and reflects the measure of the spread of data around the average value.

To estimate the variability level of a characteristic, the sample usually uses a root from an unbiased sample variance, called unbiased standard sample deviation:

$$s = \sqrt{\frac{n}{n-1}\delta^2} = \sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \tag{2}$$

It can be seen from the formulae that, with a sufficiently large sample of data, the difference between the biased and unbiased dispersion practically disappears, and since a small sample size is assumed in the assessment of the satisfaction of corporate clients with bank services in operating offices and branches, the replacement of the root-mean-square deviation by an unbiased standard is reasonable.

The interaction aspects selected on the basis of the strength and orientation of the correlation relationship were analysed in accordance with the chosen method.

After calculating the standard deviations and means, the obtained values should be displayed graphically, deferring in the X-axis the average values of the satisfaction of corporate clients by each aspect of interaction with the bank, and in the Y-axis, the corresponding values of standard deviations (Table 1).

Table 1. Average values of satisfaction with interaction aspects and their variability.

No	Interaction aspects with the bank	Average values	Standard deviations
1	Consultations on the services offered	8.33	0.91
2	Accuracy and correctness of registration of documents	8.75	0.91
3	Flexibility of the conditions for the services provided (loans, deposits, cash services)	7.44	0.90
4	Provision of services within the agreed period	8.52	1.30
5	The cost of cash services	8.02	0.64
6	The cost of remote banking services	8.04	0.65
7	The cost of credit	7.48	0.74
8	Deposit interest	7.40	0.87
9	The appearance of the bank staff	8.63	0.49
10	Competence and discipline of the bank staff	8.21	0.80
11	The rate of work of the bank staff	7.92	0.74
12	Orientation of personnel on customer problems	8.33	1.10
13	Individual approach and personal participation in solving the problems of the customer	7.67	1.33
14	Establishment of confidential relations with bank staff	7.25	1.10
15	Participation of the bank management in solving the problems of clients	7.60	1.23
16	The desire of the bank staff to correct inaccuracies and errors	7.81	1.10

All aspects were divided into four quadrants (see Fig. 1).

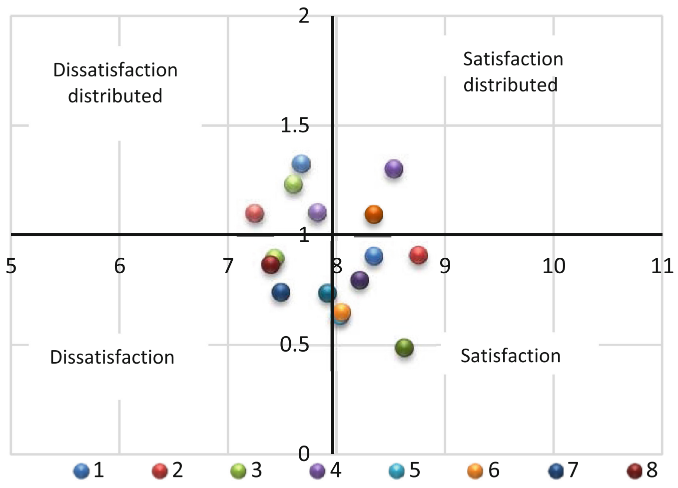


Fig. 1. Matrix of satisfaction/dissatisfaction with aspects of interaction of corporate clients with the bank.

In the right lower square of "Satisfaction" there are aspects whose assessment is above average, and the standard deviation is below one. Here the customers are generally satisfied with the quality. The following aspects of the bank interaction with corporate clients are in this quadrant:

1. *Consultations on the services offered;*
2. *Accuracy and correctness of registration of documents;*
5. *The cost of cash services;*
6. *The cost of remote banking services;*
9. *The appearance of the bank staff;*
10. *Competence and discipline of the bank staff.*

There two aspects in the upper right quadrant of "Satisfaction distributed"; their evaluation is higher than the average, and the standard deviation is higher than one:

4. *Provision of services within the agreed period;*
12. *Orientation of personnel on customer problems.*

This fact demonstrates that there is no unanimity in the assessments of clients, and this means that many of them are not satisfied with these aspects. It is necessary to identify these clients and, as a result of working with them, to understand the reason for their low satisfaction with these aspects for making corrective decisions.

The aspects of interaction with a low average score and a high standard deviation are in the left upper quadrant "Dissatisfaction distributed". The majority of corporate customers are dissatisfied with these aspects, but some of the evaluators put high marks. Usually this indicates the inconsistent nature of these aspects. The spread of estimates on these aspects in this case can be explained not so much by the inconsistency of quality, but rather by the fact that all of them in one way or another at the same time depend on the personal characteristics of the bank employees and clients, giving an extremely subjective character to the evaluation of clients. There are following aspects in this quadrant:

13. *Individual approach and personal participation in solving the problems of the customer;*
14. *Establishment of confidential relations with bank staff;*
15. *Participation of the bank management in solving the problems of clients;*
16. *The desire of the bank staff to correct inaccuracies and errors.*

The most problematic aspects are in the lower left quadrant of "Dissatisfaction"; most corporate clients have put them below average scores. This group includes:

3. *Flexibility of the conditions for the services provided (loans, deposits, cash services);*
7. *The cost of credit;*
8. *Deposit interest;*
11. *The rate of work of the bank staff.*

Customer satisfaction with these aspects is relatively low, and low values of standard deviations of these estimates in relevant aspects indicate that clients are mostly unanimous in their opinion.

When deciding on the priority of correcting aspects, it is first of all necessary to take into account the strength of their relationship with the level of satisfaction of corporate clients with the services of the whole bank.

The assessment of the satisfaction of corporate clients with the bank services based on the modified method of Lambin confirms its practical applicability and the opportunity to be the basis for an intra-bank procedure for assessing the satisfaction of corporate clients.

In general, the proposed method is reduced to the following four stages:

1. Choice of aspects of interaction of the bank with corporate clients;
2. Calculation for each of the aspects of average satisfaction values and standard deviation of satisfaction;
3. Graphical representation of results;
4. Interpretation of the results and recommendations in accordance with the Lambin method.

The practical applicability of the modified Lambin method allows proposing an intra-bank procedure for assessing the satisfaction of corporate clients, consisting of the following stages:

1. Development of a regulatory document including a method for assessing the level of satisfaction of corporate clients of the bank, agreement and approval, as well as its placement in the internal network of the bank, referral to operational offices and branches;
2. Sending the questionnaire to corporate clients through the Bank-Client system in accordance with the regulatory document, ensuring the reception and storage of questionnaires received from clients;
3. Processing the received questionnaires, evaluating and summarizing the results of customer surveys by specialists of departments dealing with corporate clients of operational offices and branches;
4. Preparation of a conclusion, including recommendations on improving the level of satisfaction of corporate customers with bank services;
5. Presentation of the conclusion to the heads of operational offices and branches, as well as to the head of the department for working with corporate clients.

5 Conclusions

1. Customer satisfaction is a multifaceted concept, and it is often treated as a subjective evaluation represents the difference between consumer's expectations and his perception of the real experience of using a product or service of the company. The higher the expectations of the client, the harder it can be to achieve his complete satisfaction. Customer satisfaction should to be measured and evaluated for making the appropriate management decisions.

2. The level of satisfaction of corporate clients with bank services is determined by people who have the authority to represent these clients in the bank on the basis of their experience of interaction with the bank.
3. There is no uniform method for assessing the level of customer satisfaction. Each organisation uses its own method and procedure for assessing the level of customer satisfaction. The peculiarity of banks is the presence of two different types of consumers of banking services – "individual customers" and "corporate clients" with different expectations and different aspects of interaction, which makes it impossible to assess the satisfaction of both types of clients simultaneously using the same method.
4. An analysis of the merits and demerits of the methodologies examined showed that Lambin method is the most useful for evaluating the level of satisfaction of corporate clients with the bank services. However, in our opinion, it requires certain improvements.
5. Lambin method has been improved, which makes it possible to propose an intra-bank procedure for assessing the level of satisfaction of corporate clients at the level of a separate office, branch or the whole bank.
6. The obtained results can be useful not only for banks in general, but they also can be applied in the operation process by managers of branches and operational offices for monitoring, timely detection and correction of problem areas of cooperation with corporate clients.

Acknowledgements. The author is grateful to Alexander Simakov for his contribution to the research.

References

1. Armstrong, G., Kotler, P.: Marketing: An Introduction, 12th edn. Prentice Hall, Upper Saddle River (2014)
2. Hill, N., Self, B., Roche, G.: Customer Satisfaction Measurement for ISO 9000:2000. Butterworth-Heinemann, Oxford (2002)
3. John, J.: An analysis on the customer loyalty in telecom sector. Special reference to Bharath Sanchar Nigam limited, India. Afr. J. Market. Manag. **3**(1), 1–5 (2011)
4. Subashini, R.: A review of service quality and customer satisfaction in banking services: global scenario. J. Internet Bank. Commer. **21**(5) (2016). http://www.icommercecentral.com/open-access/a-review-of-service-quality-and-customer-satisfaction-in-banking-services-global-scenario.pdf. Last Accessed 27 June 2017
5. Hennayake, H.M.G.Y.J.: Impact of service quality on customer satisfaction of public sector commercial banks: a study on rural economic context. Int. J. Sci. Res. Publ. **7**(2), 156–161 (2017)
6. Munari, L., Lelasi, F., Bajetta, L.: Customer satisfaction management in italian banks. Qual. Res. Financ. Mark. **5**, 139–160 (2013)
7. Blackwell, R.D., Miniard, P.W., Engel, J.F.: Consumer Behaviour. Harcourt College Publishers, London (2001)
8. Gustafsson, A., Johnson, M.D., Roos, I.: The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. J. Mark. **69**(4), 210–218 (2005)

9. Lee, G.: Customer Satisfaction and Loyalty. Internet project “Encyclopedia of Marketing” (2014). (in Russian). http://www.marketing.spb.ru/read/article/a47.htm. Last Accessed 21 Apr 2017
10. Zhukov, E.F.: Money. Credit. Banks. UNITY-DANA, Moscow (2010). (in Russian)
11. Manrai, L.A., Manrai, A.K.: A field study of customers’ switching behavior for bank services. J. Retail. Cons. Serv. **14**(3), 208–215 (2007)
12. Danchenok, L.A., Denisova, E.S.: Marketing tools of formation of loyalty of consumers of bank services in modern conditions. Market. Serv. **18**(12), 120–128 (2009). (in Russian)
13. Melnic, E.L.: Techniques for measuring customers’ satisfaction in Banks. Bull. Transilvania Univ. Braşov. Ser. V Econ. Sci. **9**(58) No. 1, 23 (2016)
14. Lambin, J.-J.: Strategic Marketing: A European Approach. McGraw-Hill Marketing for Professionals Series. McGraw-Hill Book, London (1993)
15. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A conceptual model of service quality and its implications for future research. J. Mark. **49**(4), 41–50 (1985)
16. Ramseook-Munhurrun, P.: Service quality in the public service. Int. J. Manag. Mark. Res. **3**(1), 37–50 (2010)
17. Lee, H., Lee, Y., Yoo, D.: The determinants of perceived service quality and its relationship with satisfaction. J. Serv. Mark. **14**(3), 217–231 (2000)
18. Cronin, J.J., Taylor, S.A.: Measuring service quality: a re-examination and extension. J. Mark. **56**(3), 55–68 (1992)
19. Avkiran, N.K.: Developing an instrument to measure customer service quality in branch banking. Int. J. Bank Mark. **12**(6), 10–18 (1994)
20. Dařena, F., Motyčka, A., Malo, R.: Customer satisfaction index calculation service. In: Conference ICL 2008, 24–26 September, Villach, Austria (2008)
21. Theacsi.org: About the American Customer Satisfaction Index (2016). http://www.theacsi.org/about-acsi. Last Accessed 27 July 2017
22. Epsi-rating.com About us – EPSI Rating Group (2016). http://www.epsi-rating.com/about-us/. Last Accessed 25 Mar 2017
23. Lambin, J.-J., Chumpitaz, R., Schuiling, I.: Market-Driven Management: Strategic and Operational Marketing, 2nd edn. Palgrave Macmillan, London (2007)

Socially Responsible Investing and Public Pension Fund Performance in Latvia

Irina Kuzmina-Merlino[1(✉)] and Svetlana Saksonova[2]

[1] Transport and Telecommunication Institute, Lomonosova street 1, Riga 1019, Latvia
Kuzmina.I@tsi.lv
[2] University of Latvia, 19 Raina Blvd., Riga 1586, Latvia
saksonova.svetlana@gmail.com

Abstract. Currently Latvia runs a three-pillar pension system. Despite positive dynamics of change of net assets of pension schemes, investments into pension programs are not regarded as financial products that could be popular amongst the population of Latvia. This problem is one of the key social economic issues in society and therefore its solution, which should be developed with consideration of ethical principles of investment, is socially important, which is why it should be addressed by the state. The aim of this research is to consider social and economic aspects of activities of the 2nd pillar pension fund in Latvia and to identify the attitude of Latvian residents to investments into pension programs. The authors' conclusions are based on the analysis and synthesis of scientific and economic literature, the official statistical data, and research of opinion of the Latvian population.

Keywords: Socially responsible investing · Performance
2nd pillar pension fund

1 Introduction

The importance of social responsibility in corporate activities and investment has widely been debated in academic and scientific literature because the object of research – socially responsible investment (SRI) – is linked to the concept of corporate social responsibility, which is addressed by various companies in their investment portfolios. The questions that are typically asked by investors in pertinence to such types of investments relate to the degree of safety of such investments and their rates of return in the future [1].

Other questions that are also important relate to the ethics of investment. Where will financial corporations and funds reinvest collected money in order to obtain profits in the future? Overall, this paper will examine the scope and implications of socially responsible investing by public pension funds.

Preliminary statistics show that pension fund assets grew in the Baltic countries in 2016. With the interest growing in sustainable investing, positive financial performance of many funds could be considered a positive trend [2, 3]. In the first half of 2016 the accrued pension capital increased by 7.5% or €175.4 million thus having reached the

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_57

sum of €2.5 billion by the end of June. However, the positive rate of return was demonstrated only by conservative schemes for that period (from 0.4% to 2.8%) [4]. The Bank of Latvia attributes low rates of return in pension schemes to their dependence on international financial markets, esp. stock exchange, their developments, uncertainty and high volatility, which eventually results in low interest rates on long-term investment [5]. Pension funds mainly remained exposed to equities and bonds – pension funds held more than 75% of their portfolios in equities and bonds [6].

According to the data published by OECD on Founded Pensions Indicators in 2017, the overall size of investments in the founded pension system in Latvia was assessed at 1.5% of Latvia's GDP in 2015 (for comparison, it was 14.8% in Estonia and 6.6% in Lithuania) [7]. The OECD calculates the total pension funds investment as a ration between all pension fund investments and the GDP in US dollar and as a percent of GDP. Some of the key reasons affecting these investments have been the level of salary, contributions to funds, periods of participation, rates of return in the funds, etc. [8, 9]. This paper suggests that informing the public about pension programs and the ethical aspects of investment policies of pension funds be significant factors for both natural persons and the state and this will impact the development of the market.

The aim of this research is to consider social economic aspects of activities of the 2nd pillar pension fund in Latvia and to identify the attitude of Latvian residents to investments into pension programs and their ethical aspects. In order to attain the aim, answers to the following research questions have been obtained:

1. What is social responsible investment and what criteria could be used to assess ethical aspects of investment?
2. What indicates characterize pension fund's performance?
3. To what extent is the Latvian population informed about the pension system, particularly, the 2nd pillar pension fund? To what extent is the issue of ethics of investment important for them?

Research results might be useful for investment management organizations because they might help to enhance the quality of performance management, increase the level of trust and raise the attractiveness of investment products.

2 Research Methodology

Research methodology was developed so that research questions could be addressed. The research was based on the analysis of scientific and academic literature, the results of the previous research of these authors, statistical information obtained from the following sources: OECD "Pension funds in figures", publications of the Latvian Association of Commercial Banks, etc.

A questionnaire was used to study the opinion of the Latvian population on the 2nd pillar pension fund. The target audience of the study included economically active residents of the Republic of Latvia who has participated or has been participating in the state funded or 2nd pillar pension scheme. The questionnaire was active during the period of 20 days from 25 November 2016 to 15 December 2016. The questionnaire contained 11 closed-answer questions, of which 7 questions focused on the study area and the other 4 were more of a personal nature in order to group and categorize participants. In order to calculate the confidence interval, the authors of the paper used the following formula [10]:

$$n = \frac{Z^2 * (p) * (1 - p)}{d^2}, \tag{1}$$

where: n – selection; Z – assessment (Z – factor); p – the percentage of responses (0.5 – according to the standard) and d – confidence interval.

Overall, 395 people took part in the study. According to the data published on the webpage www.manapensija.lv on 15 December 2016; the number of the participants of this study resulted in the confidence interval of ± 4.93% (the assessment 95%). The results of the study were analysed using statistical methods of information processing. The results of the analysis have been summarized as comments and conclusions consistently with the questions posed in the introduction of the study. Limitations of the research:

- The 2nd pillar pension funds managed and registered in Latvia are the open investment conservative pension funds.
- Only open (public) access enterprises were checked.
- The research period – the year 2016.
- Selection of the investigated indicators – return on equity and volatility.

3 Research Results

3.1 Socially Responsible Investment: Definition and Measurement Indexes

There is no consensus on the origin of the concept of *Socially Responsible Investment* in scientific literature. The first time the authors of the paper came across this term was when looking through European SRI Study (Eurosif) documents, in particularly, Table 1 contained various characteristics of this term that reflected both on the nature of the phenomenon and investment motivation. Traditionally the enterprises that are considered inappropriate for SRI are those that produce tobacco, alcohol and weapons, deal with gambling, pornography, experiment with animals (not for medical purposes), pollute the environment, conduct genetic manipulation and violate human rights. Obviously it is impossible to choose all criteria; therefore, the society assesses the investments of each enterprise in accord with the most vital criteria. When SRI has attained considerable volumes, the investment world has started to think about the development of the comparative index (benchmark) for SRI.

The most popular SRI indexes are listed below [12–14]:

- SAM/Dow Jones Sustainability Index (DJSI),
- EIRIS/FTSE4GOOD index,
- Vigeo/ASPI index,
- Sustainalytics index,
- Oekom index.

Table 1. Definitions of socially responsible investment [11].

Author(s)	Year	Characteristics
European SRI Study 2012 – Eurosif	2012	Terms: social, ethical, responsible, socially responsible, *sustainable* is frequently used in different areas, similar to the field of socially responsible investing (SRI). SRI is the value that incorporates various points of view, which creates a considerable challenge for an investor that is expected to act in accord with the concept
Satterfield, M.	2012	One of the key characteristics of these investors is the acceptance of the idea that the return on investment might be lower under the condition that this investment creates something good
European commission	2011	Responsibility of enterprises for their influence on society. In order to implement their corporate social responsibility, it is important to identify, avert and alleviate plausible negative effects.
Golosujs, S. & Kuzmina-Merlino, I.	2014	Socially responsible investing is a process which includes the conformity of various criteria and values that create the good for both the investor and society in general and that can ensure the simultaneous attainment of multiple goals, for example, general risk reduction and enhancement of the public image of the enterprise

Having grounds in theory, the authors of the paper developed 5 criteria that might be more suitable for socially responsible investing: production of Tobacco, production of alcohol, production of weapons, enterprises with legal problems (defence of human and employment rights), investments that are in public circulation in Latvia.

3.2 The Analysis of Responses of Latvian Participants

Most participants, that is, 173 people, were between the age of 26 and 35. They formed 44% from the overall participant number. The second largest group was the youth between the age of 19 and 25 – they were 147 people or 37%. The group of participants between the age of 36 and 45 included 54 people and those between the age of 46 and 55 were 21 persons, which was 5% of the overall number of participants. The average age of participants was 29.8 and the most frequently occurring age (moda) was equal to 25. The division based on sex was similar, that is, 195 participants (49%) were men and 200 participants (51%) were women.

The question "*Have you started to think about your pension capital?*" has received "Yes" answer from 267 participants, which is 67.6% and which means that overall people are interested in learning more about the pension system of the Republic of Latvia and the ways their pension capital has been accrued.

The criterion that has been indicated as the most valuable (Table 2) was the return and its average indicator is 4.1. The average mark for the investment structure was 3.7. This means that people find it important to know where pension schemes invest their money into. This finding is also supported by the answer to the question "Is it important

Table 2. Assessment scale of various criteria.

	Importance of criteria*					Total	Average
Criterion	1	2	3	4	5		
Fund return	15	31	56	96	197	395	4.1
Advertising of the pension scheme	132	130	98	32	3	395	2.1
Offered gifts	129	96	86	47	37	395	2.4
Investment structure	21	43	81	127	123	395	3.7
Commission expense	37	54	79	126	99	395	3.5
Total	334	354	400	428	459	1975	3.2

*Importance of Criteria: 1 – low, 5 – high.

for you to know where your pension scheme invests money?" The answer "Yes" was received from 255 participants.

Having analysed the answers to the question "What was it that mostly affected your decision to choose the 2nd pillar pension scheme?", it was concluded that the people's decision was mostly impacted by initiatives of sellers and banks. This answer was produced by 172 people, which constituted 43.5% of the overall participant number. In order to determine the overall awareness of the 2nd pension pillar, participants were presented with a direct question "Are you informed about the 2nd pension pillar?" The answers to this question showed that generally the public is informed about the state funded pension scheme or the 2nd pension pillar. Of all participants, 167 people or 42.7% answered positively and 110 or 27.8% answered they had limited knowledge about the scheme. Thus, people are aware of the existence of the 2nd pension pillar in Latvia but they do not have detailed knowledge about it.

The majority of participants, 253 people or 64% positively answered the question of "*Is it important for you where your pension scheme invests money?*" This is yet evidence that affirms the importance of SRI for people when choosing a pension scheme. Another question was "*What factors could affect your decision to choose the 2nd pension pillar*?" Two major answers were obtained. "The good reputation of the money manager" was important for 72.3% and "the investment into potentially significant return" was indicated by 60.8%. A smaller number of participants, 88 or 25%, highlighted the importance of investment in Latvia – it was a positive factor for them. The authors of the paper highly evaluate this answer because it suggests that people find SRI criteria important. The obtained responses, therefore, clearly indicate that SRI criteria are important in Latvia.

It is interesting to consider dependences between different factors. Data that are presented in Table 3 were obtained using simple correlation coefficient formula [15] and they show the dependence between the age and the assessment criteria.

The obtained results have allowed the authors to make conclusion: there is no correlation between the age and specific SRI factors – they do not depend on each other, which means that people choose criteria independently of their age. There is a moderately strong correlation between the fund return and the investment structure – when participants highly value the return, they also highly value the investment structure, in other words, the higher the fund return, the higher the investment structure.

Table 3. Correlation analysis of primary data (1) (Pearson correlation, Sig. (2-tailed), $N = 395$).

Question	Age	Fund return	Advertising of pension scheme	Offered gifts	Investment structure	Commission expenses
Age of participant	1.000					
Fund return	0.251**	1.000				
Advertising of pension scheme	0.087	0.009	1.000			
Offered gifts	–0.150**	−0.105*	0.416**	1.000		
Investment structure	0.294**	0.340**	–0.122**	–0.231**	1.000	
Commission expenses	0.170**	0.420**	–0.074	–0.128*	0.386**	1.000

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Another instance of a moderately strong correlation was presented by the correlation between the return and commission expenses – the higher value assigned to return, the higher number appears in the expenses column. Yet another moderately strong correlation was obtained for the commission expenses and investment structure – the higher points on commission, the higher points on the investment structure.

Correlation analyses results presented in Table 4 show a moderately strong correlation among such factors as the age of participants, knowledge of the pension pillar and the importance of the investment structure.

Table 4. Correlation analysis of primary data (2) (Pearson Correlation, Sig. (2-tailed), $N = 395$).

	Your age?	Are you aware of the 2nd pension pillar?	Is it important for you where your pension scheme invests money?
Age of participant	1.000		
Are you aware of the 2nd pension pillar?	–0.328**	1.000	
Is it important for you where your pension scheme invests money?	–0.344**	0.411**	1.000

** Correlation is significant at the 0.01 level (2-tailed).

In order to test the hypothesis about whether empirical distribution is consistent with the assumed theoretical distribution, the authors have used Pearson's chi-squared (χ^2) test, whose test statistic is given by the formula [17]:

$$\chi^2_{emp} = \sum_{i=1}^{n} \frac{(\chi_i - \overline{\chi}_i)^2}{\overline{\chi}_i}, \tag{2}$$

where: n – number of bins (essentially answers to specific questions); χ_i – observed frequency for bin i (answers to specific equations); $\overline{\chi}_i$ – theoretical frequency for bin i.

This criterion for hypothesis testing was applied for analysis of answers on the research question "What factors could positively affect your choice of the 2nd pension pillar?" To conduct the analysis consistently with the Pearson (χ^2) criteria, two opposite hypotheses were formulated:

- H_0 – different groups of participants have the same views on a specific question,
- H_1 – different groups of participants have different views on a specific question.

Participants were grouped as follows:

- Group 1 – men who are 35 or younger,
- Group 2 – men who are older than 35,
- Group 3 – women who are 35 or younger,
- Group 4 – women who are older than 35.

Using formula (2) the real value of the criteria $\chi^2_{emp} = 25.91 \ = 25.91$ was obtained and the theoretical value corresponds to $\chi^2_{teor} \ = 21.03$. Because $\chi^2_{\mathrm{emp}} > \chi^2_{\mathrm{teor}}$ $(25.91 > 21.03)$ with the significant level $\alpha \ = 0.05$ the alternative hypothesis H_1 was accepted, because all four groups had different views on this question. The obtained results suggest:

- Men who are older than 35 are prepared to choose a pension scheme with a potentially greater rate of return significantly more often than other groups. Another criterion that is more important to them than to other groups is the reputation of the managing company.
- Investment into enterprises with excellent reputation is significantly more important to people who are younger than 35 irrespective of their sex.
- Offered gifts have more importance to men who are older than 35 and to women who are younger than 35.
- Investments into Latvia are equally important to all groups, which means that age and sex do not affect people's choice in this case.

3.3 The 2nd Pension Pillar Performance

Since July 2001, there has been a 3 pillar pension system in Latvia. According to www.manapensija.lv, it includes:

- The 1st pillar (state compulsory unfunded pension scheme),
- The 2nd (state funded pension scheme),
- The 3rd pillar (private voluntary pension scheme).

Depending on the risk level of each system, three levels of funds are typically distinguished: active, conservative and balanced. This paper has focused only on the

2nd pillar pension fund and on indicators of the conservative scheme. The conservative scheme is a plan which invests money into stable and safe securities. This investment plan is with small profitability, but it is not as volatile as investments in market shares. At the beginning of 2017, there were 21 pension schemes of the 2nd pension pillar in Latvia, 8 of which were conservative. They are presented in Table 5. As Table 5 shows, the value of assets of conservative plans equals €716.7 million or 25.9% of the 2nd pension pillar's total value of assets. The number of participants in conservative plans equals 26.5% of the total number of participants in the 2nd pension pillar.

Table 5. The 2nd pension pillar's conservative plans.

2nd pension pillar	Total value of assets		Number of participants	
Conservative plans	EUR	% of total	Person	% of total
IPAS CBL Asset Management	105 620 979	14.7	51 317	15.2
IPAS DNB Asset Management	71 082 691	9.9	29 296	8.7
INVL 58+conservative plan	20 647 571	2.9	19 157	5.7
Nordea Luminor Investment Plan	33 825 071	4.7	11 813	3.5
Norvik IPS Plan "Daugava"	30 692 562	4.3	25 072	7.4
SEB Conservative Plan	17 116 978	2.4	9 569	2.8
SEB Latvian Plan	162 437 051	22.7	75 650	22.5
Swedbank Pension Plan "Stability	275 259 983	38.4	114 789	34.1
Total	716 682 886	100.0	336 663	100.0

Source: estimates of the authors of this article based on the data published on www.manapensija.lv (02.01.2017).

According to OECD data on the Baltic Sea region, all pension funds exhibit positive real investment rates of return, but average returns on pension fund investment are in Baltic countries (0.9% in Latvia; 1.0% in Estonia; 2.7% in Lithuania). For comparative analysis, data of Teachers Insurance and Annuity Association of America (TIAA) was taken. The TIAA analysis of leading US responsible investing equity shows that average annual return (performance) as of 31.12.2016 for the top performed SRI companies ranged from 5.96% to 7.05%; the average annualized standard deviations for the RI indexes ranged from 16.21% to 17.55% and the average annual Sharpe ratio ranged 0.69 over the past 10 years [19].

The key indicators of the investment attractiveness of a pension fund are return on equity and volatility. In contrast, volatility is an indicator that characterizes price changes of financial instruments and the degree of risk when using financial instruments at a particular point in time. According to formula (3) the fund return was calculated as a relation of a scheme part value at the end of a period to its value at the beginning of the same period. This relationship is expressed in annual percentage of a degree, where the remainder is 365, but the divisor is the number of days over the calculated period [20, 21]:

$$I = \left(\frac{C_b}{C_s}\right)^{\frac{n}{365}}, \tag{3}$$

where: I – fund return; C_b – part value (or price) at the end of the period; C_s – part value (or price) at the beginning of the period; n – number of days from the beginning to the end of the period.

Main performance indicators of the pension conservative funds activities are presented in Table 6.

Table 6. Pension conservative plans performance at the end of 2016 [22] (prepared by the authors using data published on www.manapensija.lv).

	Yearly performance in % as at 31.12.2016	Volatility	Risk level (2 –low)
IPAS DNB Asset Management	0.75	1.3%	2
IPAS CBL Assets Management	3.01	0.5–2%	2
INVL 58+Conservative Plan	2.08	1.55%	2
Nordea Luminor Investment Plan	2.16	1.5%	2
NORVIK IPS plan “Daugava”	4.95	1.0%	2
SEB Conservative Plan	2.07	1.0%	2
SEB Latvian Plan	0.84	0.5–1.99%	2
Swedbank Pension Plan “Stability	0.36	–0.22%	2

The indicators of the average return of conservative plans range from 0.36% to 4.95%. These data are somewhat different from those published by the Financial and Capital Market Commission of Latvia. As it was mentioned earlier, their data showed the average return of conservative plans within the range of 0.4% to 2.8% [23]. The calculated indicators of volatility suggest low risk of investment into the 2nd pension pillar schemes.

4 Conclusions

The following conclusions were made based on the findings of this research:

1. Socially responsible investing is a process which includes the conformity of different criteria and values, which creates the good for the investor and society overall and which can ensure the simultaneous attainment of multiple goals.
2. SRI portfolios can include multiple goals simultaneously, for example, reduction of overall investment risks and the enhancement of the public image of the enterprise.
3. Overall awareness of the 2nd pillar pension fund in Latvia in 2016 was lower than the average.
4. Latvian people are able to identify factors that affect the results of 2nd pillar pension fund in a very direct way using indirect factors.
5. Latvian people irrespective of their age and sex consider investments into the Latvian economy important.

6. Five significant SRI assessment criteria have been identified: (1) production of tobacco; (2) production of alcohol; (3) production of weapons; (4) enterprises with legal problems (defence of human and employment rights); (5) investment into the local market (in public circulation in Latvia).
7. Average return on pension fund investment in Latvia is low (0.9% in 2016), but positive. Over the period of 2016 the difference between return and volatility was non-significant. This means that socially responsible investments can ensure similar returns to compared volatility.

References

1. Weber, O.: The financial sector impact on sustainable development. J. Sustain. Financ. Invest. **4**(1), 1–8 (2014)
2. Krosinsky, C., Robins, N.: Sustainable Investing: Revolutions in Theory and Practice, 2nd edn. Routledge, New York (2016)
3. Schroder, M.: Financial effects of corporate social responsibility: a literature review. J. Sustain. Financ. Invest. **4**(4), 337–350 (2014)
4. LKA, Association of Latvian Commercial Banks: Pension Fund Review of the Association of Latvian Commercial Banks (ALCB) for 2016, No 20, Riga (2017)
5. Latvijas Banka homepage. Financial Stability Report (2017). https://www.bank.lv/en/publications-r/financial-stability-report. Last Accessed 07 June 2017
6. Paakkinen, R.: Pension in CEE: New dawns for Baltic pensions in: Investment and Pensions in Europe, IPE (2017). https://www.ipe.com/pensions/pensions-in/cee/pensions-in-central-and-eastern-europe-new-dawns-for-baltic-pensions/10016931.article. Last Accessed 10 July 2017
7. OECD: Pension Funds in Figures, Pension Funds assets grew in most economies in 2016 (2017). https://www.oecd.org/daf/fin/private-pensions/Pension-Funds-in-Figures-2017.pdf. Last Accessed 22 July 2017
8. Lieksnis, R.: Evaluating the financial performance of Latvian and Estonian second-pillar pension funds. Res. Econ. Bus. Cent. East. Eur. **2**(2), 54–70 (2010)
9. Mavlutova, I., Titova, S., Fomins, A.: Pension system in changing economic environment: case of Latvia. Procedia Econ. Financ. **39**, 219–228 (2016)
10. Shenoy, G.V., Srivastave, U.K.: Business Statistics. New Age International, New Delhi (2005)
11. Kuzmina-Merlino, I., Golosujs, S.: Social responsible investing in Latvia. In: Sauka, A., Rivza, B. (eds.) Latvian Economics and Business Challenges, Monograph, pp. 64–81. Ventspils University of Applied Sciences, Ventspils (2014)
12. Cheng, J.F.: Domini Social Investments: A CSR Case Study. University of California, San Diego (2007)
13. Domini, A.: Socially Responsible Investing: Making a Difference and Making Money. Kaplan Business, Chicago (2001)
14. Trog, U.: SRI – Socially Responsible Investments, Integrating personal values and societal concerns with investment decisions. Eco Design Foundation (2001). http://www.changedesign.org/Resources/EDFPublications/Articles/Papers/SRI.pdf. Last Accessed 01 July 2017
15. Peck, R., DeVore, J.L.: Statistics: The Exploration & Analysis of Data. California Polytechnic State University. Cengage Learning (2011)

16. Hill, T., Lewicki, P.: Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining. StatSoft, Vermont (2006)
17. Jackson, S.L.: Statistics: Plain and Simple. Jacksonville University. Cengage Learning (2009)
18. ManaPensija: 2nd pension pillar (2017). www.manapensija.lv. Last Accessed 29 July 2017
19. O'Brien, A., Liao, L., Campagna, J.: Responsible Investing: Delivering Competitive Performance. NUVEEN TIAA Investment (2017)
20. Frensidy, B.: Financial Mathematics. Salemba Empat, Jakarta (2008)
21. LR Financial and Capital Market Commission homepage: 2011.gada 11.novembra Legal regulations Nr. 240 for investors (2011). https://m.likumi.lv/doc.php?id=239867. Last Accessed 20 June 2017
22. ManaPensija: 2nd pension pillar, Current statistics (2017). http://www.manapensija.lv/en/2nd-pension-pillar/statistics/. Last Accessed 10 July 2017
23. LR Financial and Capital Market Commission homepage: Financial and Capital Market Report in the 2nd quarter of 2016. State pension plans' resources management (2016). http://www.fktk.lv/attachments/article/5948/Parskats_Q2_2016_09092016.pdf. Last Accessed 11 May 2017

The Role of Productivity in Increasing Latvian Competitiveness

Ilze Sproģe[1(✉)], Sandra Jekabsone[2(✉)], and Irina Skribane[2(✉)]

[1] Transport and Telecommunication Institute, Lomonosova street 1, Riga 1019, Latvia
sproge.i@tsi.lv
[2] University of Latvia, Aspāzijas bulv. 5, Riga 1050, Latvia
{sandra.jekabsone,irina.skribane}@lu.lv

Abstract. Productivity is the main key factor for Latvia to increased prosperity, ensuring efficient use of resources. Due to limited availability of resources and the increasing competition in the global markets, it is important for Latvia to support its economic growth through increased productivity.

One of the main challenges for Latvia is the creation of new competitive advantages that are associated with investments in the latest technologies, innovation, research, human capital, efficient allocation of resources and redistribution that comes with the behavioural changes of economic subjects. Increasing entrepreneurs' motivation is a major structural change in policy making. Economic structural transformation process is largely dependent on the quality of the institutional framework (legislation, state aid and economic and political institutions), which provides goods and resources market efficiency, minimizing the redistribution process costs and risks, thereby strengthening the country's competitive benefits.

The aim of the research is to assess the potential for productivity growth to improve competitiveness of Latvia to identify the main obstacles that limit the attraction of resources and redistribution of higher value-added products and prepare recommendations for policies on the micro, sectoral and macro level.

Results of the analysis highlighted the potential economic policies to improve productivity on the micro, sectoral and macro level. One of the main conditions for a balanced development of the economics is the ability to reduce the productivity gap, achieving the most rapid productivity convergence with the EU average, while maintaining high wage increase rates. This can be done either by attracting additional investments, which is currently problematic considering investors' uncertainty, or to increase total factor productivity through structural reforms, improving innovation capacity, efficient use of resources and developing the high-tech industry.

Keywords: Added value · Labour cost · Productivity · Competitiveness

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_58

1 Introduction

Since 2010 the productivity of Latvia's economy has been at the level of 40–45% of the EU average [3]. Although in recent years productivity growth rate was faster than the EU average, but labour costs grew almost twice the rate and this can adversely affect competitiveness of Latvia. A further increase in labour costs is inevitable in the open labour market conditions, therefore, strengthening the competitiveness of Latvian is largely determined by the ability to reduce the productivity gap with the advanced economies.

During the period from 2011 to 2014, the growth of the Latvian economy on average reached 3.8% annually, but in the last two years the growth has slowed down [13]. During post-crisis period Latvia's economy has experienced significant macroeconomic adjustments (including stabilization policies) that prevented economic imbalances, but they were insufficient to increase productivity and strengthen competitiveness, which are the main prerequisites for a stable growth momentum to ensure the prosperity and convergence. Economics of Latvia is macroeconomically balanced now, but there is lack of incentives for rapid growth. There are no explicit trends, which allows to define main development impulses. In short term growth speed is defined by different short-term conjuncture obstacles and they are not promising lasting contribution to provisions of further growth.

Economic activities in general have not reached the pre-crisis level. Whereas dynamics of competitiveness rates shows, that model of economy of Latvia has not changed. It is based on benefits of the low labour costs. Remaining at the current situation, economic growth rates in the medium term can be achieved only 2–3% per year, which is not sufficient to ensure a stable income convergence with the highly developed countries and that the Latvian economy has the risk to get into the middle-income-trap.

Macroeconomic stability of economics is one of the important conditions for sustainable development, but is not sufficient to increase the productivity and competitiveness, which are the main conditions for stable growth rates and welfare. In the future Latvia needs a sustainable economic model where the ability of production of goods and services with higher added value is essential and that would generate higher income [8]. Economy with high productivity level, which cannot be based on benefits of the low labour costs. It should be considered, that in open labour market increase of labour costs is inevitable. Latvia can lose competitiveness in low labour cost field faster than benefits in goods production with high added value. Macroeconomic stability of Latvia simultaneously with weak growth rates is defined as "bad balance" and delay in this situation increases possibility to get in medium-income-trap with minimal convergence chances [6]. Therefore, stable convergence of Latvia and competitiveness could be ensured only by decreasing the productivity gap between modern economics. Productivity increase is actual modern problem in many, including highly developed countries. Productivity growth rate is slowing around the world - in the period from 2010 to 2015 the EU average productivity growth was 0.7% per year [3]. Although Latvia's productivity growth rate in recent years was faster than the EU average, labour costs are above almost twice the rate of EU and therefor with negative impact on the competitiveness of local producers. Since labour cost inevitable will rise in Latvia, to keep international competitiveness it is necessary to increase productivity of Latvia and reach at least half of EU level within next 3 years.

The aim of the research is to assess the potential for productivity growth to improve competitiveness of Latvia to identify the main obstacles that limit the attraction of resources and redistribution of higher value-added products and prepare recommendations for policies on the micro, sectoral and macro level.

Research methods to be used in the research: Literature review on the productivity, empirical analysis of data. The main research sources include the information available in the databases of the CSB and Eurostat, as well as the studies and publications on the productivity by the Ministries of Economics Republic of Latvia, Bank of Latvia, European Commission, OECD, World Bank and International Monetary Fund.

Results of the analysis highlighted the potential economic policies to improve productivity on the micro, sectoral and macro level. One of the main conditions for a balanced development of the economics is the ability to reduce the productivity gap, achieving the most rapid productivity convergence with the EU average, while maintaining high wage increase rates. This can be done either by attracting additional investments, which is currently problematic considering investors' uncertainty, or to increase total factor productivity through structural reforms, improving innovation capacity, efficient use of resources and developing the high-tech industry.

2 Theoretical Analysis of Productivity Growth Impact Factors

Productivity is commonly defined as a ratio of a volume measure of output to a volume measure of input use [15]. Productivity is not just about doing things more efficiently by "doing things right", it is also about achieving maximum effectiveness by "doing the right things". Thus, higher productivity can be achieved through efficient and effective use of resources such as labour, capital and materials in the production of various goods and services. The objectives of productivity measurement include [15]:

- Technology (a frequently stated objective of measuring productivity growth is to trace technical change);
- Efficiency;
- Real cost savings;
- Benchmarking production processes;
- Living standards.

Productivity can be measured across various levels (e.g. national, industry, organization and operational) and different sectors (e.g. manufacturing and services). On the national economy level, productivity can be estimated and expressed in GDP per person employed, which shows how much of the total income in a specified period is generated from one worker. Industry level productivity is measured as a value added per employee (by dividing the industry added value with the number of people employed), while the level of individual companies often uses a variety of physical parameters, such as the number of parts produced per 1 employee. In general, for small and open economies productivity on macro level is determined by average value added of exports per one worker. Exports determine countries competitiveness – either it is based on high technologies or low cost competitiveness [20].

Literature review shows that productivity is linked to many influencing factors, starting with R&D spending and innovations [10], structural changes in the economy [4, 5, 12, 16], ecological perspective [9, 11] and even quality management systems [1]. Factors that influence productivity are related to technologies, knowledge, exclusive resources (specific natural resources), prices of the resources (labour, electricity etc.), management, brand name and many other (see Fig. 1).

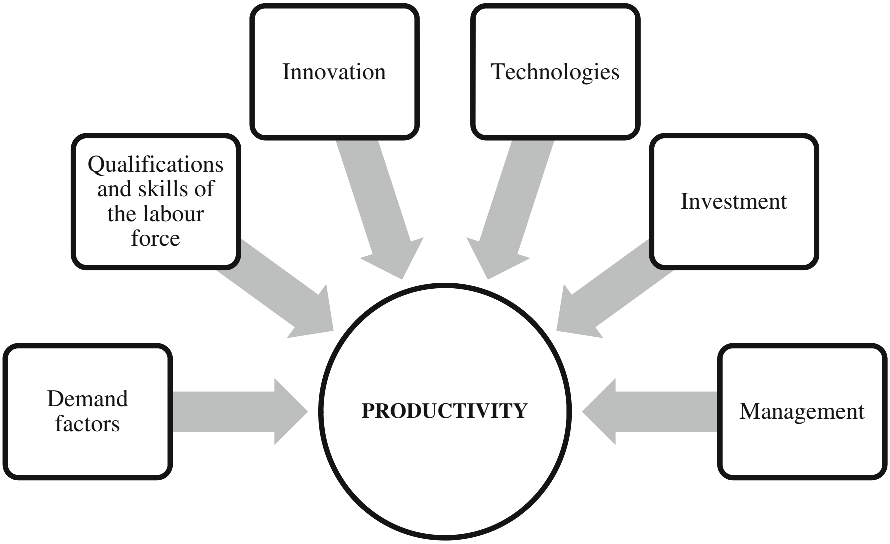


Fig. 1. Productivity influencing factors. Source: author's construction.

The increase in productivity is determined by several factors, such as:

– structures that are related to scientific and technical progress in the role of intensification of production;
– socio-economic, which is mainly related to investment in human capital (education of human capital, training, knowledge, motivating people to be productive);
– organizational, which are related to the production process organization and management, production specialization and concentration of production territorially, as well as horizontal and vertical cross-link establishments [7].

The main problem of all these above-mentioned activities is how to allocate investments to increase productivity between employers, workers, and the state. Technology development key contributions, of course, are done by operators. State aid is related to the promotion and scientific research base. However, it was the state that had a key role to play in development, but it also increases the individual contribution. Less developed is the collaboration between business and vocational education and lifelong learning programs and has its own reasons. Return from investment in business education is not clear and has a higher risk (the workers can change jobs, employee qualifications obtained by visiting these or other training programs may not meet a host of needs, it requires time). Organizational factors are mostly corporate responsibility. Latvia has currently poorly developed such macro-level measures, such as clusters and all related activities are not conscious of their role in increasing productivity.

In a well-functioning economy productivity growth drives:

1. Global frontier firms innovate and these technologies diffuse to other firms, raising *within-firm* productivity;
2. Efficient reallocation underpins the growth of productive firms, also via new entry and the downsizing and exit of less productive firms;
3. As the most productive firms gain market shares aggregate productivity grows.

But the process of reallocating resources does not always happen automatically, there are several factors that hinder these processes, such as unfair competition, administrative cross-border etc.

3 Productivity and Competitiveness of Latvia

After the global financial crisis, the model of economic growth in Latvia changed, the economy has become more stable and balanced as a result of macroeconomic adjustments and decreasing internal and external proportions. Nevertheless, in the rankings of competitiveness which are regularly published by the World Economic Forum (WEF), Latvia significantly lags behind other recent EU countries (the Czech Republic, Poland, Slovenia, Slovakia and the other Baltic States), and particularly in indicators related to the development of innovation systems [8]. This is mostly because manufacturing is a small proportion of Latvian GDP and because of the industrial sub-sector's technological structure, where low technology industries are dominating (they amount to 60% of total manufacturing added value), altogether this is why there is such low level of productivity in manufacturing and in whole national economy (see Fig. 2).

In 2016 GDP per person employed was 44.2% from the EU-28 average, meanwhile GDP per capita was 65% [3]. Even though Latvia's productivity has grown significantly since 2000, it's still far behind EU-28 average level. At the same time, Latvia's productivity is one of the fastest growing in the EU and more than five times faster than EU average (see Fig. 3).

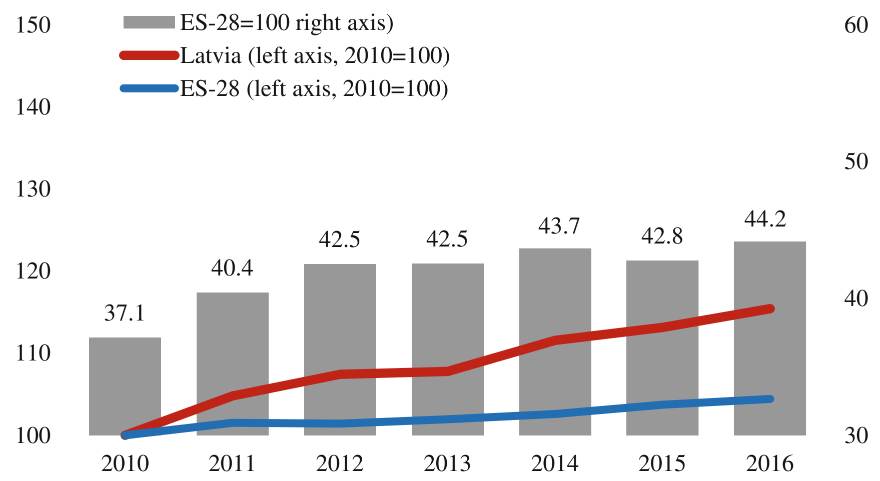


Fig. 2. Productivity (GDP per person employed) in Latvia and EU. Source: author's construction based on. Eurostat databases.

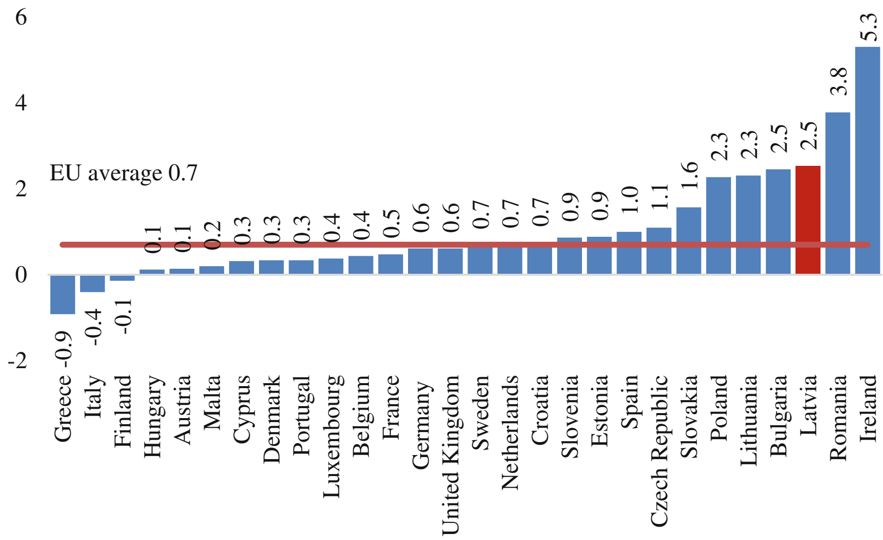


Fig. 3. Productivity growth in EU countries 2010–2015 (% changes, annual average) Source: author's construction based on. Eurostat databases.

From 2010 to 2015 Latvia's lagging behind in terms of productivity index in the national economy fell by 6% points in total, but in the industry – by 1.5% points. At the same time, we can observe rapid growth of labour cost (see Fig. 4). It is mainly related to low cost level (in 2016, labour costs per employed in the economy of Latvia were 41.7% of the EU average in total). From 2010 till 2016 average growth of wages in Latvia reached three times larger than in EU average [3].

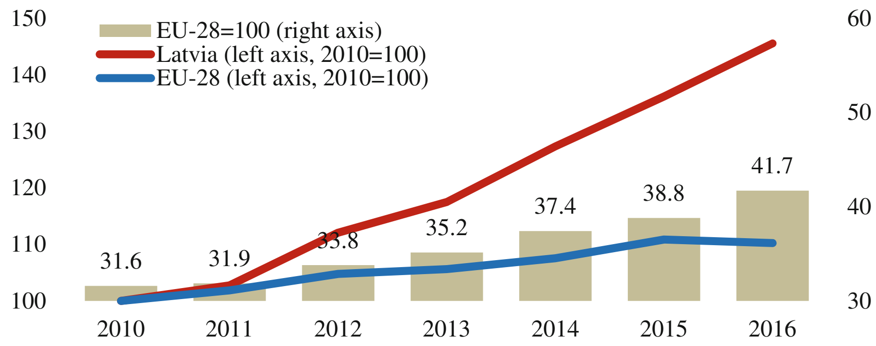


Fig. 4. Labour cost in Latvia and EU. Source: author's construction based on. Eurostat databases.

In recent years, the dynamics of labour costs and productivity were largely determined by factors of structural nature. With the economic growth resuming, wage growth is becoming more rapid, substantially due to the growing competition in the EU labour market and the low competitiveness of Latvia in the said market. By contrast, growth of productivity has been more moderate. It means that the advantages of cheap labour cost competitiveness are being gradually lost.

As shown by the unit labour cost (ULC[1]) dynamics, wages increased almost five times faster than productivity from 2004 to 2007, negatively affecting the international competitiveness of Latvia. Adjustments in product and labour markets due to the crisis reduced the gap between productivity and labour cost dynamics, resulting in Latvia's competitiveness improving gradually in foreign markets. However, there has been an increase in the ULC in recent years, which have been more rapid than the EU average (see Fig. 5).

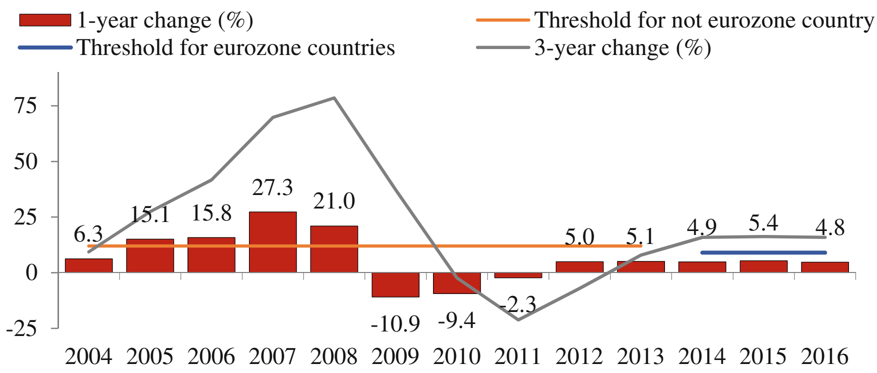


Fig. 5. Nominal unit labour costs (3-year % change). Source: author's construction based on. Eurostat databases.

[1] ULC is the ratio between labor costs and productivity. If productivity grows faster than wages, then ULC decreases, which indicates that the national cost competitiveness is increasing, and vice versa.

Dynamics of ULC in last year's show, that the risk for Latvian producers into lose competitiveness in EU markets. The question about productivity increasing becomes more actual. Low productivity level is evidence, that economical system does not ensure effective redistribution resources to produce good with highly added value. Therefore, for increasing of productivity it is important to support the innovations, to ensure efficiency of goods and the resource market, by minimizing of costs and risk related to redistribution processes, so benefits of competitive country will be strengthened.

Shift-share analysis of labour productivity growth shows that within-sector productivity increase largely explain overall productivity development. However productivity gains resulting from the movement of labour from less productive to more productive sectors (the so-called "shift effect") were relatively small (about 0.5% points of labour productivity growth in the period of 2010–2015).

4 Conclusions and Recommendations

Since 2010 the productivity of Latvia's economy has been at the level of 40–45% of the EU average. Although in recent years productivity growth rate was faster than the EU average, but labour costs grew almost twice the rate and this can adversely affect competitiveness of Latvia. Wages and salaries are the key component of labour costs. A further increase in labour costs is inevitable in the open labour market conditions, therefore, strengthening the competitiveness of Latvian is largely determined by the ability to reduce the productivity gap with the advanced economies.

One of the main challenges for Latvia is the creation of new competitive advantages that are associated with investments in the latest technologies, innovation, research, human capital, efficient allocation of resources and redistribution that comes with the behavioural changes of economic subjects. Increasing entrepreneurs' motivation is a major structural change in policy making. Economic structural transformation process is largely dependent on the quality of the institutional framework (legislation, state aid and economic and political institutions), which provides goods and resources market efficiency, minimizing the redistribution process costs and risks, thereby strengthening the country's competitive benefits.

Key productivity challenges: 1. overcoming « medium income trap » - promoting rebalancing the economy towards the higher value added sectors, 2. labour market (shrinking population and aging labour force, mismatch of skills demand and supply), 3. addressing the weaknesses in the business environment.

References

1. Albulescu, C.T., Drăghici, A., Fistiş, G.M., Truşculescu, A.: Does ISO 9001 quality certification influence labor productivity in EU-27? Procedia Soc. Behav. Sci. **221**, pp. 278–286 (2016). http://doi.org/10.1016/j.sbspro.2016.05.116
2. Central Statistical Bureau (CSB) of Latvia Statistic Database (2017). http://www.csb.gov.lv/dati/statistikas-datubazes-28270.html. Accessed 16 May 2017
3. Eurostat 2017. Statistic Database. http://ec.europa.eu/eurostat/data/database. Accessed 10 March 2017

4. Harada, T.: Structural change and economic growth with relation-specific investment. Struct. Change Econ. Dyn. **32**, 1–10 (2015). https://doi.org/10.1016/j.strueco.2014.12.002
5. Hartwig, J.: Structural change, aggregate demand and employment dynamics in the OECD 1970–2010. Struct. Change Econ. Dyn. **34**, 36–45 (2015). http://doi.org/10.1016/j.strueco.2015.06.001
6. Im, F.G., Rosenblatt, D.: Middle-income traps: a conceptual and empirical survey. J. Int. Commer. Econ. Policy **6**(3), 1550013 (2015). http://doi.org/10.1142/S1793993315500131
7. Jekabsone, S., Skribāne, I.: Structural weaknesses and challenges of economics growth of Latvia, Social Research Nr.1, Siauliai University, pp. 74–85 (2014)
8. Jēkabsone, S., Skribāne, I.: Smart specialisation strategy: realisation opportunities and problems in Latvia. Humanit. Soc. Sci. Latvia **24**(1), 153–161 (2016). University of Latvia Press
9. Yörük, B.K., Zaim, O.: Productivity growth in OECD countries: a comparison with Malmquist indices. J. Comp. Econ. **33**(2), 401–420 (2005). http://doi.org/10.1016/j.jce.2005.03.011
10. Lopez-Rodriguez, J., Martinez-Lopez, D.: Looking beyond the R&D effects on innovation: the contribution of non-R&D activities to total factor productivity growth in the EU. Struct. Change Econ. Dyn. **40**, 37–45 (2017). https://doi.org/10.1016/j.strueco.2016.11.002
11. Mahlberg, B., Luptacik, M., Sahoo, B.K.: Examining the drivers of total factor productivity change with an illustrative example of 14 EU countries. Ecol. Econ. **72**, 60–69 (2011). http://doi.org/10.1016/j.ecolecon.2011.10.001
12. Maudos, J., Pastor, J.M., Serrano, L.: Explaining the US-EU productivity growth gap: structural change vs. intra-sectoral effect. Econ. Lett. **100**(2), 311–313 (2008). http://doi.org/10.1016/j.econlet.2008.02.017
13. Ministry of Economics, The National Economic of Latvia Macroeconomics Review #69/2016-4, Riga: Ministry of Economics of the Republic of Latvia
14. Minniti, A., Venturini, F.: The long-run growth effects of R&D policy. Res. Policy **46**(1), 316–326 (2017). http://doi.org/10.1016/j.respol.2016.11.006
15. OECD, Measuring Productivity (2011). https://www.oecd.org/std/productivity-stats/2352458.pdf. Accessed 16 June 2017
16. Padilla-Pérez, R., Villarreal, F.G.: Structural change and productivity growth in Mexico, 1990–2014. Struct. Change Econ. Dyn. **41**, 53–63 (2017). http://doi.org/10.1016/j.strueco.2017.02.002
17. Pashev, K., Casini, P., Kay, N., Pantea, S.: EU Structural Change (2015). http://doi.org/10.2873/64651
18. Priede, J.: Quality competitivenes of Latvia's wood industry. Eur. Integr. Stud. **1**(5), 229–236 (2011). http://dx.doi.org/10.5755/j01.eis.0.5.1101
19. Priede, J.: Quality competitiveness of Latvia's food industry in the fish products group. J. Econ. Bus. Manage. **1**(2), 192–196 (2013). https://doi.org/10.7763/joebm.2013.v1.41
20. Priede, J., Pereira, E.T.: European Union's competitiveness and export performance in context of EU – Russia political and economic sanctions. Procedia Soc. Behav. Sci. **207**, 680–689 (2015). http://doi.org/10.1016/j.sbspro.2015.10.138
21. Priede, J., Skapars, R.: Quality competitiveness of Latvia's metal industry in the iron and steel product groups. Econ. Manage. **17**(1), 202–208 (2012). http://doi.org/10.5755/j01.em.17.1.2268

Adjustment of Banking Activity According to Basel III Requirements: Experience and Problems of Eastern Europe Countries

Natalia Konovalova[1] and Nina Trubnikova[2(✉)]

[1] RISEBA University, Meža street 3, Riga, Latvia
natalija.konovalova@riseba.lv
[2] RUDN University, Miklukho-Maklaya street 6, Moscow, Russian Federation
ninavadimovna@mail.ru

Abstract. Adjustment of banking activity is one of key problems in our times and is acquiring a particular importance both for banks themselves and their shareholders and also for customers and depositors. Therefore in order to ensure reliable operation of commercial banks and prevent their vulnerability to economic insecurity, the supervisory bodies are continuously improving the methods of and approaches towards the management of bank risks accenting a paramount importance of capital adequacy and liquidity. The objective of the research is to identify the impact of Basel III regulation methods on the banking system stability and economic growth. The main results of the study are the identification of positive and negative aspects of banking activity regulation according to Basel III requirements and the obtaining of evidences from Eastern Europe countries about consequences of nonmonetary regulation of banking activity.

Keywords: Capital · Liquidity · Nonmonetary regulation
Growth rates of GDP

1 Introduction

Transition towards regulation of banking activity based on Basel III requirements was caused by consequences of global financial and economic crisis 2007–2009 that endangered the financial system stability in many countries. The banks were forced to form a large volume of accumulations to cover bad debts and could not deal with absorption of losses. It means that the system of banking regulation and supervision existing at that time did not fully reflect the banking sector risks during the periods of economic and financial shocks. Banking sector in most Eastern Europe countries within the period 2009 to 2012 was loss-making. Thus, maximum losses of banking sector in Romania were fixed in 2011 and amounted to 800 mln RON [1]. Year 2011 was the most loss-making also for Hungary where banking sector had suffered losses nearly 300 mln HUF [2]. Banking sector in Baltic countries – Lithuania, Latvia and Estonia – had maximum losses in 2009. Estonian commercial banks in 2009 had suffered losses 600 mln EUR [3, 4] while losses in banking sector of Lithuania and Latvia within the same period had

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_59

exceeded 1 bln EUR [5–8]. Although the banking sector of such Eastern Europe countries as Poland, Czech Republic, Slovakia during the crisis and postcrisis periods was in general loss-free, many banks of these countries also incurred losses and faced problems with risk management. Similar situation was observed also in other European countries. It all demanded to review the bank regulation, which was reflected in Basel III requirements. With the objective to increase the banking sector stability, reduce the systemic risk and prevent the systemic crises in future, Basel III toughens the requirements for the capital adequacy and liquidity as well as implements the financial leverage ratio.

The goal of the research is to identify the influence of Basel III requirements on the banking system stability and economic growth in Eastern Europe countries.

Research tasks are:

1. To study the Basel III requirements related with bank's capital and bank's liquidity management;
2. To do forecast of changes of GDP growth rates;
3. To show the dependence between the rigid Basel III requirements towards the capital and liquidity and GDP growth rates in Eastern Europe countries;
4. To reveal the positive and negative consequences after Basel III requirements introduction;
5. To develop suggestions for combination of monetary and nonmonetary methods of regulation towards the banks' stability and economic development.

2 Theoretical Background

The Basel III Accord introduces new requirements to bank capital adequacy and liquidity. New requirements of the Basel Committee will be fully implemented in banking system in 2019. Under the agreements the minimum requirement for common equity as highest form of loss absorbing capital will be raised to 4.5% after application of stricter adjustments. Implementation of new minimum requirements began in January, 2013. Banks should provide new minimum requirements in relation to risk-weighted assets. Minimum common equity and Tier 1 requirements are phased in between 1 January 2013 and 1 January 2019. On 1 January 2013, minimum common equity requirement grew to 3.5% and Tier 1 capital requirement grew from 4% to 4.5%. On 1 January 2014, banks had 4% minimum common equity requirement and Tier 1–5.5%. On 1 January 2015, banks – 4.5% common equity and 6% Tier 1 requirements. The indicator of capital conservation buffer was introduced in banks' practice on 1 January 2016. The capital conservation buffer indicator is increasing from 0.625% in 2016 to 2.5% in 2019. But total capital requirement remains at the existing level of 8%. The difference between the total capital requirement of 8% and the Tier 1 requirement can be met with Tier 2 and higher forms of capital [9].

Acknowledging the necessity for an increasing level of bank's liquidity risk management and control, the Basel Committee on Banking Supervision (BCBS) developed a new version of Basel III. It provides for the introduction of uniform requirements for the maintenance of a sufficient amount of liquid resources reserve in order to prevent the in the future periods of crisis the high level of insufficiency financial resources. In this case, for commercial banks are offered two new ratios, which regulate the condition of liquid assets: LCR – Liquidity Coverage Ratio and NSFR – Net Stable Funding Ratio [10].

The stability of a bank depends on a bank's capital, its quality and size. A bank's capital is a mandatory and integral part of its financial resources, and its development in the form of core capital is a required step even before establishing a commercial bank [11]. Practically every stage of a bank's business is directly or indirectly linked to the capital at the bank's disposal and its value. A bank's capital serves as one of determinants in the evaluation process of its stability. In case of sudden capital adequacy problems a bank may lose its competitiveness [12]. The main function of a commercial bank's capital is generation of bank's income and profit respectively, and provide for a possibility to cover unexpected operating losses of a commercial bank [13]. American scientists Schooner and Taylor in [14] offer an identical definition, but in addition to that they stress the possibility to use capital of a commercial bank to cover possible losses caused by credit risk.

Liquidity and liquidity risk management are the key factors for the safety of business operations in any commercial banks [15]. Recently, many banks are facing the problem of liquidity strain when severe competition about how to attract deposits forces the banks to find other sponsors [16]. Unreasonable liquidity is the first sign of financial instability [17]. For some financial companies, the problem is not just one liquidity, but that there is also a threat to their solvency [18]. Liquidity risk is a term widely used now in the popular press, but the truth is that few practitioners or academics seem to understand this risk well. Perhaps not surprisingly, because until just a few years ago, there was very little work being done to analyse this risk factor [19]. Together with the development of finance market, opportunities and risks in liquidity management of commercial banks will also meet a correlative increase. This shows the importance of planning the liquidity needs by the methods with high stability and low cost in order to sponsor for business operations of commercial banks in the global growing competition [20]. Liquidity risk is difficult to measure and depends on so many factors that a capital requirement is unsuitable to prevent it [21].

Basel III regulation is meant to improve banks' capital solvency, liquidity quality and risk management. It overcame the limits of Basel II and provided a more accurate capital definition with new leverage and liquidity ratios and two capital buffers. Under Basel III, banks are facing severe regulation challenges. In terms of capital adequacy, the capital definition is stricter while the risk weighting is higher for counterparty credit risk. Consequently, banks should reduce their risk exposure and increase high-quality capital. This is difficult since reducing risk exposure would have a negative impact on profitability and reduce investors' appetite for banks. In terms of liquidity, banks are encouraged to invest in liquid assets and increase stable funding including customer deposits. However, a big issue during the sovereign crisis is the low appetite for long term debt in banks, and thus high related funding costs. Deposits are a good source of stable funding which may be chased by banks. However, the fierce competition for deposits and other stable funding will push up the funding cost, leading to a lower profitability [22].

3 Impact of Basel III Standards on Banking System Stability and Economic Growth of Eastern Europe Countries

Undoubtedly, the increasing requirements for the capital and liquidity reduce the probability of banking crises and increase the banking system soundness. Thus, according to the results of the model of the Bank for International Settlements, the increase in the capital liquidity ratio just by 1% (from 7% to 8%) reduces the probability of crisis occurrence more than by one-third [23]. Furthermore, the lower is the initial capital level, the higher is the effect of crisis probability reduction. Implementation of liquidity standards and particularly the net stable funding ratio (NSFR) helps to increase the banking system stability in the long run.

All these measures are aimed to restrain the range of credit boom in the periods of economic growth and to restrict a sharp decline in the volume of credit operations during the crisis recession. However, with this effect of non-monetary regulation there will be a slowdown in economic growth. Why? When improving the quality of the capital items and increasing the specific weight of equity capital in the bank resources as well as when increasing the share of liquid assets and reducing the volume of transformation of short-term resources, a reduction in the banking activity can take place, which in turn will lead to a slowdown in economic growth. The reason for a slowdown in GDP rates will be that while maintaining the previous level of assets the banks will be required to have a bigger capital, but when increasing their assets with the rising share of risky assets, the banks will be obliged to increase the volume and share of equity in the total capital. As a result, the cost of resources will increase, which can lead to a reduction in the efficiency of banking activity, namely, to a reduction in the rates of return on assets and on capital. Reduction in return on capital will entail a reduction in the dividend payments and in the attractiveness of bank shares for investors. As a result, inflow of financial resources into the economy will decrease and this in turn will negatively affect the growth indicators. Let us consider what impact the implementation of minimum capital requirements under Basel III had and will have in future on economic growth (Fig. 1).

During the period from the beginning of implementation of minimum capital requirements (2013) until their full introduction (2019), the increase in the ratios is accompanied by a slowdown in GDP growth rates (both actual and predicted). Forecast of GDP growth rates until 2020 is presented by experts of “Trading Economics”. It is clear from Fig. 1 that upon the increase in the ratios of minimum capital requirements a slowdown in GDP growth rates takes place in Eastern Europe, which confirms conclusions and forecasts made by experts and analysts of OECD and IMF.

Since Basel III first of all puts forward requirements for the increase in the share of stable capital elements, in the analysis of impact on economic growth it is expedient to consider the relationship between tier 1 capital adequacy and GDP growth rates (Figs. 2, 4 and 5). It is evident from presented diagrams that essential drop in GDP growth rates in Eastern Europe countries was observed twice during the last decade: in financial crisis period (2007–2009) and also in the period of supervisory monitoring (2011–2012) recommended by Basel III when the banks had to prepare for the implementation of new, more severe requirements in relation to the capital. Thus, it can be concluded that Basel III requirements are nonmonetary methods of regulation oriented

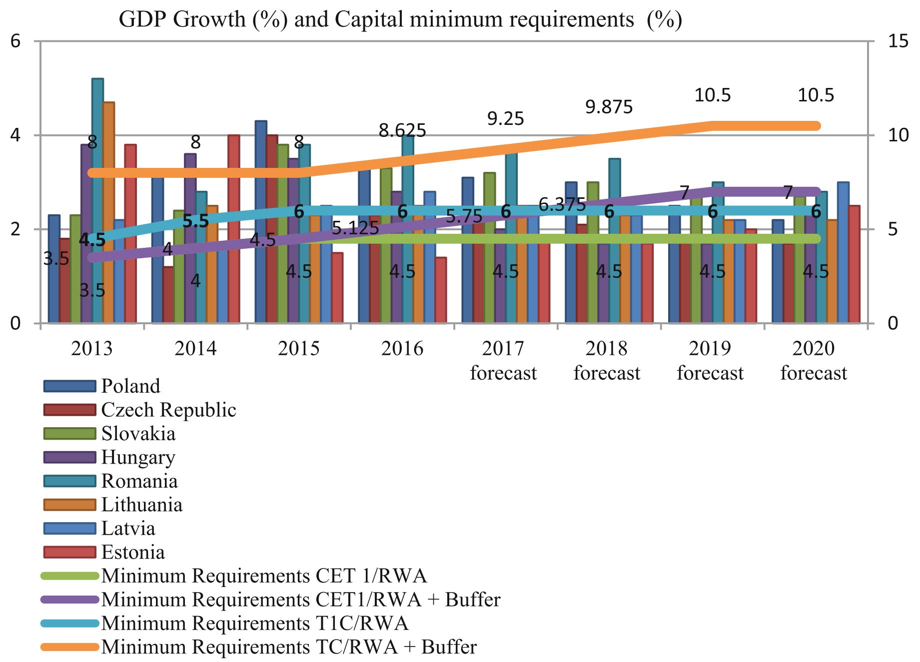


Fig. 1. Dependence between changes of the minimum requirements of capital and growth rates of GDP (developed by the authors basing on [9, 24]).

on the one hand towards the stability and sustainability of banking system and on the other hand to the constraint of economic growth rates, i.e. nonmonetary restriction. Experts of "Trading Economics" predict that GDP growth rates will slow also in future since the effect from implementation of Basel III requirements only begins. Thus according to expert evaluations, it is expected that a slowdown in GDP growth rates will be more tangible in such economically strong Eastern Europe countries as Czech Republic, Poland, Slovakia. Thus, for example, while in 2015 GDP growth rate in Poland was 4.3%, then according to forecasts of "Trading Economics" GDP growth rate in Poland by 2020 will slow to 2.2%. In Czech Republic, a slowdown to 1.7% in GDP growth rates is expected by 2020 compared to 4% growth in 2015. But in Slovakia GDP growth rate 3.8% in 2015 will slow down to 2.7% by 2020. In other Eastern Europe countries, a slowdown in GDP growth rates will be less considerable. It should be also noted that the severization of requirements for the capital will be associated not only with a slowdown in GDP growth rates but will also lead to a reduction in GDP volatility and in fluctuation range within the business cycle (Figs. 2, 4 and 5).

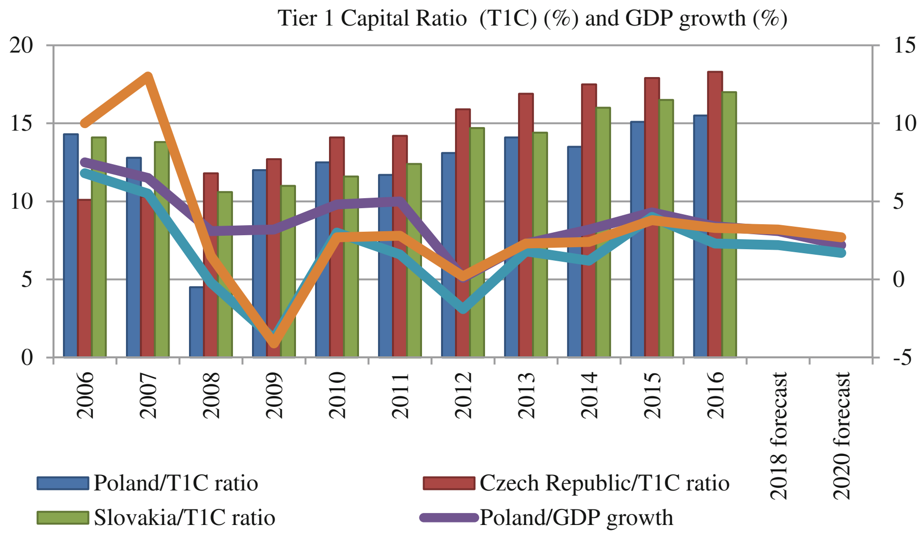


Fig. 2. Dependence between T1C ratio and growth rates of GDP in Poland, Czech Republic and Slovakia (developed by the authors basing on [25–27]).

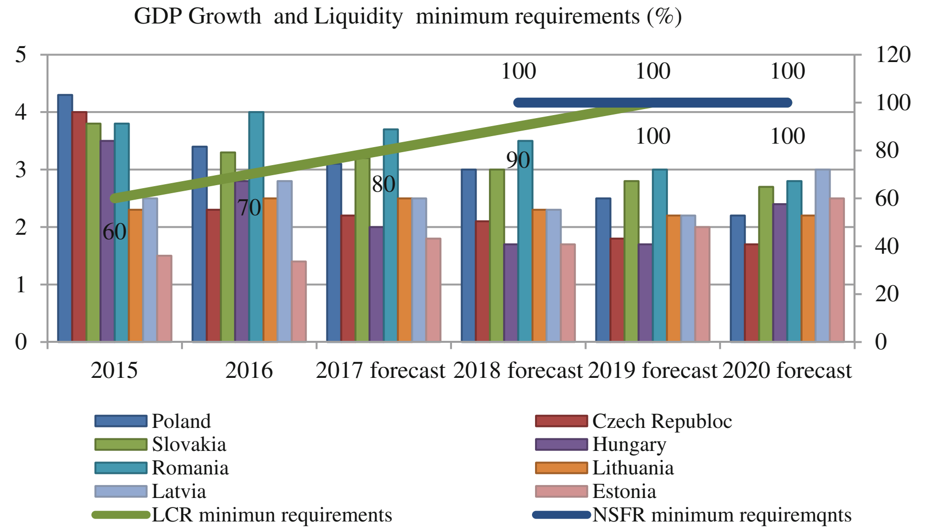


Fig. 3. Dependence between changes of the minimum requirements of liquidity and growth rates of GDP (developed by the authors basing on [9, 24]).

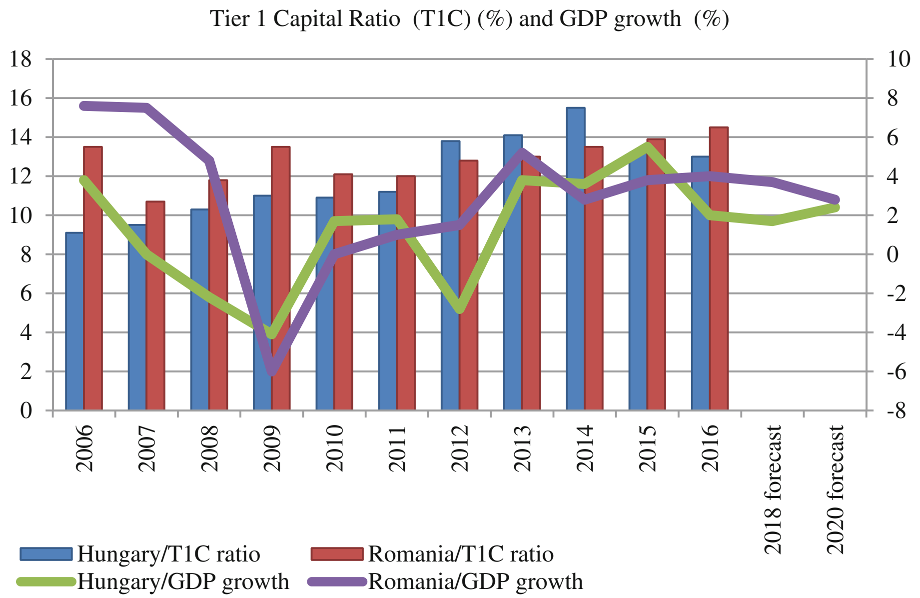


Fig. 4. Dependence between T1C ratio and growth rates of GDP in Hungary and Romania (developed by the authors basing on [1, 2]).

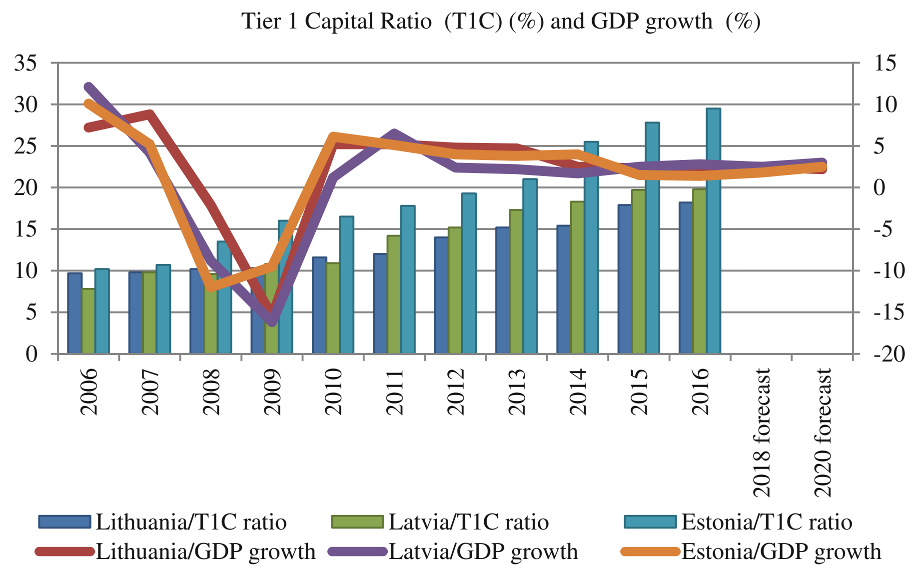


Fig. 5. Dependence between T1C ratio and growth rates of GDP in Lithuania, Latvia and Estonia (developed by the authors basing on data [4, 7, 8]).

What will the new liquidity standards bring? According to expert evaluations, the implementation of liquidity standards as well as the increasing capital requirements will also lead to a reduction in GDP growth rates. According to the results of study carried out by the Microenvironment Analysis Group and the agency “Trading Economics” [24], the banks’ performed increase in the level of liquid assets by 25% and extension of bank liabilities maturity in order to maintain the net stable funding ratio can lead to a reduction in annual GDP growth rates by 0.08% [24].

Estimates of the Bank for International Settlements confirm this conclusion: increase in the net stable funding ratio by 1% point leads to a downward deviation of GDP from the basic level by 0.08% [9].

Another consequence of implementation of additional liquidity standards can be a reduction in the volumes of bank crediting and overall deterioration of bank operation efficiency. The thing is that said measures hamper the process of transformation of assets and liabilities, which is a necessary element of dynamic bank management policy. It can entail a deterioration of ROE and ROA ratios.

Impact on economic growth by implementation of new liquidity requirements under Basel III is shown on Fig. 3. After implementation in 2015 of the Liquidity Coverage Ratio (LCR) and with its subsequent rise, a slowdown is also expected in GDP growth rates. Implementation from 2018 of the Net Stable Funding Ratio (NSFR) will also adversely affect the economic growth in general. According to forecasts of “Trading Economics”, by 2019 (target deadline for full compliance of Basel III) a slowdown in growth rates can be maximal within a 5-year period (2015–2019).

How to withstand possible negative consequences of heavy Basel III regulation? It appears that a negative impact of Basel III nonmonetary regulation on economic growth can be reduced or prevented by methods of monetary policy, which will be oriented at the money-and-credit expansion. However, it will only be possible in those Eastern Europe countries that have retained their monetary independence and their national currency (such as Poland, Czech Republic, Hungary, and Romania). In Eastern Europe countries being the members of the euro area, realization of such countermeasures will be strongly restricted or impossible at all since the legal minimum reserve ratio and ECB refinancing rate are currently on zero level while ECB deposit rate is even negative. Good opportunities for prevention of Basel III negative impact due to reduction in legal reserve ratio can be used by Romania with its effective legal reserve ratio 10% [1, 24], Poland with effective legal reserve ratio 3.5% [24, 28] as well as Czech Republic and Hungary with effective legal reserve ratio 2% [2, 24, 25]. Possibilities of downward maneuvering with the refinancing rate in these countries will be insignificant since currently the refinancing rate in Romania is 1.75% [1], in Poland – 1.5% [28], in Hungary – 0.9% [2], but in Czech Republic – only 0.05% [25]. Nevertheless, Poland, Czech Republic, Hungary and Romania will be capable to a greater degree withstand a slowdown in their economic growth by realization of monetary policy than such analyzed Eastern Europe countries as Slovakia, Lithuania, Latvia and Estonia.

4 Conclusions

Bank regulation based on Basel III requirements may have both positive and adverse aspects and consequences. As positive aspects we consider the followings:

- Growing requirements for the capital and liquidity will increase the borrowing power and solvency of banks and, therewith, the sustainability of the entire banking sector.
- Banking system and economy in general will be more resistant to financial shocks.
- Regulation based on Basel III will also contribute to reduction in systemic risk and prevention of systemic crises in future.

And negative consequences:

- Increase in the capital of banks as well as improvement of its structure and quality will lead to growing expenditures of banks, which in turn can entail growth in credit rates and reduction of banking activity.
- As a result, economic growth will slow and reduction of banking activity will have adverse impact on profitability of banking business. At the same time, it can affect the involvement of banks in high-risk transactions.

Proposals towards the banks' stability increasing and mitigation of Basel III strict requirements:

1. Use possibilities to increase the core capital by means of share issue for public offering and to be offered to the existing shareholders.
2. With the increase of profit of commercial banks use the possibility to increase the equity capital from internal sources of the commercial bank, for example, by means of capitalisation.
3. To reduce the credit risk, which has the greatest impact on capital adequacy ratios, the quality of credit portfolio management should be improved by strengthening the credit monitoring and introducing stricter evaluation requirements of borrowers' creditworthiness.
4. To achieve consistency between a bank's risks and capital required to cover them, commercial banks as they accumulate historical data must develop and approve risk evaluation methods based on internal ratings of the commercial bank.
5. It is recommended to do the regular stress-testing, undergo simulations of problematic situations, as well as verify banks' liquidity, solvency and durability against various stress situations.
6. The commercial banks should perform the short-term liquidity planning in accordance with the cash flows based on the requirements of Basel III.
7. Negative impact of Basel III nonmonetary regulation on economic growth can be reduced or prevented by methods of monetary policy, which will be oriented at the money-and-credit expansion.
8. Good opportunities for prevention of Basel III negative impact due to reduction in legal reserve ratio can be used by Romania with its effective legal reserve ratio 10%, Poland with effective legal reserve ratio 3.5% as well as Czech Republic and Hungary with effective legal reserve ratio 2%.

References

1. National Bank of Romania. http://www.bnr.ro/Home.aspx. Accessed 28 July 2017
2. Central Bank of Hungary. http://www.mnb.hu/en. Accessed 28 July 2017
3. Eesti Pank. http://www.eestipank.ee/en. Accessed 28 July 2017
4. Estonian Financial Supervision Authority. http://www.fi.ee/?lang=en. Accessed 28 July 2017
5. Bank of Lithuania. http://www.lb.lt/en_index.htm. Accessed 28 July 2017
6. Bank of Latvia. https://www.bank.lv/en/. Accessed 25 Aug 2017
7. Securities Commission of the Republic of Lithuania. http://www.vpk.lt/. Accessed 28 July 2017
8. Latvian Financial and Capital Market Commission. http://www.fktk.lv/en/. Accessed 28 July 2017
9. Basel Committee on Banking Supervision. http://www.bis.org/. Accessed 15 June 2017
10. Konovalova, N., Zarembo, J.: Imbalanced liquidity risk management: evidence from Latvian and Lithuanian commercial banks. Copernican J. Finan. Account. **4**(1), 109–130 (2015)
11. Saksonova, S.: Banku darbība. Latvian Association of Commercial Banks counseling and training center, Riga (2006)
12. Greuning, H., Brajovic Bratanovic, S.: Analyzing banking risk: a framework for assessing corporate governance and risk management. World Bank, Washington (2009)
13. Chorafas, D.N.: Economic Capital Allocation with Basel II: Cost, Benefit and Implementation Procedures. Elsevier Butterworth – Heinemann, Oxford (2004)
14. Schooner, H.M., Taylor, M.: Regulation of Global Banking: Principles and Policies. Academic Press, Chicago (2009)
15. Bertham, M.: Corporate Cash Balanced and the Bottom Line. Paradigm Shift to Dynamic Discounting, 1st Edn. Amazon (2011)
16. Rose, P.S.: Commercial Bank Management. Texas A&M University, Boston (2002)
17. Schinasi, G.: Financial-stability challenges in European emerging-market countries. World Bank **57**(73), 35–57 (2011). Policy Research working paper
18. Allen, W.A.: International Liquidity and the Financial Crisis. Cambridge University Press, Amazon, New York (2013)
19. Chacko, G., Evans, C.L., Gunawan, H., Sjoman, A.L.: The Global Economic System: How Liquidity Shocks Affect Financial Institutions and Lead to Economic Crisis. FT Press, New Jersey (2011)
20. Kochubey, T., Kowalczyk, D.: The relationship between capital, liquidity and risk in commercial banks. In: 20th Dubrovnik Economic Conference Proceedings, pp. 21–32. Elsevier, Dubrovnik (2017)
21. Ruozi, R., Ferrari, P.: Liquidity Risk Management in Banks: Economic and Regulatory Issues, pp. 1–54. Springer, Heidelberg (2012)
22. Yuting, F., Yuanyuan, X., Patrick, L.: The impact of Basel III on the European banking industry. HEC, Paris (2012)
23. Kyriaki, K., Zopounidis, C.: Management of bank performance in Greece. South-Eastern Europe J. Econ. **1**(1), 79–95 (2008)
24. Trading Economics. http://www.tradingeconomics.com/. Accessed 25 Aug 2017
25. Czech National Bank. http://www.cnb.cz/en/. Accessed 28 July 2017
26. Národná banka Slovenska. http://www.nbs.sk/en/home. Accessed 28 July 2017
27. Polish Financial Supervision Authority. http://www.knf.gov.pl/en/index.html. Accessed 28 July 2017
28. Narodowy Bank Polski. http://www.nbp.pl/. Accessed 28 July 2017

Assessment of the Competitive Environment of Passengers in Rail Transport: Case Study of Lithuania

Aldona Jarašūnienė(✉), Nijolė Batarlienė, and Kristina Čižiūnienė

Vilniaus Gedimino Technikos Universitetas, Transporto inžinerijos fakultetas, J. Basanavičiaus g. 28B, 03224 Vilnius, Lithuania
{aldona.jarasuniene,nijole.batarliene, kristina.ciziuniene}@vgtu.lt

Abstract. The paper presents various scientific interpretations of the concept of competition, the concept of a competitive environment, methods and models of competitive environment assessment and the competition factors that influence the activities of the company are presented as well. According to this literature assessment, the analysis of competitive environment of passenger transportation of JC "Lithuanian Railways" was carried out and the forecast of passenger transportation was provided for all modes in 2017–2021.

Keywords: Competition · Competitive environment · Factors
Rail transport · Road transport · Market · Market participants

1 Introduction

Market participants who are profit-seeking and who want to determine the user's choice for their own benefit must constantly raise the quality of the product or service provided, improve the company's technical capabilities, integrate modern technologies, raise staff qualifications, search for new markets, and to keep competitive prices [1, 2]. Meanwhile, the consumer receives a quality product or service at an economically reasonable cost. Competition obliges market participants to constantly monitor the environment, analyze the economic situation and make strategic decisions.

According to the data of the Lithuanian Statistics Department, in recent years about 0.4 billion passengers have travel by all kinds of transport in the country. The majority – about 98% chose the road transport, the rest of the passengers used rail, water and air transport [9].

2 Concept of Competition

Summarizing the analysis of scientific literature, it can be argued that competition is a particular struggle, competition or attempt to get more financial or other benefits to market participants engaged in economic activity.

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_60

In the Lithuanian encyclopedia [9], competition is defined as follows: Competition in the economy is understood as a competition in which legal and natural persons engaged in business, independently acting in the market, restrict each other's ability to dominate the market and encourage the production of the goods and services necessary for consumers and the desire to increase its efficiency.

A short and clear definition is given in Economic Terms Dictionary [6]: Competition is a competition between consumers who want to buy a product or service, as well as manufacturers who are seeking to sell the product or service. According to I. Beniušienė, "in the general sense, competition is perceived as an opportunity to win a battle, which strengthens the position of an enterprise in relation to competitors" [2, 6]. "Competition is the ratio between any number of companies that sell identical goods or services at the same time to a defined range of consumers", according to Goydard's book Competition Law of the European Communities [14, 15]. R. Wish offers the following definition: "competition is a fight or competition for excellence, while in the business world it is an advantage to conquer the consumer or business on the market". The concept of competition is also included in the Law on Competition of the Republic of Lithuania (adopted in 1992 and lost power in in 1999). It states that competition is "a competition in which economic operators, independently acting in the market, restrict one another's ability to dominate the market and promote the production of the goods needed by consumers and increase their efficiency" [5].

In summary – competition is a fight and in the operation of a competition mechanism, the winner is the consumer of goods or services [4].

3 The Competitive Environment and Its Evaluation Methods

Participation in any market is based on competition. Not only producers of goods and services compete with consumers, but also consumers. As a result, the user gets better quality and also is offered a more affordable price. Even the scientific literature, that has been analyzed, does not allow us to define a competitive environment unambiguously. Usik N. also states that "the competitive environment is the institutional conditions for coordinating the activities of market participants". It follows from this definition that the term "competitive environment" refers to a state or regional entity with developed infrastructure, more or less standardized budgetary, fiscal and fiscal policy management tools, and relevant institutions [13]. This definition is different from Walter Eucken, who stated that "competitive environment is a competitive system based on legal norms and principles of economic management that are changing under the influence of monitoring the results of the realization of operational goals". By combining these two definitions into one, it can be argued that the notion of "competitive environment" is based on a geographically defined area with developed infrastructure, more or less standardized budgetary, fiscal and fiscal policy management tools, and relevant institutions. It is important to mention that a competitive environment is a temporary object that is exposed by a particular competitor to the competition.

However, it should be kept in mind that the objective of a competitive environment is to ensure the functioning of the competition mechanism [10, 12].

The theory states that it is very important for each company to know the unknown possibilities of competitors predict their further actions and actually assess the level of competitive threat. As a result, in conducting a competitive environment analysis, the following objectives and requirements are set:

- The determination of a competitive market type;
- Determination of the existing and potential competitors type and size, their quantity, competing companies;
- Assessment of the market share occupied by competitors;
- Assessment of the direction and intensity of competition;
- Determination of competitors' weaknesses and strengths, strategies and the competitiveness of their products or services;
- Analysis of competitors' behavior in the market, forecasting of further actions and their reaction to various marketing actions.

In order to monitor market competition in real time, the Federal Antimonopoly Service (Russia) has developed a methodology for analyzing and evaluating competitive environment [12]. Based on this methodology, the analysis and evaluation of competitive environment consists of the following steps:

- Determining the time interval for market research;
- Determining the range of goods or services and the scale of the geographic market;
- Determining the number of companies operating in the market of goods or services;
- Determining the size of the entire market and the market shares of the companies;
- Determining the density of goods or services offered on the market;
- Identification of potential market entry barriers;
- Assessment of a competitive market environment;
- Preparation of an analytical report.

For the implementation of the above steps, information from various statistical organizations (for example, Lithuania Statistics Department, etc.) can be used.

There are also other ways of analyzing and assessing the competitive environment of the market, which are described in the scientific literature. The most popular are the SWOT analysis and Porter's method [7].

According to Porter's methodology, the company must set its competitive position, i.e. the company's competitive forces, which are characterized by the driving force of competition. Porter argued that the position of a particular company or, more precisely, the place of economic activity determine the advantages over the competition. In the final result, the absolute leader will become the company with the greatest competitive advantage. There are two basic types of competitive advantage: lower operating costs and product or service differentiation.

Lower costs reflect the company's ability to develop, manufacture and sell goods or services at a lower cost than a direct competitor. Differentiation is the ability to provide the consumer with a product or service, improving its quality, adding additional features or capabilities, or providing free after-sales service. In other words, differentiation is the value added of a service or product without raising its monetary price.

The main focus of Porter's analysis of competitive environment is the ability of the company to exploit its competitive advantage in a market that depends not only on direct competition but also on other factors of competition. These factors are: potential competitors; substitute goods; clients; suppliers.

The objective of Porter's competitive analysis is to identify the benefits and threats that the company may face. The five competitive forces determine the profit of the economic activity, under the influence of prices that can be set by the enterprise itself; company costs; the amount of investment required to compete in a particular economic activity.

The essence of the SWOT analysis method is to highlight the firm's strengths and weaknesses, as well as analysis of opportunities and threats and their interconnections. The purpose of the SWOT analysis is to identify and select the key factors that will be relevant for company strategy development. This analysis analyzes six key areas: a product or service; processes; buyers; distribution; finance; administration.

After analyzing these trends and identifying significant factors, the strategy of the company's activity is being developed.

Summarizing the whole theory analysis it is possible to formulate the concept of a company's competitive environment. Thus, the competitive environment is the whole of the conditions under which the operators who offer or acquire a particular good or service, the raw materials required for its production or their relations during a certain period function. Firms often use three main methods for evaluating a competitive environment: Federal Anti-trust Service, SWOT analysis, and Porte's competitive force model.

4 Competitive Environment in the Transport Sector

Transport activity has always been very important in the economy of the country. All companies, irrespective of the mode of transport, provide a practically identical service – transportation of passengers [3, 7]. Therefore, the quality of the services provided, the additional conditions or discounts granted by the carriers play an exclusive role in this business. Also, the cost of the transportation service, the travel time and the infrastructure are also very important.

According to the data published by the Department of Statistics, in 2016, all the main indicators of the Lithuanian transport and logistics sector arose [8]. Compared to 2015, gross value added (GDP) was generated in this sector (4.3%), corporate income (7.1%), number of employees (5.2%), exports of transport services (5.1%). These positive changes indicate a high degree of competitiveness of this industry. The country's progress in the logistics sector was demonstrated by the World Bank in 2016. Logistics Performance Index (LPI). In it, Lithuania has risen from the 46th place to the 29th, bypassing Estonia, Latvia and Poland [9].

According to the data of the Lithuanian Statistics Department, in 2015, and 2016 Transport, storage and communications (according to the Classification of Economic Activities) accounted for 13.1% of the gross domestic product (GDP) [8]. The contribution of transport and storage to GDP amounted to 11.4% and 11.6% (see Table 1).

Analyzing the passengers (including freight), Lithuania distinguishes five modes: rail, road, maritime, inland waterways and air transport.

Table 1. The main economic indicators of the country's transport sector [8].

	2013	2014	2015	2016	Change 2015/2016
Transport, storage and communication (bln. EUR)	4 453	4 547	4 394	4 517	2.8%
Contribution of transport, storage links to GDP (%)	14.5	13.9	13.1	13.1	
Transportation and storage (bln. EUR)	3 885	3 993	3 832	3 997	4.3%
Transport and storage contribution to GDP (%)	12.3	12.1	11.4	11.6	

Source: Official Statistics Portal of Lithuania.

Observing the period from Lithuania's accession to the European Union until 2017, it is possible to distinguish the following stages of economic development: economic growth, decline, stabilization and rise. With regard to the state of Lithuania's economy, the passenger transport market responded to changes in different ways. By trying to adapt to the current situation, acting under the laws of economics and competitiveness, developed strategies for ongoing activities, tried to maintain the existing customer segment and attract new ones. During the 2008–2010 economic downturn due to a shrinking market, competition has reached the highest level. A large part of the companies providing passenger services did not keep up and go bankrupt, part of their operations were loss making, hoping for better times [7].

The total number of carried passengers is very different compared for the different modes of transport. Number of passengers traveling by rail in 2016 increased by about 5% compared to 2015. The number of passengers traveling by air also increased (Table 2).

Table 2. Passenger transportation by all types of transport, thsd. [8].

	2012	2013	2014	2015	2016
All types of transport	*394 998.1*	*411 097.8*	*426 433.8*	*401 045.4*	*386 241.8*
Rail transport	4 802.1	4 844.1	4 577.1	4 226.5	4 427.8
Road transport	387 533.2	403 580.1	418 973.9	393 819.3	378 880.3
Maritime transport	322.6	327.9	336.5	349.2	366.3
Inland waterway transport	1764.2	1901.1	2028.9	1993.3	1850.9
Air transport	576.0	444.7	517.4	657.1	716.5

According to the official Lithuanian statistic portal, 386.2 million passengers were transported by all types of transport in 2016. This is 3.7% less than in 2015. The majority – 98.1% passengers were transported by road (of which: 22.6% were transported by trolleybuses, 75.5% were transported by bus), by rail transport – 1.15%, by inland waterway transport – 0.5%, by maritime transport – 0.09%, by air transport – 0.19%. International bus transport accounted for 36%, total international passenger transport, rail transport – 28.1%. Thus, the dominance of road transport over other modes of transport is indisputable.

The main economic indicators of the transport sector have increased in 2016: the transport and logistics sector generates GDP by 4.3%, revenue is 7.1%, and the number of enterprises is 1.4%, the number of employed is 5.2%, which demonstrates the high competitiveness of this sector of the economy.

In Lithuania, passengers who travel by rail are only transported by the company – JC "Lithuanian Railways". Increased passenger transportation by rail was the result of completed infrastructure repairs, additional train journeys and improved service quality. In 2016 completed repair works of Vokė bridge, during which 204 journeys were canceled and 6835 routes reduced. As a result, the passengers' flows on the routes Vilnius–Kaunas, Vilnius–Varėna–Marcinkonys, Vilnius–Trakai returned. Upon completion of Rail Baltica repairs, the route Kaunas–Kazlų Rūda–Marijampolė was restored. After analyzing the market demand, additional trains began to run on Radviliškis–Šiauliai–Mažeikiai, Vilnius–Ignalina–Turmantas, Vilnius–Šiauliai routes.

5 Competitive Environment for Passengers Traveling by Rail

Examining the competitive environment of the passenger transport market, which consists mainly of rail and road transport companies, one can distinguish the main factors contributing to the competitiveness of these enterprises. The factors most relevant to this are: the cost of travel, the provision of services, travel time, reliability, opportunities, security, specialization of services, enabling the satisfaction of new, special customer needs [12].

With the growth of the Lithuanian economy and the emergence of companies offering new passenger services, in order to avoid losing competitiveness, Lithuanian railways must pay attention to every aspect of the theory of competitiveness. Especially important is the cost of travel and additional opportunities, as well as the discount program, according to the categories of passengers. Much attention must be paid to improving the quality of already provided services.

Regarding the competitive environment of JC "Lithuanian Railways", the main competitors are companies that carry passengers by road. It is a company providing passenger transport by bus, minibus or taxi. A very important aspect is the use of personal vehicles, because it helps to save time, gives comfort to the movement, freedom and the opportunity to go from door to door. Therefore, in a sense, private transport becomes a competitor of Lithuanian railways [14].

To win the competition and thus to preserve existing and even attract new customers, Lithuanian railways must know very well what competitive environment they surround, the objects and entities involved in it. Only then it is possible to create long-term business plans and strategies. A SWOT analysis of rail passenger transport activities was carried out, and results are presented in Table 3.

A comprehensive analysis of the competitive environment is allows us to identify not only competitive disadvantages but also the competitive disadvantages of competing companies. Identifying its own weak and strong points, Lithuanian railways could concentrate on the first ones, thus improving their situation in the market for services provided.

Table 3. SWOT analysis of JC "Lithuanian Railways" passenger transport activities*

Strengths	Weaknesses
High level of security of passenger service	Insufficient train speed
Strict compliance with the approved route schedule	There is no possibility to smoke
The ability for passengers to relax and rest	Outdated park of wagon
Food service	During the trip there is a smell of fuel and stove burning
Train service staff help	The train does not go from door to door
In train is a lot of space, you can walk, lie	During the trip, noise and wagon vibrations
The ticket price is based on various discount programs	It is not always a smooth-functioning air conditioning system
Opportunities	Threats
To improve railway station infrastructure	Unfavorable political decisions
Update the wagon park	Sudden changes in economic
Take advantage of the geographical situation in Lithuania	Lack of qualified staff
Coordinate train and bus timetables	Financial problems
Increase the number of possible routes	Cooperation with neighboring countries

*Source: created by authors.

It is also important to analyze any other mode of transport for passengers, and for example, to make passenger traffic forecasts. Passenger transportation forecasts were done as a linear extrapolation of the time series, i.e. prediction of passenger volumes using the least squares method, taking actual data from Table 2.

This prediction was made according to individual modes of transport for 2017–2021. Table 4 shows passenger transport forecasts for 2017–2021.

Table 4. Passenger transportation forecasts for all modes of transport in 2017–2021.

	2017	2018	2019	2020	2021
All types of transport	395 693.88	392 937.38	390 180.88	387 424.38	384 667.88
Rail transport	4 165.66	3 996.38	3 912.002	3 827.55	3 629.625
Road transport	388 437.38	385 730.72	383 024.06	380 317.4	377 610.74
Maritime transport	373.11	386.668	399.6294	411.349	422.397
Inland waterway transport	1 987.36	2 013.92	2 040.48	2 067.04	2 093.6
Air transport	730.36	779.7	829.04	878.38	927.72

Source: created by authors

According to the predicted results of Regression method it can be argued that passenger transportation tends will decrease during the forecast period. After analysing the individual modes of transport, it is seen that passenger transportation is projected to increase by maritime, inland and air transport. Rail transport during the analysed period (2017–2021) has a tendency to decrease (see Fig. 1).

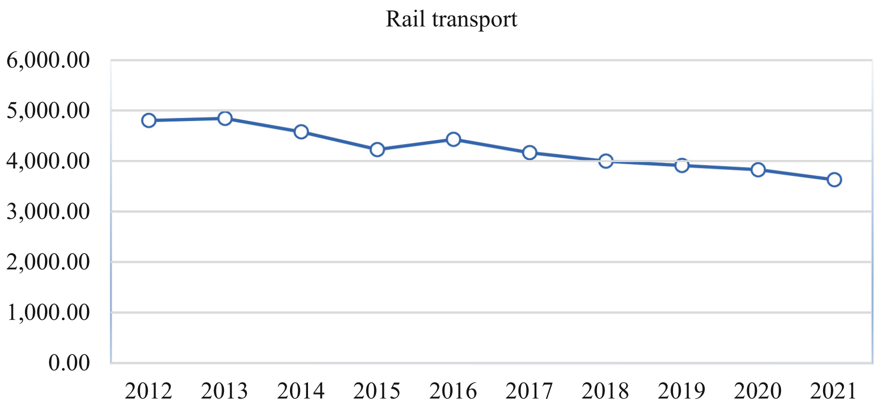


Fig. 1. Rail passenger transport actual data and forecasts 2017–2021 m. Source: created by authors.

According to a survey carried out by the Lithuanian Carrier Association LINAVA, the transport sector is developing its competitiveness through its fleet of vehicles, technology and employee experience. After accepting the decision to start the modernization of existing rolling stock, the Lithuanian railways in 2014 bought 15 new TEM TMH locomotives. Every year, employees of the company organize qualification trainings, examinations of already available professional knowledge and safety requirements. Introducing electronic ticketing and return systems that allow future travelers to save valuable time.

One of the main goals of JC “Lithuanian Railways” is Rail Baltic. He should connect Helsinki, Tallinn, Riga, Vilnius and Warsaw, extending it to Berlin. The creation and modernization of the modern north-south axis connecting the Baltic States with Poland will be able to meet the growing needs of trade and services in the EU [14].

The railway line from the Belarusian border to Vilnius, from Vilnius to Kaunas, and the stretch from Klaipeda to Kaliningrad should be electrified and arranged in accordance with the highest technical parameters within 10 years. It is hoped that a rail link from Rail Baltic, which meets the European requirements, will already be constructed from Kaunas, where the planned speed is 210 and more km/h.

6 Conclusions

1. The analysis of scientific literature has shown that competition can be defined as a certain struggle for the benefit of oneself, an attempt to gain an advantage over a direct opponent.
2. After analyzing the methods and models of the company’s competitive environment assessment presented in the scientific literature, SWOT analysis and the model of competitive forces of M. Porter can be distinguished.

3. The statistical analysis carried out has shown that in 2016, the main transport sector's economic indicators have increased: the transport and logistics sector generated 4.3% of GDP, revenue – 7.1%, the number of enterprises – 1.4%, the number of employees – 5.2%, which indicates a significant increase in this the competitiveness of the economy.
4. After analyzing SWOT of the Competition Environment of JC "Lithuanian Railways", the strengths of the company were named: High level of security of passenger services, strict compliance with the approved route schedule, various discount programs that determine the ticket price, etc.
5. The forecasts for 2017–2021 have shown that rail passenger transport tends to decrease, while water and air transport – increase.

References

1. Baublys, A.: Transporto ekonomika. Technika, Vilnius (2003)
2. Beniušienė, I., Svirskienė, G.: Konkurencingumas: teorinis aspektas. "Ekonomika ir vadyba: aktualijos ir perspektyvos" mokslinės konferencijos medžiaga, Šiauliai (2008)
3. Butkevicius, J.: Keleivių vežimo Lietuvos geležinkeliais tyrimai ir plėtra. Transportas. Technika, Vilnius (2011)
4. Klimašauskienė, D.: Konkurencijos samprata ekonomikos teorijoje (2007)
5. Lietuvos Respublikos konkurencijos įstatymas. Valstybės žinios **29**(1), 841 (1992)
6. Maksvytienė, I.: Ekonominis konkurencingumas: metodologinis aspektas. Organizacijų vadyba: sisteminiai tyrimai (2002)
7. Porter, M.: Competitive Strategy: Techniques for Analyzing Industries and Competitors, New York (1980)
8. Official Statistics Portal of Lithuania. Oficialus statistikos portalas. Transporto rodikliai, http://osp.stat.gov
9. Visuotinė lietuvių enciklopedija t. 10, Mokslo ir enciklopedijų leidybos institutas, Vilnius (2005)
10. Whish, R.: Competition Law. Butterworths (2003)
11. Ovodenko, A.A.: The Important Economics Problems in Modern Russia. Scientific Works Collection (2007) (in Russian)
12. Voronkova, O.V., Zavražina, K.B., Tolstiakov, R.R.: Marketing – Science Tool (2009) (in Russian)
13. Usik, N.I.: Competitive policy and the formation of a competitive environment. J. Theor. Econ. **4** (2012) (in Russian)
14. Gamova, V.: Analysis of competitive environment of rail transport. Master thesis (2016)
15. Goyder, D.G.: EC Competition Law, Oxford (1988)

Education and Training in Engineering

Career Management in a Technical University as an Essential Factor Influencing Its Competitiveness

Yulia Stukalina(✉)

Transport and Telecommunication Institute, 1 Lomonosova Street, Riga 1019, Latvia
Stukalina.J@tsi.lv

Abstract. Career management is often regarded as a peripheral activity in a university compared to education and research, which are traditionally considered to be central activities of a higher education institution. As a result, career guidance services are not properly coordinated with other services provided by a university. However, contemporary higher education trends (a bigger focus on research and innovation, growing popularity of lifelong learning, intensified university – business collaboration, increased contribution to the society and social responsibility, etc.) make education managers create competitive strategies aimed at generating sustainable competitive advantage. It is also determined by the fact that nowadays, students expect to have a more rewarding return on their investment in higher education in terms of both academic quality and employability, as the transition from higher education to the international labor market may sometimes become a problem. The aim of the paper is to discuss some basic issues related to career management in a contemporary technical university in the context of achieving competitive advantage in the international education market. The research question can be formulated like this: "Should career management be viewed as a special focus area of the comprehensive organizational strategy, career guidance and counselling being an essential factor that influences competitiveness of a modern university"? The paper is based on the analysis of (1) theoretical literature on strategic management; (2) official EU and OECD documents on the main issues of higher education in the agenda of students' professional career management; (3) career management activities performed in some leading technical universities in different parts of the globe.

Keywords: Technical university · Competitive advantage
Career guidance and counselling services

1 Introduction

According to Smaliukiene et al. [1], social and economic changes have their impact on the career awareness and its management, global economy and labor migration creating changes on the macro-level. The Lisbon Strategy focused, among other things, on employment and labor market reform, much attention being paid to the concept of "employability" and the activation of labor market policies [2]. The EU idea of being the leading knowledge economy (the Lisbon agenda) depends much on many factors

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_61

including (1) how EU citizens are prepared for entering the modern labor market; (2) what opportunities they have for developing necessary skills; (3) how they can be provided with career guidance and counselling needed for managing their careers in the context of greater social cohesion [3]. That is why the present EU policy and strategy framework for lifelong learning (which embraces general education, vocational education and training, higher education, adult learning) and employment is aimed at setting "holistic and coherent lifelong guidance systems" for supporting career guidance needs of various target groups [4]. Because of demographic ageing, the sustainability of Europe's economic and social model is now under great pressure, so the phenomenon of the second career labor market also looms large in the agenda of lifelong learning and employment [5].

Career guidance and counselling is vital for making right education and career choices, being fundamental strategic components for executing lifelong learning and employment strategies both at regional and national levels [6]. Significant economic and social changes have made education re-assess its role; now, it is not possible to consider education in isolation from other main public policies, so closer links with the labor markets are very important for educational institutions [7].

The need for career guidance and counselling in modern higher education increases owing to a few noteworthy factors including the following [8]: (a) universities operate in a globalized and extremely competitive international environment, which calls for creating novel market-oriented strategies; (b) more efficient career management would encourage graduates' participation in lifelong learning.

Career management is often regarded as a peripheral activity in a university compared to education and research, which are traditionally considered to be central activities of a higher education institution. As a result, career guidance services are not properly coordinated with other services provided by a university. However, contemporary higher education trends make education managers create competitive strategies aimed at generating sustainable competitive advantage in the international education market. It is vital for the development of a modern university, since global competitive forces are now reformatting the higher education industry [9].

In this context, career management should be viewed as a special focus area of the university's comprehensive strategy. It is also determined by the fact that nowadays, students expect to have a more rewarding return on their investment in higher education in terms of both academic quality and employability, as the transition from higher education to the international labor market may sometimes become a problem. Consequently, modern universities must be more flexible for meeting the expectations of new populations of learners [10], adapting their policies and tools to the requirements of knowledge-based economies [11]. Collaboration between universities and industry should be strengthened at national and regional levels [12].

As contemporary technical universities play an essential role in supporting knowledge creation and knowledge transfer in the knowledge-based economy, they can contribute to the economic development of society by providing numerous services to business, which include applied research, professional training, consultancy, career guidance and counselling [13].

The aim of the paper is to discuss some basic issues related to career management in a contemporary technical university in the context of achieving competitive advantage in the international education market. In the paper, it is argued that career guidance and counselling can be regarded as an essential factor influencing its competitiveness. The research question can be formulated like this: "Should career management be viewed as a special focus area of the comprehensive organizational strategy, career guidance and counselling being an essential factor that influences competitiveness of a modern university"?

The paper is based on the analysis of (1) theoretical literature on strategic management; (2) official EU and OECD documents on the main issues of higher education in the agenda of students' professional career development; (3) career management activities performed in some leading technical universities in different parts of the globe. The main findings and conclusions are presented below.

2 Career Management in the Frame of the Overall Strategy of a Technical University

2.1 Identifying Areas of Competitive Advantage in a Modern University

For competing effectively, organizations try to gain sustainable competitive advantage implementing value-creating strategies [14]. Today, universities have to adopt some strategic management practices initially created for the private sector, provided that they are adapted properly for the educational sector [15–17]. As a company can gain competitive advantage if it is able to create value for its buyers [18], so a university can accomplish competitive advantage if it is able to create value for its stakeholders, including its students.

Thus, in order to stay competitive in the international education market managers who work in higher education search for the new ways to gain competitive advantage, focusing on the main determinants of competitiveness in this field. Today, education managers have to re-examine the internal structures and processes of organizations, creating new managerial and decision-making systems [19].

Contemporary universities can establish multiple strategic goals [20]. For creating a competitive strategy of a modern university it is vital to understand and identify the areas of competitive advantage (functional domains). In the frame of broad strategic goals, managers may set more detailed (functional area-specific) objectives, and develop different strategies in the context of an overall strategy. In a university, these strategies are built around the following domains (functions) – education, research, staff, university services and facilities [21].

The above functional domains are closely related to the basic activities that are traditionally performed by higher education institutions. McBurnie and Ziguras [22] characterize them as teaching, research and community services. These functional areas embrace different organizational entities (for example, academic departments, administrative support units, career management centers/offices, etc.), each carrying out specific activities or processes. Every functional area is associated with a specific task to be performed in complex university's settings (Table 1). Due to the significant role that career guidance and counselling play in a modern higher education institution, we regard career management as a separate functional domain making its contribution to gaining competitive advantage.

Table 1. Main tasks performed by different functional domains in the frame of the overall university's strategy aimed at achieving competitive advantage

Functional domain	Main task
1. Education	1. To support sustainable academic activities in the agenda of lifelong education
2. Research	2. To support sustainable research process
3. Staff	3. To ensure sustainable personal and professional development
4. Services and facilities	4. To provide sustainable institutional performance
5. Career management	5. To provide sustainable career guidance and counselling services

Thus, strategic planning embraces multiple levels and functions of a higher education institution; different strategies included in the university's all-inclusive strategic plan. They are all aimed at enhancing student experience and increase their employability in the international labor market.

In the frame of the strategic planning model developed by Thompson and Strickland [23], the following elements can be included in the strategic plan at the functional level: setting functional area-specific strategic objectives and developing functional area-specific initiatives for supporting their accomplishment (Table 2).

Table 2. Area-specific strategic objectives and initiatives related to different functional domains of a university (adapted from Stukalina [13])

Functional domain	Area-specific objectives	Area-specific initiatives
1. Education	1. Enabling the acquisition of transferable skills necessary for the knowledge-based economy	1. Developing educational programs responsive to the changes in the international labor market
2. Research	2. Cultivating the reputation of a research-based university	2. Creating innovative products in partnership with domestic and international companies
3. Staff	3. Increasing discipline-specific proficiency of the teaching staff	3. Establishing consistent recruitment procedures aimed at improving quality of the teaching staff
4. Services and facilities	4. Ensuring ICT-based education, research and career guidance	4. Investing in cost-effective ICT-based infrastructure
5. Career management	5. Integrating career services throughout the student lifecycle	5. Establishing career management centers in the agenda of lifelong learning

2.2 Career Management Activities in a Technical University: An International Perspective

In order to gain an international perspective on the career management in higher technical education, the author has summarized various career management activities executed in some leading technical universities in different parts of the globe.

Table 3. Career management activities in modern technical universities: an international perspective

University	Specialization	Main focus of activities
1. University of Strathclyde (UK) [24]	1. Architecture, IT, Electronic and Electrical Engineering, Mechanical and Aerospace Engineering, etc	1. Offering strategies to support self-awareness and personal skills development
2. IE University (Spain) [25]	2. Architecture, Information Systems Management, Law, Business Analytics and Big data, International Business Administration. Political Studies, Corporate Communication, etc.	2. Launching graduates' careers across industries in the international labor market; directing them through the decision-making process
3. Texas Tech University (USA) [26]	3. IT, Energy Commerce, Architecture, Agricultural and Applied Economics, Data Science, Information Systems and Quantitative Sciences, Accounting, etc.	3. Fostering student success through enabling students for their future careers: helping them prepare, connect and get employed
4. Colorado Technical University (USA) [27]	4. Engineering and Computer Science, Electronics Technology, IT, Business Administration, Software Engineering, Cybercrime and Security, Computer Systems Security, Information Systems Management etc.	4. Assisting graduates more effectively compete for career opportunities by providing access to reliable resources; offering students high quality career coaching
5. Vilnius Gediminas Technical University (EU) [28]	5. Information Systems, IT, Software Engineering, Informatics, Architecture, Transport Engineering Economics and Logistics, Engineering Economics and Management, etc.	5. Assisting graduates by creating individual career plans and developing strategies leading to a successful career
6. University of Oulu (EU) [29]	6. Engineering, Architecture, Wireless Communications Engineering, Natural Sciences, Environmental Engineering, Computer Science, etc.	6. Establishing channels for students to do internships through stakeholder and employer co-operation
7. University of Technology Sydney (Australia) [30]	7. Engineering, Architecture and Design, Environmental Engineering Management, Business and Technology, Electronics, Business Administration, etc.	7. Providing students with basic employability skills and knowledge vital for being successful in the international labor market

As one can see from Table 3, career management activities in different technical universities are intended to provide their students and graduates with a wide range of services supporting the acquisition of modern skills necessary for successfully launching and developing their professional career in the local or international labor market. These skills include as career management skills, as transferable skills specified in the Recommendation of the European Parliament and of the Council on key competences for lifelong learning [31]: the ability to communicate in different communicative situations, intercultural understanding, basic competences in science and technology, digital competence, learning to learn, the ability to demonstrate tolerance and understanding of different viewpoints and cultures, sense of initiative and entrepreneurship, etc. This would enhance both graduates' employment prospects and personal development.

So the total student experience in a university involve not only academic learning, but also educational and development programs delivered by careers servicecentres. Such approach can change the university's overall strategy, making employability a central focus in the highly competitive modern environment.

3 Career Guidance Services in a Technical University

Thus, it is apparent from the foregoing that modern universities are committed to assisting their students and graduates in all aspects of their professional career development. They can provide them with a wide range of career guidance and counseling resources (Table 4).

Table 4. Career guidance and counselling resources available in a technical university

Career guidance resources	Description
1. Career guidance and counselling center	1. Provides free counselling, guidance and coaching for job search efforts and career development
2. E-library	2. Contains recommended literature needed for supporting individual career development
3. Professional development courses	3. Provide the acquisition of a discipline-specific professional skills
4. Handbooks and self-guided tools	4. Help plan and manage individual career development
5. Career guidance and counselling network	5. Provides individual career development assistance from confidential career counselors
6. Career events	6. Present current career trends in the international labor market
7. Alumni association	7. Helps alumni provide guidance to current students

Career management is regarded as a continuing problem-solving process, in which information is collected, and career goals and strategies are created [32]. Thus, career information is the basis for the provision of career guidance services [33], career-related information being obtained from multiple sources. Advances in ICT have considerably improved the quality and accessibility of informational resources related to career management. Modern higher education institutions can benefit from web-based career services management systems that integrate various career guidance and counselling services including those described above. Unique innovative programs, products, and services enable educators to support the success and professional growth of their students, graduates and alumni. So they will take advantage of free access to online career service platforms, through which they can explore job postings, stay informed about forthcoming programs, download helpful career resources, interact with potential employers and schedule counseling appointments, etc. In turn, employers may use such platforms to register for the various recruiting programs, post jobs and maintain an information profile available to students, graduates and alumni etc.

Let's consider the following example. Today, many educational institutions, employers and professional associations utilize the *Symplicity* Career Services Management System (CSM) to manage all aspects of the recruitment process through providing a number of online services to students, alumni and employers (Table 5).

Table 5. Symplicity Career Services Management System: main features [34]

Feature	Description
1. Student CMS	1. Used to communicate with students to guide them through the job search process including giving career advice, evaluating resumes, reviewing job applications and interviews, accessing student activity data
2. Experiential learning	2. Used for posting internship positions, managing applicants, processing evaluations, and tracking offers
3. Recruiting toolkit	3. Used to connect students and employers; it includes easy-to-use job posting system, enhanced employer profiles, on-campus interviews and resume books
4. Advanced metrics and outcomes reporting	4. Used to collect and report on first-destination and outcomes
5. Career fair and workshop manager	5. Used to organize and host career fairs and workshops with simple registration pages, email reminders and payment processing for busy employers
6. Swipe card kiosks	6. Used to offer students convenient access to job opportunities

As one can see from the table above, web-based career management systems may include the management of: job posting, externships and internship postings; internal and external communications; career guidance, counselling and mentoring; the recruitment process, recruitment events and programs; post-graduate employment opportunities; job application documents; employer registration, etc.

Therefore, the internal web-based career management system may become the key to effective career services integration in a technical university. The ability to manage information sharing would enable a more collaborative environment as inside as outside a university (facilitating interactions between employers and students/alumni).

In a higher education institution, such systems can be developed on the basis of the existing e-learning platforms, which support the educational process and stimulate social interaction.

4 Conclusions

This paper has identified and discussed some central issues related to career management in a higher education institution in the context of achieving sustainable competitive advantage in contemporary higher education settings.

The analysis of theoretical literature and official EU and OECD documents provided in the paper has allowed the author to draw the following conclusions. For making their organizations more competitive higher education managers have to seek for the new ways to gain competitive advantage, focusing on the main determinants of competitiveness in the field. In the agenda of creating competitive strategies, career management activities should be integrated into the strategic plan of a technical university, career guidance and counselling being an essential factor affecting its competitiveness in the international education market. Thus, career management ought to be viewed as a special focus area of the comprehensive organizational strategy, and career management activities should be given more emphasis in the framework of the university's overall strategy. The transition from higher education to the global labor market may sometimes become a problem, so it is important that the total student experience in a university would include not only academic learning, but also professional career development programs offered by careers service centres, as employability is given increasing emphasis in the extremely competitive modern environment.

In order to gain an international perspective on the career management in higher technical education, different career management activities performed in some leading technical universities have been summarized. This has allowed the author to conclude that modern higher education institutions are committed to supporting their students, graduates and alumni in all aspects of their professional career development by providing them with a wide range of career guidance and counselling resources and services.

As advances in ICT have significantly improved the quality and availability of informational resources related to career management, technical universities can now benefit from web-based career services management systems integrating various career guidance and counselling services. Students, graduates and alumni, as well as potential employers, can take full advantage of free access to online career service platforms offering useful tips and information.

References

1. Smaliukiene, R., Korsakiene, R., Tvaronaviciene, M.: Career management opportunities in international labour market: a theoretical perspective. Procedia Soc. Behav. Sci. **110**, 293–300 (2014)
2. The Lisbon Strategy 2000–2010: An analysis and evaluation of the methods used and results achieved: Final Report. Brussels, European Parliament (2010). http://www.europarl.europa.eu/document/activities/cont/201107/20110718ATT24270/20110718ATT24270EN.pdf. Accessed 14 April 2017

3. Career Development at Work – A Review of Career Guidance to Support People in Employment, Office for Official Publications of the European Communities, Luxembourg (2008). http://www.cedefop.europa.eu/en/publications-and-resources/publications/5183. Accessed 15 Aug 2017
4. Lifelong Guidance across Europe: Reviewing Policy Progress and Future Prospects (Working paper No. 11), Publications Office of the European Union, Luxembourg (2011). http://www.google.lv/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=0CC8QFjAC&url=http%3A%2F%2Fwww.cedefop.europa.eu%2Ffiles%2F6111_en.pdf&ei=3edZVfa4IouWsgHXiIGoDA&usg=AFQjCNFKUbK4euUYXvM7cR7A9CK33DEjdQ&bvm=bv.93564037,d.bGg. Accessed 22 June 2017
5. Second Career Labour Markets: Assessing Challenges – Advancing Policies. Bertelsmann Stiftung and European Policy Centre, Gütersloh, Germany (2013). https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/GP_Second_Career_Labour_Markets.pdf. Accessed 10 July 2017
6. Establishing and Developing National Lifelong Guidance Policy Forums: A Manual for Policy-makers and Stakeholders, Office for Official Publications of the European Communities, Luxembourg (2008). http://www.cedefop.europa.eu/EN/Files/5188_en.pdf. Accessed 23 May 2017
7. Career Guidance: New Ways Forward. In: OECD Education Policy Analysis, Chap. 2 (2003). http://www.keepeek.com/Digital-Asset-Management/oecd/education/education-policy-analysis-2003_epa-2003-en#page1. Accessed 24 May 2017
8. Roskosa, A., Stukalina, Y.: The new challenge for higher education institutions of Latvia: directing students' professional career development. In: Proceedings of the 9th International Scientific Conference on Rural Environment. Education. Personality, pp. 239–246 (2016)
9. Rust, V.D., Kim, S.: The global competition in higher education. World Stud. Educ. **13**(2), 5–20 (2012)
10. The State of Higher Education 2014: Executive Summary. Introduction (2014). http://www.oecd.org/edu/imhe/SOHE2014-Executive%20Summary.pdf. Accessed 11 May 2017
11. COM(2007) 61 Final: A Coherent Framework of Indicators and Benchmarks for Monitoring Progress towards the Lisbon Objectives in Education and Training, Communication from the Commission of 21 February 2007 (2007). http://europa.eu/scadplus/leg/en/cha/c11099.htm. Accessed 22 April 2017
12. COM(2003) 58 Final: The Role of the Universities in the Europe of Knowledge, Communication from the Commission of 5 February 2003 (2003). http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11067_en.htm. Accessed 21 April 2017
13. Stukalina, Y.: Management of a Technical University in the context of preparing students for the 21st century careers in science and technology. Procedia Eng. **178**, 249–257 (2017)
14. Barney, J.: Gaining and Maintaining Competitive Advantage, 3rd edn. Addison-Wesley Publishing Company, Reading (2007)
15. Marginson, S., Considine, M.: The Enterprise University: Power, Governance and Re-invention in Australia. Cambridge University Press, Sydney (2000)
16. Malony, J.: Australian universities today. In: Coady, T. (ed.) Why Universities Matter. Allen & Unwin, Sydney (2000)
17. Pisani, A.M., Filkins, J.W.: Strategic planning in a transitory higher education environment: building historical foundations and industrial parallels. In: Smart, J. (ed.) Handbook of Theory and Research, vol. XIII. Agathon Press, New York (2000)
18. Porter, M.: From competitive advantage to corporate strategy. Harv. Bus. Rev. **3**, 43–59 (1987)

19. Felt, U.: The new meaning of leadership in autonomous universities. In: Managing the University Community: Exploring Good Practice, pp. 11–12. European University Association, Belgium (2007). Chapter I: Change management
20. Bain, G.: Leading universities. In: Managing the University Community: Exploring Good Practice, pp. 13–15. European University Association, Belgium (2007). Chapter I: Change management
21. Stukalina, Y.: Strategic management of higher education institutions. Manag. Organ. Syst. Res. **70**, 80–90 (2014)
22. McBurnie, G., Ziguras, C.: Trends and future scenarios in programme and institution mobility across borders. In: Cross-border Higher Education: Trends and Perspectives, Higher Education to 2030, vol. 2, pp. 89–108. Globalisation. OECD Publishing (2009). http://www.mfdps.si/Files/Knjiznica/higher%20educational%202030%20OECD.pdf. Accessed 14 May 2017
23. Thompson, A.A., Strickland, A.J.: Strategic Management: Concepts and Cases. McGraw-Hill, Irwin (2003)
24. University of Strathclyde Homepage. https://www.strath.ac.uk/mba/careermanagement/. Accessed 29 June 2017
25. IE University Homepage. https://www.ie.edu/university/ieu-experience/career-management-center/. Accessed 30 June 2017
26. Texas Tech University Homepage. http://www.depts.ttu.edu/rawlsbusiness/about/cmc/. Accessed 28 June 2017
27. Colorado Technical University Homepage. http://www.coloradotech.edu/career-development. Accessed 28 June 2017
28. Vilnius Gediminas Technical University Homepage. http://www.vgtu.lt/studies/career-services/68362?lang=2. Accessed 29 June 2017
29. University of Oulu Homepage. http://www.oulu.fi/university/node/35195. Accessed 29 June 2017
30. University of Technology Sydney Homepage. http://www.uts.edu.au/current-students/opportunities/career-development/careers-service. Accessed 30 June 2017
31. Recommendation of the European Parliament and of the Council on key competences for lifelong learning (2006). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32006H0962. Accessed 25 May 2017
32. Greenhouse, J.H., Callanan, G.A., Godshalk, V.M.: Career Management, 4th edn. SAGE, USA (2009)
33. Hansen, E.: Career Guidance: A Resource Handbook for Low- and Middle-Income Countries. International Labour Organisation, Geneva (2006)
34. Symplicity Homepage. https://www.symplicity.com/higher-ed/solutions/csm. Accessed 11 July 2017

Learning Analytics and Software Engineering Competences

Boriss Misnevs[1(✉)] and Aliaksandr Puptsau[2]

[1] Transport and Telecommunication Institute, Lomonosova str., 1, Riga, Latvia
Misnevs.B@tsi.lv

[2] European Humanities University, Tauro g. 12, 01114 Vilnius, Lithuania
alexander.puptsev@ehu.lt

Abstract. This paper examines some issues and questions regarding the use of Learning Analytics (LA) in the measurement, collection, analysis, and reporting of data related to Software Engineering (SE) competences obtained by master students at a university. The paper's authors review some results of iSECRET Project with the goal of introducing Data Mining approach for better control of the teaching process based on remote competences evaluation. This shift gave the opportunity to apply 'business intelligence' techniques to educational data. The research focuses on progress and professional standards for accountability systems used in Software Engineering Master Program with an emphasis on professional competences defined by e-CF Framework. The authors pay particular attention to the information infrastructure and mathematical background needed to conduct effective Learning Analytics at SE Master Program. Accordingly, the paper attempts to develop an original mathematical model was done. The approach is based on the pilot remote measurement of the components of competences using SECEIP Portal. The developed technique of using Learning Analytics for measuring the competences at SE Master program may be used for reengineering of Legacy University Management Information Systems and University executive staff training.

Keywords: e-CF framework · 'Big Data' · Study progress
University Information System · Knowledge discovery

1 Introduction

This paper examines some issues and questions regarding the use of Learning Analytics (LA) in the measurement, collection, analysis, and reporting of data related to Software Engineering (SE) competences obtained by master students at a university [1]. The authors review some results of iSECRET Project with the goal of introducing Data Mining (DM) approach for better control of the teaching process based on remote competences evaluation [2].

The learning outcomes (LO) of Software Engineering/Software Technologies Master Program relevant to Level 7 are the following [3]:

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_62

Knowledge – a highly specialised knowledge applicable in work and study contexts, that may function as the basis for original thinking and/or research, critical rethinking of knowledge in the field and in the cross-disciplinary domains. Skills – specialized problem-solving skills that are essential for research and/or innovation, as well as for producing new knowledge and procedures, and integrating knowledge from different fields.

Competence – that are supposed to manage and transform complex and unpredictable work or study contexts that require new strategic approaches, as well as to take responsibility for contributing to professional knowledge and practice and/or assisting the strategic performance of teams.

"Competence" here means a proven ability to use knowledge, skills, as well as personal, social, and/or methodological skills in work or study contexts, professional and personal development. In the context of the European Qualifications Framework, competence is described in terms of responsibility and autonomy.

The term "Learning Analytics" has been significantly popularized through the International Conferences on Learning Analytics and Knowledge (LAK) [4], where learning analytics was described as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". However, the definition and aims of Learning Analytics still are disputable. As George Siemens puts it [5], "Learning analytics is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections for predicting and advising people's learning". Many researchers also employ EDUCAUSE definition ("The use of data and models to predict student progress and performance, and the ability to act on that information") as clearer and practically applicable.

2 Methodology

2.1 Existing Research Overview

Rebecca Ferguson [6] links the progress of using analytics in teaching and learning environment to the increasing interest in 'big data' for business intelligence and the rise of online education focused on Virtual Learning Environments, Content Management Systems, and Management Information Systems for education, that increased the amount of students' digital data (i.e. regarding the background of students or log data). This shift gave the opportunity to apply 'business intelligence' techniques to educational data.

A systematic overview of learning analytics and its key concepts is provided by Chatti et al. [7, 8] through a reference model for learning analytics based on four dimensions: data, environments, context (what?), stakeholders (who?), objectives (why?), and methods (how?).

The Key Questions of Analytics are described in the paper of Adam Cooper [9].

Questions of information and fact [9]:

1. What happened? Analytics produces reports and summarized descriptions of data, (the past).
2. What is happening now? Analytics provides alerts in near-real time, (the present).
3. Where are trends leading? Past data is extrapolated, (the future).

Questions of understanding and insight [9]:

1. How and why did something happen? Analytics builds models and explanation, (the past).
2. What is the best next action? Analytics provides one or more recommendations, (the present).
3. What is likely to happen? Analytics provides prediction, simulates the effect of alternative courses of action, or identifies an optimal course of action, (the future).

An intervention is any action that is taken with the intention of improving student outcomes.

An intervention is ideally supported by analysis with measurable outcomes.

Interventions can be passive or intrusive and can be facilitated by a person or be fully automated by a system. For example, Degree Compass (developed and implemented at Austin Peay State University in Tennessee, USA) determines courses, which are needed for a student to graduate, and ranks them according to how they fit with the sequence of courses in the student's degree program and their centrality to the university curriculum as a whole [10]. That ranking is then overlaid with a collaborative filtering model that predicts the courses in which the student is most likely to achieve the best grades.

Dashboards for faculty, advisors, and tutors support instructional staff by identifying trends and enabling early intervention [10].

Universities acknowledge the potential that large amounts of data produced by their IT systems have for addressing strategic challenges that academia is facing today. Whether analytics is used to identify the ways to reduce educational costs or to provide early interventions that could help a student to succeed in a course, the system employs a combination of embedded administrative and academic technologies.

At the same time it is essential for institutional authorities to understand LA as a part of a larger process: deploying sophisticated analytical tools and predictive models as such would hardly influence the study process. It is important to determine the position of Learning Analytics on the path of a student's success (see Fig. 1).

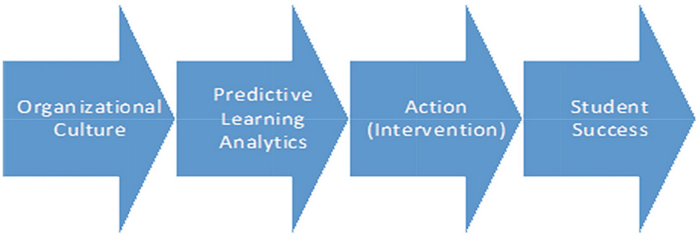


Fig. 1 Position of predictive learning analytics on the path of student success [11].

With modularized course content, students could test their understanding of certain concepts they are already familiar with and then take only those course modules that they haven't previously been exposed to. Different approaches to knowledge validation could theoretically shorten time to degree completion and cut the overall cost of education. Now universities are looking for complex classes where adaptive learning could help personalize students' educational experiences and move them forward. These universities hope to dispel the myth that the return on investment in human resources and predictive analytics is low [11].

2.2 Applicable Learning Analytics Methods

Data about learning and learners is being generated today on an unprecedented scale. The fields of learning analytics have emerged with the aim of transforming this data into new insights that can benefit students, teachers, and administrators. LA must have a robust technology to channel and analyze data, as well as effective action plans and procedures based on the results of analysis.

Learning analytics methods include [12]:

- Content analysis, particularly of resources that students create (i.e. essays).
- Discourse analytics, which aims to capture meaningful data on student interactions where a special focus (unlike social network analytics) is made on exploring the discursive properties of the language used, as opposed to the analysis of the network of interactions, or forum-post counts only, etc.
- Social learning analytics, which is aimed at exploring the role of social interaction in learning, the importance of learning networks, discourse used to making sense, etc. [12].
- Disposition analytics, which seeks to capture data regarding student's dispositions to their own learning, and their relationship towards their learning [13]. For example, "curious" learners may be more inclined to ask questions, and this data can be captured and analyzed as a part of learning analytics [14].

The processes that learning analytics aim to serve and enhance in the development of e-CF competencies are more problematic because there is no consensus on how they fit together, how they operate in learning and teaching, and how they relate to the traditional competencies and skills measured in education.

Business analytics can be used as an instrument to address the problem of economic and technical validation of factors that influence students' success, recruiting success, and institutional decision-making with the purpose of increasing the efficiency of resource allocation.

OLAP can be used for data mining or uncovering previously undiscerned relationships between data items. OLAP database does not need to be as large as a data warehouse, since not all transactional data is needed for trend analysis. Using Open Database Connectivity (ODBC) data can be imported from existing relational databases to create a multidimensional database for OLAP. In this case various reports may be generated from many different perspectives. Research [15] employs the following star scheme that consists of the central part, which is data warehouse agglomerating large amount of data, and smaller peripheral tables, which are called dimension tables (see Fig. 2).

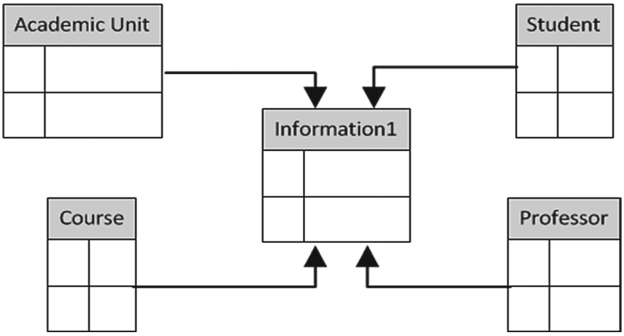


Fig. 2 View of a education system using star schema [15].

Using the integrated data together with data warehousing and online analytical processing application procedures automatically or semi-automatically could provide suggestions to improve teaching and learning procedures.

2.3 Selected Research Model

The current research is focused on enhancing progress and professional standards for accountability systems used in Software Engineering Master Program with a special emphasis on professional competences defined by e-CF Framework [16].

On the technology side, one of the key decisions when designing an LA application is which data to use as predictors and indicators of student progress [13].

Indicators should be considered as factors determining the participation of students in teaching, and as factors that allow for a qualitative analysis of pedagogical technologies and private teaching methods that are employed by teachers. Indicators are used for the analysis of work of both trainees and students.

Dispositional Indicators are the factors or dispositions brought to learning context by a learner. They are in place before a course begins and thus provide some indications of how a student is disposed to and prepared for his learning, and they can also reveal to what extent a learner is likely to respond to any course-related interventions.

Many of these factors are factual and are readily quantifiable, such as age, gender, ethnicity, current grade point average (GPA), prior learning experience, first-in-college, or even financial status.

The usage of dispositional information helps to depict a more complete portrait of students.

Activity and Performance Indicators represent the measures that are the digital fingerprints left by learners as they engage in their learning activities and make their way through the course sequence. Many of these are quantitative in nature. Most projects are gathering these fingerprints where they are most plentiful: via the learning management system (LMS).

Examples include the number and frequency of LMS logins, the amount of time spent on the course website, the number of discussion forum posts, grades, and formative quiz scores.

These quantitative estimates indirectly can also show the shortcomings in the effectiveness of pedagogical technologies and teaching methods employed by teachers (i.e. whether students are overloaded with theoretical materials and other educational activities), as well as the discrepancy between the amount of time spent for students' training and final teaching and learning outcomes [17].

These indicators are relatively straightforward to collect and can be readily analyzed, and the results displayed through visualizations.

Student Artifacts are students' material learning outcomes – essays, blog and discussion forum posts, media productions, etc. Some researchers contend that direct analysis of such artifacts can provide indications of whether students are achieving the needed level of expertise and whether they are demonstrating higher-order thinking skills in their work.

This approach is far less common but has the potential to put LA "closer" to the actual learning by detecting indications of competence and mastery.

An example of a possible mathematical model for quantitative assessment of the obtained level of student's competencies in accordance with the requirements of the educational program for the developed system is presented below.

Calculation of the level of student competencies in the educational program can be performed as follows:

LG – Level Global
$LG = \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} (k_i + s_i + a_i)$, where n is the number of competencies tested,
C_i – level of i-th competence formation,
k_i – level type of C_i competence Knowledge,
s_i – level type for C_i competence Skills,
a_i – level type for C_i competence Attitudes.

The restriction $LG \leq 1$ is convenient for switching to a ten-point rating system, but the limit of the restriction may vary with the national assessment system in a country.

In accordance with the rubric tables, each type of competencies should be checked after performing an automatic computer check, for example in the Moodle distance learning system, it can take the following numerical values:

$k_i = kb_i$ or kd_i or ka_i or ke_i, where kb_i – beginning knowledge level, kd_i – developing knowledge level, ka_i – accomplish knowledge level, ke_i – exemplary knowledge level.

$s_i = sb_i$ or sd_i or sa_i or se_i, where sb_i – beginning skills level, sd_i – developing skills level, sa_i – accomplished skills level, se_i – exemplary skills level.

$a_i = ab_i$ or ad_i or aa_i or $ae_{i,}$ where ab_i – beginning attitudes level, ad_i – developing attitudes level, aa_i – accomplished attitudes level, ae_i – exemplary attitudes level.

The level of each competency under consideration is determined by the results of computer tasks/tests performed by students in accordance with the i-th competence, and is calculated by the following formula:

$$C_i = k_i + s_i + a_i, \text{ where } C_i \leq 1, \tag{1}$$

$k_i = \sum_{j=1}^{m} Tk_{ij}$, where m is the number of tasks/tests Tk_j of *knowledge* level for competence C_i;

$s_i = \sum_{j=1}^{l} Ts_{ij}$, where l is the number of tasks/tests Ts_i of skills level for competence C_i;

$a_i = \sum_{j=1}^{k} Ta_{ij}$, where k is the number of tasks/tests Tai of attitudes level for competence C_i.

The general formula for calculating the level of competencies developed by a student in the educational program can be presented as follows:

$$LG = \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} (k_i + s_i + a_i) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} Tk_{ij} + \sum_{j=1}^{l} Ts_{ij} + \sum_{j=1}^{k} Ta_{ij} \right). \tag{2}$$

Despite the importance of obtaining an individual assessment of the level of competence for each student in the educational program according to the formula (1) for pedagogical studies, the greatest value is the results of the formation of these competencies for a group of randomly selected students of the program from the general population.

In this case, research can be carried out in two ways:

- Comparison of achievements of the same group of students at different time periods (dependent samples). For example, before the beginning of the formation of a certain competence in a group of students and before studying the academic subject (s), and after studying them. Typically, such studies are conducted to determine the effectiveness of techniques used by a particular teacher to study a particular academic subject (s);
- Comparison of the achievements of different groups of students (independent groups randomly selected from the general population) in the development of competencies for analyzing the effectiveness of different teachers using various private methods after studying the same academic subject.

In this case, research procedures should include comparative analysis of given group results, characterized by similar properties (s). For example, for each type of knowledge, skills, or attitudes for each competence separately.

The following options are most likely to be applicable.

If the level of each type of competences will be expressed by a limited numerical interval of points (i.e. on the basis of tests carried out), an interval scale may be used.

However, the task can be refined and consequently give more information for pedagogical analysis if each completed task or test would record achievements of different levels (beginning, developing, accomplished, and exemplary). In this case, a nominal scale can be used.

Each case presented above needs its own mathematical methods that allow for testing a hypothesis and either confirm or reject it in a given research scope.

3 Research Results and Discussion

3.1 Implementation of SECEIP for Competence Evaluation Purpose

Methodology for remote evaluation of competences in Software Engineering (Software Technology) has been developed on the basis of existing European standards (i.e. European Qualifications Framework, European eCompetence Framework and others.) and adapted for online Educational Outcome evaluation [3]. Methodology recommends a set of Outcome models for different levels of Educational Outcomes (knowledge, skills, and competence) measurement and evaluation in SE&ST professional area. The Methodology defines main requirements for planning evaluation, resource alignment, testing implementation, results reporting, and mapping results of evaluation.

The Methodology has the following four general steps [2]:

1. Create simple Template for partners to describe the set of Competences for each dedicated course as a collection of Knowledge, Skills and Attitude/Proficiency Level.
2. Write Rubrics for each item of the Competence evaluation (criteria and grade scale).
3. Create tests (or any other type of assignments) to measure each item of the Competence (separately for Knowledge, Skills and Attitude/Proficiency Level).
4. Calculate the final Competence Evaluation Mark having an integrative formula (using formula with weights).

The SECEIP Portal (http://moodle.kic.teiep.gr/) consists of the following functional parts:

1. The Master Program structure (24 subjects list with a selection feature).
2. Program's Educational Outcome (for each subject).
3. Subjects' Learning Outcomes (for each subject with references to e-CF competences).
4. Information Section (guidance on the use of tests for verification purposes, actual learning outcomes achieved, guidance on the use and interpretation of results, etc.).
5. Administrative section (managing access and user registration, accounting results, instruments of addition and modification of materials from the site, etc.).
6. Community forum (to obtain feedback from students and academic staff).

Based on LMS Moodle platform the SECEIP portal has a big potential for LA implementation using plugins from Moodle library like SmartClass™ Learning Analytics [18].

3.2 Field Studies

The main emphasis has been done on information infrastructure needed to perform effective Learning Analytic at SE Master Program level.

During implementation of the iSECRET project 57 persons (teachers and master students) from 6 countries (Bulgaria, Greece, Latvia, Lithuania, Poland, and Spain) were trained in SECEIP Portal usage. They all completed special online Training Course built in the SECEIP Portal [19].

The project partners created an interactive resource, which consisted of two parts – one on-line training course “How to Use SECEIP” for academic personnel and another on-line training course “How to Use SECEIP” for “Software Engineering” master program graduates. Teaching materials were developed for this online course by project partners. Course materials were uploaded into LMS Moodle, tested, and supplied for localization by selected partners. As a result, three copies of SECEIP portal as Open Educational Resource (OER) are installed in Arta (Greece), Murcia (Spain), and Riga (Latvia).

4 Conclusion

In this paper the authors made an attempt to develop a model and approach of using Learning Analytics in measuring, collecting, analyzing, and reporting of data related to Software Engineering competences obtained by master students at the university. The approach is based on the pilot remote measurement of the components of competence using SECEIP Portal. The developed approach of Learning Analytics usage for measuring competence in SE Master program may be used for re-engineering of Legacy University Management Information Systems and University management executive staff training.

Modern universities are preparing students for the jobs that do not exist yet; their graduates will use the technologies that have not been invented yet in order to solve problems that still are not actualized as problems.

The research is part of the project “Implementation of Software Engineering Competence Remote Evaluation for Master Program Graduates (iSECRET)” run by TTI, contract No. 20151-LV01-KA203-013439, co-financed by EC ERASMUS+ program.

References

1. Misnevs, B., Yatskiv, I.: Data science: professional requirements and competence evaluation. Baltic J. Mod. Comput. **4**, 441–453 (2016)
2. Misnevs, B., Jusas, V., Fernandez Aleman, J.L., Kafadarova, N.: Remote evaluation of software engineering competences. Procedia Comput. Sci. **104C**, 20–26 (2017)
3. The European Qualifications Framework for Lifelong Learning (EQF): Recommendation of the European Parliament and of the Council (2008). http://relaunch.ecompetences.eu/wp-content/uploads/2013/11/eqf_broch_2008_en.pdf. Accessed Nov 2017
4. Shum B., Crick R.: Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In: Proceedings of the 2nd International Conference on Learning Analytics & Knowledge, Vancouver, 29 Apr–2 May 2012, pp. 92–101. ACM, New York (2012). https://doi.org/10.1145/2330601.2330629
5. Siemens, G.: What Are Learning Analytics? Elearnspace, 25 August 2010. http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/. Accessed Nov 2017

6. Ferguson, R.: The state of learning analytics in 2012: a review and future challenges. Technical report, Knowledge Media Institute: The Open University, UK (2012). http://kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf. Accessed Nov 2017
7. Chatti, M., Dyckhoff, A., Schroeder, U., Thüs, H.: A reference model for learning analytics. Int. J. Technol. Enhanced Learn. (IJTEL) **4**(5/6), 318–331 (2012)
8. Chatti, M.A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A.M.F., Wahid, U., Greven, C., Chakrabarti, A., Schroeder, U.: Learning Analytics: Challenges and Future Research Directions. eleed, Iss. 10. http://eleed.campussource.de/archive/10/4035. Accessed Nov 2017
9. Cooper, A.: A Brief History of Analytics A Briefing Paper. CETIS Analytics Series. JISC CETIS, November 2012. http://publications.cetis.org.uk/wp-content/uploads/2012/12/Analytics-Brief-History-Vol-1-No9.pdf. Accessed Nov 2017
10. The Predictive Learning Analytics Revolution: ECAR Working Group Paper, 7 October 2015. https://library.educause.edu/~/media/files/library/2015/10/ewg1510-pdf.pdf. Accessed Nov 2017
11. Roscola, T.: Using Predictive Analytics, Adaptive Learning to Transform Higher Education. http://www.govtech.com/education/Using-Predictive-Analytics-Adaptive-Learning-to-Transform-Higher-Education.html. Accessed Nov 2017
12. Shum, S., Ferguson, R.: Social learning analytics. Educ. Technol. Soc. **15**(3), 3–26 (2012). (Siemens, G., Gašević, D. (eds.) Special Issue on Learning & Knowledge Analytics) http://www.ifets.info Open Access Eprint. http://oro.open.ac.uk/34092. Accessed Nov 2017
13. Brown, M.: Learning Analytics: Moving from Concept to Practice. EDUCAUSE Learning Initiative Briefing (2012). http://www.educause.edu/library/resources/learning-analytics-moving-concept-practice. Accessed Nov 2017
14. Shum, S., Crick, R.: J. Learn. Anal. **3**(2) (2016) http://dx.doi.org/10.18608/jla.2016.32.2
15. Mirabedini, S., Nourani, S.: Int. Res. J. Appl. Basic Sci. (2014). http://irjabs.com/files_site/paperlist/r_2130_140421181323.pdf. Accessed Nov 2017
16. e-Competence Framework e-CF. http://www.ecompetences.eu/. Accessed Nov 2017
17. Puptsau, A.: Professional development of faculty in distance learning at European Humanities University. In: Advanced Learning Technologies, ALTA-2016, pp. 34–36. TSU Kaunas, Lithuania (2016)
18. General plugins (Local): SmartKlass™ Learning Analytics Moodle. https://moodle.org/plugins/local_smart_klass. Accessed Nov 2017
19. Kafadarova, N.: On-line training course “How to use SECEIP for academic personnel and master program graduates”. In: Abstracts of the International Scientific and Educational Conference “Actual Problems of Education (MIP 2017)”, 1–2 June, 2017, Riga, pp. 25–26 (2017). http://www.tsi.lv/sites/default/files/editor/science/Conferences/MIP2017/mip_2017_ed.pdf. Accessed Nov 2017

Ensuring the Academic Workforce Age Balance as a Personnel Management Tool

Oksana Pozdnyakova(✉) and Anatoly Pozdnyakov

Transport and Telecommunication Institute, Lomonosova 1, Riga 1019, Latvia
Pozdnakova.O@tsi.lv

Abstract. One of the important tasks of the process of personnel management is to ensure personnel succession. Considering the fact that in the last decades there has been a certain imbalance in the age and qualification structures of the pedagogical potential and first of all, it manifests itself in the process of the academic staff aging, each institution faces a task of rejuvenating the academic staff. Authors have carried out a research the purpose of which is to consider and characterize the range of problems associated with the age-related imbalance of the academic personnel, to analyse the causes and to suggest the ways of solving these problems in order to optimize the process of personnel reproduction. The study involved the lecturers and students of the Transport and Telecommunication Institute. The results show that the older academic staffs reluctantly interrupts their work after retirement, but the main reason for the process of age imbalance of the academic staff is the unwillingness of the young people to take up an academic career. This fact is related to the absence of interest for teaching among the graduates, also the financial and social conditions of the lecturer's activity. The authors come to the conclusion that it can be recommended to form a competent policy of personnel succession planning this process at the level of the top management of the university, develop financially supported programmes and a set of measures aimed at levelling off the age structure of the academic personnel.

Keywords: Academic workforce · Aging of academic staff
Imbalance of the age level

1 Introduction

The higher education institutions adopt the control systems used in modern organisations. Management policy is formed up by the top management of the university and affects all the processes of the institution performance. Staff management is an especially important area of organisation management, since personnel are the most significant part of the organisation and it represents one of the most valuable development resources.

The quality of the educational services provided by the higher education institution is supported by the activity of the entire staff of the education institution, but it is the lecturer who, in the process of training the specialist, transfers knowledge, skills, competences, social experience, socio-cultural norms and values to the students in

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_63

accordance with modern requirements. Therefore, academic staff is the most important element of the system of higher professional education, determining all other components, as well as the quality of the scientific research and the level of the specialists' training.

The most important tasks in the process of personnel management are the issues of ensuring the efficiency of the interaction with employees, satisfying their needs, comprehension of their motivational attitudes, their ability to form these attitudes in accordance with the tasks which the university faces. Nevertheless, the necessity to create an efficient mechanism for managing the process of personnel reproduction and first of all the reproduction of the academic staff is no less important.

The process of reproduction of this social stratum is ongoing, but in the last decades there has been a certain imbalance in the age and qualification structures of the pedagogical potential. First of all, it manifests itself in the process of the academic staff aging.

The process of the faculty staff aging is a problem of many countries. A number of researchers work upon the aging issue: even taking into account the general trend of aging the workforce, the academic institutions are at the forefront of this process. In many countries the average age of professors now exceeds all other professional groups [1–3]. This problem is of primary importance for Latvia, as well. Certainly, different universities and various departments present slightly different data, but the figures demonstrate, that, in general, 47.5% of Latvian academic staff are over 50 years old, and 26% are older than sixty [4]. Among the academic staff with Doctoral degrees these indicators are even higher – every second of them is over 55 years old (51.5%) [5].

The education system, in general, and each institution, in particular, faces an urgent task of rejuvenating the academic staff by supporting young scientists and doctoral students.

The authors of this article have carried out a research the purpose of which is to consider and characterize the range of problems associated with the age-related imbalance of the academic staff, to analyse their causes and to suggest the ways of solving these problems to optimize the process of personnel reproduction.

2 Methodology

The research was conducted on the basis of the Transport and Telecommunication Institute (Riga, Latvia). The study involved the lecturers of various age categories (about half of the total academic staff), as well as, 103 students of the first and fourth years of studies (73 and 30 respondents respectively), including full-time and part-time students.

While choosing the format of the part of the study with lecturers' participation, the preference was given to informal interviews as the most flexible method of collecting information. The discussed issues concerned the motivation of the employees to continue working after retirement, the degree of satisfaction with academic work, and also the existence of elements of age discrimination at the university.

One of the research tasks was to get information about whether students are connecting their further professional life with the process of teaching at the university and what are their value orientations. To conduct the survey of the students' interests and ambitions, the method of questionnaires with closed and semi-closed questions was mainly used. The results were checked in in-depth interviews with individual students.

3 Ensuring the Academic Workforce Age Balance

3.1 Age Imbalance in the Academic Environment

To a certain extent, the age imbalance in the academic environment is formed by representatives of the older generation of the Faculty who are reluctant to interrupt their work after retirement. There are numerous reasons for this unwillingness, including financial, demographic, and personal ones. It is possible to agree with Druzhilov [3] who suggests the reason for the aged faculty staff to continue their academic activity after reaching the retirement age, and he also states that the length of the productive stage of the professional activities grows together with the increase in the life expectancy of people who have reached the age of 60–65. Therefore, the representatives of the so-called baby-boom generation (born from 1945 to 1964) are healthier, more active and productive than their peers from previous generations.

The authors of this article also agree with the Yakoboski's statement that a rather high degree of satisfaction of the academic staff with their work (both with content and working conditions) is also quite important. Yakoboski [6] examined a large group of aged faculty staff and found out that satisfaction with their academic work was one of the main motivating factors to continue working after the retirement. This conclusion coincides with the results of interviewing the lecturers of the Transport and Telecommunication Institute, implemented by the authors of this article (see Fig. 1). Thus, 38.5% of respondents noted interest in their work among the most important incentives to continue working after reaching the retirement age; the same percentage of respondents noted the opportunity to communicate with people. The habit is important for 18% of respondents. 30% mentioned the opportunity to work in a team where their knowledge and experience are valued as an important circumstance.

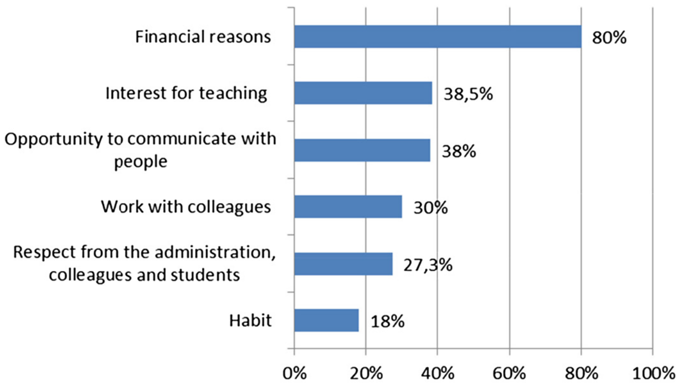


Fig. 1. The main reasons to continue working after retirement.

It could be added that aged academic staff feel psychologically more comfortable in the workplace than people of their age in other areas of the national economy, since there is no such thing as ageism (age discrimination) in respect of older employees at universities. Moreover, faculty staff in adulthood, as a rule, are perceived as highly qualified specialists with great life and professional experience and skills, and therefore worthy of well-deserved respect.

Indirectly, this is confirmed by the survey conducted by the authors of this article, according to which the age of the academic staff was considered by students as an insignificant factor for assessing the quality of the academic performance: only 5% of the interviewed students admitted that they would prefer a young tutor, precisely because he is closer to them by age and, therefore, would better understand their aspirations and problems. The rest of the interviewed students consider the qualification of the academic staff and their benevolent attitude towards their students to be of primary importance.

On the contrary, some authors believe that it is possible to speak about the age discrimination in the sphere of professional interaction in respect to young assistants on the part of honourable professors [7].

Undoubtedly, of exceptionally favourable value for older faculty staff was the abolition of the discriminatory provision of the Higher Education Law by the Constitutional Court of Latvia in 2003; which provided that the elective positions of a professor, an associate professor, and an assistant professor in higher education institutions can be occupied up to the age of 65.

Also, the decision of the Constitutional Court on the discrepancy between the Constitution (the Constitution of Latvia) and the prohibition for working pensioners to receive pensions in full amount, and its abolition played a positive role for retired academic staff. Considering that it is quite difficult to provide the replacement quickly enough so that it would be adequate in terms of qualification and experience (more information on that matter could be found below), educational institutions would eventually face a complex personnel problem.

There must be mentioned the issue greatly influencing the fact that many academic staff, approaching the retirement age, do not even want to think about leaving for a deserved rest. There are the financial reasons for that, including such factors as noticeable reduction in the level of income and, accordingly, in the standard of living after the termination of employment, the rising costs of medical services, and fears of general economic instability, inflation, etc. 80% of the interviewed lecturers mentioned the financial incentive to continue working after retirement as the main one (see Fig. 1).

However, according to the authors of this article, the main reason of a large number of preretirement and retirement aged staff at the university is that there is virtually no one to replace them: there is almost no middle generation of Doctors of Sciences, as well as, the academic staff of the most promising group of 25–34 years of age. This is the situation in the Transport and Telecommunication Institute, and at many other institutions of higher education in Latvia [4], as it was shown by the studies conducted by foreign scientists (for example, [8, 9]).

How can the current situation with staff imbalance be assessed? A number of authors describe it quite dramatically, speaking of the impending collapse of the system of higher education unless urgent measures are taken to prepare the replacement of the older generation of the academic staff [3, 10]. On the other hand, other authors believe that pedagogical activity, the process of teaching is currently changing due to the breakthrough in the technological development, so the higher education institutions do not need as many employees as it used to be earlier, so universities will cope with the training load even after the retirement of the old generation staff [8]. In addition, the fact in support of this is that there is a demographic decline in many countries, and in the case of Latvia, there is also such a phenomenon as high level of the immigration of young people, which results in the drop of the students' number, and this factor, though negative in itself, at least, partially improves the situation with the academic staff.

The issue of the optimal age and status of the academic staff has always provoked discussion. There is no univocal answer to it, because there are a lot of many "strong" and "weak" points taking into consideration each age category of the academic staff. The studies on the problem of aging of academic staff mention such strengths of the so-called "baby-boom" generation (either approaching retirement age or those who have already reached it), as the deepest theoretical knowledge and vast experience. However, the use of outdated methods, as well as, natural fatigue and "burnout" are mentioned as deficiencies [11]. In addition, it is noted that the dominance of the aged academic staff may cause significant financial costs for the university (higher wages, more frequent payment of the sick leave, etc.), not to mention the fact that there will come a time when, by virtue of the natural causes there will be no one to teach new students [9, 10]. In addition, the "washing away" of the middle generation and young people from pedagogical and scientific environment and the imbalance of the age structure of the staff leads not only to the "aging" of the teams, but also to the discontinuity in the reproduction of pedagogical personnel, the difficulties of preserving scientific schools, in the transfer of traditions and research skills.

In turn, young lecturers have new teaching methods, they still have a lot of strength and enthusiasm, and it is easier for them to find a common language with students because of the lesser age difference. Also, the young tutor is more profitable for the university: the salary is lower than the professor's one, and he gets ill less often. At the same time, he/she does not have enough theoretical and practical experience (there is, little reading of the scientific literature, only few publications, a smaller number of the techniques applied, etc.). In addition, the percentage of obtained degrees among the representatives of the younger generation, as a rule, is low or even "zero".

The number of middle age tutors in the structure of the academic staff who, on the one hand, would have experience and skills, and at the same time, retain enthusiasm and motivation for improvement are represented, as it has already been mentioned above, is extremely limited. This layer of academic staff in the university turned out to be thin, first of all, because of the non-competitiveness of wages with the business sector, IT companies, banks, etc. As a result, Doctors of sciences – representatives of the so-called generation X (i.e. those who are now 35–50 years old), who are professionally established in other areas not related to education, are not likely to be motivated to take up teaching work due to a number of objective reasons. Not without a reason this generation is often called as the lost (for the university) generation in the research literature.

3.2 Staff Succession Planning at the University

Considering the fact that the third to the fifth part of the university staff will retire in the next decade, and some universities may lose half of their leadership in the next five years [10], ensuring staff continuity should become one of the most important tasks in personnel management. Certainly, a number of objective circumstances, for example, such as the ongoing demographic changes, are beyond the control of the employer; however, employer can control a lot of factors in order to increase the attractiveness of the workplace. So, the most important task for the personnel management is intensive involvement of young people in the department operation.

However, as researches show, the prestige of scientific and pedagogical activity among young people is low. There are only 0.4 Doctors of sciences per 1 000 inhabitants in the age group from 25 to 34 years in Latvia [5]. The survey and selective interviewing of the first and fourth year students of the Transport and Telecommunication Institute conducted by the authors of this article also showed that the number of students wishing to pursue an academic career does not exceed 6% of the total number of respondents. Moreover, among the first-year students there were 6.8% people considering academic career, while among the fourth year students they constituted only 3.3% (see Table 1). The reasons are obvious: the graduate students are more clearly and consciously oriented on the sphere of professional activity after graduating from a higher education institution; they are better informed and have more reasonable and sensible approach to the assessment of their abilities to work in the academic sphere.

Table 1. Students' orientation on the academic career.

	The first-year students		The fourth-year students		Total	
	N	%	N	%	N	%
Thought about being a lecturer	5	6.8	1	3.3	6	5.8
Thought about being a lecturer but under certain conditions	12	16.4	4	13.3	16	15.5
Did not think about being a lecturer	56	76.8	25	83.4	81	78.7

To understand why students abandon the career of university lecturer they were asked questions about the reasons for choosing this profession or refusing it. The answers of the surveyed and interviewed respondents confirm that their choice of further careers is influenced by the interconnected external and internal factors.

Those respondents who expressed a desire to become a lecturer admitted that in the teaching work they were primarily attracted by the opportunities for self-actualization, namely, by the creative nature of work (72%), the possibility of professional growth (50%), constant communication with the young (37%). There was also a desire to share their knowledge with others (50%). The social bonuses of the teaching work are also important factors for those surveyed: flexible working time (28%), long summer vacation (37%), autonomy in work (35%) (see Fig. 2).

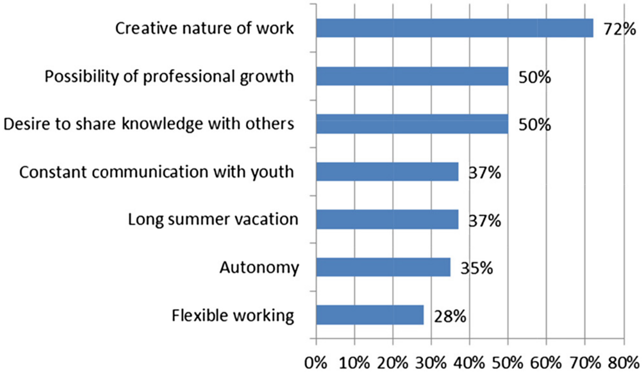


Fig. 2. Reasons to become a university lecturer (Students' opinion).

The reluctance to become a lecturer among the interviewed students is most often explained by the lack of interest in this type of activity (45%). An important role is also played by the fear of the audience (26%) and the lack of personal qualities, which, from their point of view, the lecturers need (28%). As the interviewed students admitted:

– "I cannot be a lecturer, because I am very shy to speak to a large audience".
– "I will never have the patience to explain for the hundredth time what I myself understood a long time ago".
– "I am afraid that students will not listen to me, and I will not be able to cope with them."

An important point that should also be noted is related to the financial and social conditions of the profession of a university lecturer. In the interview students voiced their vision of the future:

– "I like many other young people, want to have my own car, my apartment or house; but I understand that the lecturer's salary will not allow me to implement these plans. By getting a job in a bank or a good IT company. I can earn enough money to make my dream come true."
– "I have certain ambitions; I want to become a leader. Working at a university. I cannot have such opportunities. Most lecturers are engaged in the same thing from year to year – lecturing or doing scientific work because the scientific and pedagogical career of the lecturer at the university includes quite a few steps. For example, at a manufacturing enterprise specialists can move both vertically and horizontally along the career ladder, which may include a large number of different stages."

Thus, the second most important factor for those respondents who do not consider the possibility of teaching is low salary (30%) of the young teaching staff. The apparent lack of career opportunities is also important (23%) (see Fig. 3).

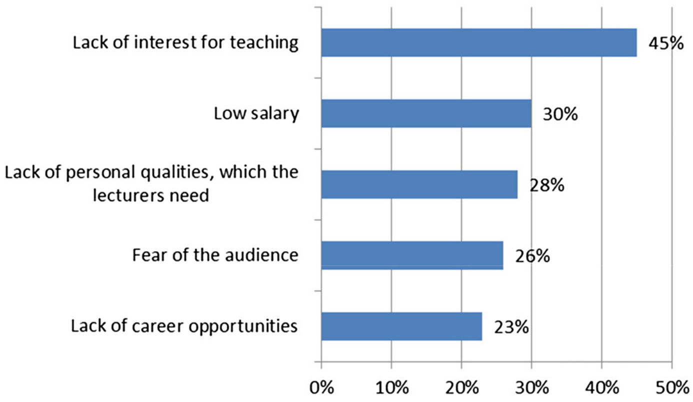


Fig. 3. Reasons not to become a university lecturer (Students' opinion).

However, a fairly large proportion of the surveyed students (15.5%) are not so categorical. They do not exclude the possibility of teaching, but they agree to consider this as a hobby, i.e. as an addition to the main work carried out outside the walls of the Institute. Alternatively, they voice the conditions, observance of which will attract them to self-actualization in pedagogical field.

So, among the factors that could positively influence their choice to become a lecturer, first of all, they mention financial ones, i.e. attractive salary (38.5%). There were also expressed wishes for certain scientific conditions for creativity (work with modern equipment, participation in the scientific projects, grants. etc.) – (46.2%) and the preferential conditions of the study for Doctoral degrees (38.5%). However, researches conducted in Latvia show that a third (35.4%) of Doctoral students and young scientists do not want to be engaged in academic activities [12]. In this regard, it is important for the personnel management to develop the programmes for advance professional selection, and to "recruit" the talented and motivated young people for the future pedagogic career starting from the students' years.

The task of securing and retaining young faculty at the departments seems to be even more difficult and costly than the task of attracting and even targeted training of young people. The turnover of young employees is higher than among staff of other ages. The main reason for the circulation is the dissatisfaction of the employees with their position. In particular, it is the dissatisfaction with wages, conditions and the organization of work, the instability of official position and the inability to make a career. Unfortunately, it is impossible to offer a competitive salary as a factor of academic work attractiveness. A young tutor, who graduated from, for example, the computer technology department, will receive a salary significantly lower than his classmate, who is employed by a large computer company. In addition, financial costs of receiving the postgraduate education are also considered. Therefore, emphasis should be placed, for example, on additionally paying not only for conducting the training sessions, but also for research development. Thus, the income of a new lecturer will become a competitive with the income at the initial position in a large company. The young tutor should be able to continue his study on PhD programmes; therefore, the load can be reduced while retaining wages. Active involvement in the work on the scientific projects will increase young staff's creative potential and will satisfy the need for scientific activity. The creation of a clear and modern academic model of career development will contribute to the retention of young academic staff. Perhaps there is a reason to encourage a balance between work and family life.

At the same time, there should be carried out the work to attract representatives of the middle generation to the teaching process, for example, in practical activities with special emphasis on remuneration. It should be noted that the specialists who have replaced the production activity for the teaching possess the true knowledge of the real situation and can share it with the audience, making the lectures more relevant and informative. However, in order not to lose novelties, it is necessary to encourage a reasonable combination of the academic work at the university and in the industry business. At the same time, according to the authors of this article, it is useful to provide business professionals, for example, with the appropriate additional pedagogical training for the development of teaching and research skills, since not only practical experience, but also comprehension of the corresponding theory and pedagogical methods is required to teach at the university level.

Certainly, it is mandatory to use the rich experience and knowledge of the retired professionals who will help to cope with the shortage of academic staff. Attraction of young lecturers and preservation of the experienced specialists of the older age groups seems to be an incompatible task, however, they do not mutually exclude each other. The key role of these employees, as the authors suggest, is mentoring. At the same time, giving the academic load to a young or inexperienced colleague, the tutor should not lose in the earned payment. One of the ways out is to involve specialists of the traditional retirement age in developing online textbooks. Also, some of the most efficient strategies for retaining the older generation of the academic staff include part-time contracts with retention of retirement benefits, decent rewards for the work done. Perhaps this process is financially burdensome for the university, but none the less, use of the retired specialists proves ultimately to be more profitable than recruiting foreign specialists or beheading the department, since the older generation of academic staff at the department are those who have not only experience and knowledge, but also are holders of academic degrees and titles.

Authors of this article would like to draw attention to another important aspect. Aging of the labour force and its reproduction must become a paramount problem not only for the Heads of departments who are forced to cope with personnel policy problems independently, but, first of all, for the educational institution, as a whole. There must exist an officially approved, clear policy of planning the succession in personnel as a means of managing human resources; there also should be developed the approved and financially supported programme and integrated system of measures designed to level off the age structure of the academic staff.

4 Conclusion

The sphere of education, like any other branch, is able to function normally if it is provided with trained specialists. However, in recent decades there has been noted a process of age imbalance of academic staff due to the increase in the proportion of lecturers of pre-retirement and retirement age. The main reason for this process (besides financial, demographic and personal reasons) is an extremely low number of young and middle-aged lecturers at the departments.

However, in authors' opinion, the crisis situation with the academic personnel is also caused, to a large extent, by the shortcomings in the personnel policy, since the solution of the problems of the reproduction of pedagogical personnel is greatly determined by the efficiency of the management mechanisms functioning, which, in their turn, are formed and used by the university.

The imbalance of the age structure of the staff creates certain problems because leads not only to the "aging" of teams, but also to the discontinuity in the reproduction of pedagogical personnel, the difficulties of preserving scientific schools, in the transfer of traditions and research skills.

Considering the fact that according to the forecasts of specialists, the problem of aging of the academic staff will be relevant for, at least, more than one decade, it is necessary to plan the reproduction of pedagogical personnel proactively, taking certain measures, namely:

1. The use of a variety of measures to attract, reinforce and retain the young lecturers to the university, using material incentives, encouragement of scientific activity, ensuring the career growth; it is also important to develop the programmes for advance selection of students for the future professional career of the lecturer;
2. Involvement of the representatives of the middle-age generation, who have successful experience in the practical professional field, into the teaching process, and simultaneously providing them with the opportunity to improve the level of their pedagogical skills;
3. Development and use of effective strategies for retaining the older generation of lecturers, development of mentoring practice;
4. Formation of a competent policy of personnel succession planning it at the level of the university, the development of financially supported programmes and a set of measures aimed at levelling off the age structure of the academic personnel.

References

1. Larkin, J., Neumann, R.: Ageing academics: workforce priorities for universities. Int. J. Employ. Stud. **20**(1), 3–24 (2012)
2. Loomes, S., McCarthy, G.: The aging workforce: how can Australian universities address future workforce challenges? In: 25th Annual ANZAM Conference 2011. ANZAM, Wellington, New Zealand (2011)
3. Druzhilov, S.A.: Demograficheskaya "yama" i ugroza bezrabotitsy prepodavateley vuzov (in Russian). Mod. Sci. Res. Innov. 7 (2011). http://web.snauka.ru/issues/2011/11/5325. Accessed 28 Sep 2017

4. IZM Parskats par Latvijas augstako izglitibu 2016.gada, Riga (2016) (in Latvian). http://www.izm.gov.lv/lv/. Accessed 28 Sep 2017
5. IZM Sakotnejais novertejums. Pecdoktoranturas petniecibas atbalsts. Eiropas Savienibas strukturfondi 2014.-2020.gadam, Riga (2015) (in Latvian). http://www.izm.gov.lv/images/strukturfondi/IZM_SN_SAM111_Post-doc_160315.pdf. Accessed 28 Sep 2017
6. Yakoboski, P.J.: Understanding the faculty retirement (non)decision: results from the faculty career and retirement survey. TIAA-CREF Institute: trends and issues (2015). https://www.tiaainstitute.org/publication/understanding-faculty-retirement-non-decision. Accessed 28 Sep 2017
7. Onopriyenko, V.I.: Pokoleniya dolgozhiteley v nauke: vzglyad sotsiologa. Sociol. Stud. **4**, 75–85 (2007). (in Russian)
8. Winchester, H.: Staffing issues for universities. In: 3rd Annual Higher Education Summit, Melbourne, Australia, 17–18 March 2005 (2005). http://www.unisa.edu.au/cha/mgmt&gov/smgprofiles/wnchester.asp#corpres. Accessed 20 June 2017
9. Clark, R., d'Ambrosio, M.: Recruitment, retention, and retirement: compensation and employment policies for higher education. Educ. Gerontol. **31**(5), 385–403 (2005)
10. Hugo, G., Morriss, A.: Investigating the ageing academic workforce: Stocktake. University of Adelaide (2010). https://www.universitiesaustralia.edu.au/ArticleDocuments/210/Investigating the Ageing Academic Workforce.pdf.aspx. Accessed 28 Sep 2017
11. Gingras, Y., Lariviere, V., Macaluso, B., Robitaille, J.P.: The effects of aging on researchers' publication and citation patterns. PLoS ONE **3**(12), e4048 (2008)
12. AIP Jauno zinatnieku integracija augstskolas, Riga (2013) (in Latvian). http://www.aip.lv/ESF_par_projektu.htm. Accessed 28 Sep 2017

Mobile Education: Review of Literature on Negative Effects of Multitasking

Kristīne Užule(✉)

Transport and Telecommunication Institute, Riga 1019, Latvia
uzule.k@tsi.lv

Abstract. Mobile education has permeated contemporary higher education market. One of the pillars of mobile education is the usage of mobiles devices, such as computers, tablets and smart phones, in the training process, which has significantly enhanced the quality and accessibility of education due to the continuous ability of students to access training materials through the Internet connection as well as their ability to open resources in download folders at any point in time. Yet, despite the continuous availability of a wide range of training resources, students' performance sometimes is below set expectations. Therefore, the paper has attempted to identify a factor that could possibly interfere with learning. The conducted literature review has pointed towards multitasking as one such plausible culprit. By examining students' performance in tests and critical thinking tasks under the conditions of multitasking and without such exposure, the reviewed studies suggested the link between lower academic performance and multitasking. The reviewed literature also suggested that certain student groups are more prone to multitasking than others, which potentially put them in higher risk of poorer academic performance.

Keywords: Mobile education · Multitasking · Lower academic performance

1 Introduction

Mobile learning refers to learning assisted by mobile applications and resources accessible via the Internet, which could be accessed through various mobile devices, such as computers, tablets and smart phones [1] and which has emerged as a result of availability of wireless technologies and mobile computing [2]. Mobile learning has become a global phenomenon and in fact has been recommended by UNESCO to universities as a tool that could considerably enhance the quality of the training process [3]. These new technologies provide easier and faster access to educational resources which could be used day and night irrespective of a location. Mobile technologies have also increased the speed of academic communication among students, instructors and administration, and as it is commonly known, communication is a foundation of academic progress. Mobile learning is also a solution for distance learning. Students who are in full-time employment or reside in another region are able to pursue their education at a chosen university

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4_64

via mobile solutions, which should promote motivation among students and offer an additional source of revenues to universities.

However, as many other phenomena, mobile learning might carry some potential risks, particularly, for students and therefore it is important to discuss them and develop educational policies that could minimize negative effects of such risks. One observation of the student behavior at Transport and Telecommunication Institute, Riga, has revealed that there is a potential conflict between occasionally low academic results and the expectations for students' attainments, which are high considering an easy access to a wide range of educational materials. Another observation has been such that at least some students appear to have difficulty disconnecting from their mobile devices or messaging. On such grounds the aim of the paper is to explore this conflict by asking the following question – is there a link between academic performance and multitasking? To attain the aim, the following two questions will be addressed. First, does multitasking affect academic performance? Second, considering the fact that not all students appear to have been permanently connected to their mobile devices or messaging and that some of them produce results in accord with the set expectations, are there some student groups that are potentially more vulnerable to multitasking than others? The answers to these questions are accompanied by recommendations aimed at enhancing the quality of the educational process consistently with the two questions posed. The research method used in this study is a secondary data analysis, which is based on literature review.

2 Negative Effects of Multitasking on Academic Performance

As mobile learning is based on using mobiles devices for an easy access to the Internet and any other form of virtual reality, naturally it has opened the door to multitasking [4]. Multitasking represents a concept of simultaneous engagement in multiple activities [5], for example, simultaneous usage of at least two media sources [6], which requires appropriate attention shifts and prioritization of tasks [7]. Examples of multitasking could be listening to music while completing a report or working on two documents at the same time. While under some circumstances the ability to work on a few projects simultaneously might be considered an advantage and in fact is often required [8, 9], in some other contexts, multitasking might result in serious errors, which, according to some researchers [7], have become a common phenomenon at the workplace. Multitasking errors can be divided into two categories – errors by commission, caused by wrong actions that result from interference of unnecessary information in working memory, and errors by omission, caused by not taking any actions as a result of inadequate perceptual speed of triggering switches between tasks [7]. Further research has demonstrated that multitasking errors point to some limits in the cognitive capacities of attention – if a second task is introduced in parallel, attention becomes divided, which is why an error emerges [10]. Some researchers have attributed multitasking errors to rapid switches that underline multitasking activities which often result in the imperfect selection of information and subsequent imperfect processing of information [11]. This explanation should be further specified by adding that prior to imperfect information selection and processing, multitasking might cause erroneous encoding of information caused by improper or incomplete understanding of presented information due to divided attention.

Research into academic performance has supported these theoretical claims. Multitasking in academic settings has been found to lead to errors in understanding of lectures and creation of informational gaps in the long-term memory [12], all of which ultimately results in poorer performance in academic tests [13], but more importantly impedes the development of critical, creative or deep thinking [14] that ultimately distinguishes an average level professional from a future-to-be expert. More specifically, in one study it has been found that spontaneous and volitional multitasking in contrast to undivided attention on one task – lecture comprehension – led to poorer understanding of the lecture content, which was assessed through an academic task distributed to students in the aftermath of the lecture [13]. Similar results have been obtained in an experimental study in which at some point during a lecture some students were asked to complete a distracting task on their laptops [10]. Similar results of poorer academic performance in multitasking conditions have been obtained in simultaneous reading and messaging conditions [15] and in the condition of mobile media usage while listening to a lecture [16]. Yet, in another experimental study, it has been discovered that although simultaneous reading for academic purposes and texting resulted in comparatively similar results with the group that did not engage in texting, the time that was required to complete reading tasks in the multitasking group was significantly longer – by 22–59%, which could be interpreted as a sign of less proficient academic performance [17]. Finally, multitasking has been found to negatively affect academic attainments even of those students who do not participate in the activity but observe their peers indulging in it [10].

One of the questions to ask is, if multitasking has been found to lead to errors, why could students simply not restrain from it at least in academic settings? One obstacle is such that students often do not understand that multitasking is a problem. Many people have been found to believe that by engaging in multitasking, multiple goals can be attained quicker, which is a sign of work efficiency [18], and that multitasking does not interfere with work quality [9, 18]. In another study, students have been found to assert that multitasking does not impede the completion of primary academic activities, which, however, was experimentally proved to be a false assumption [17]. Another reason why multitasking might not be considered a problem is its popularity – it has become a widespread trend and has been on the rise in recent years [5]. Thus, there are three major reasons that stop students from acknowledging the existence of the multitasking problem, which consequently blocks any attempts at limiting the practice – the firm belief that multitasking is efficient, not harmful and fashionable.

The lack of realization of the multitasking problem as well as an easy access to the virtual reality [6] have led to the widespread use of multitasking when studying – the more multitasking is used, the more frequent it becomes [5], which in some cases has transformed into the behavior with addiction-like attributes. In one study, adolescents and young adults have been found to switch to another task, such as checking a media source when completing tasks on a computer, every six minutes [19]. Similar results have been obtained for university students and working adults, albeit without precise intervals for between-task switches [17, 20]. Apart from the high recursive frequency of multitasking, another feature of the addiction-like behavior is the immediate gratification or pleasure that users report to be able to obtain from multitasking [8], which, might subsequently develop into the urge to engage in pleasure-inducing multitasking in the midst of classes [21]. Overall, the lack of acknowledgement of the multitasking problem and resulting behavior that could be characterized as having signs of addiction-like urge make it extremely difficult for students to limit or abandon the practice without external assistance.

To eliminate multitasking at colleges and universities seems next to impossible considering its widespread use across various domains of life. However, colleges and universities should strive to limit the scope of indulgence in multitasking, which, should ultimately raise the levels of students' academic achievements, including the proper encoding, integration and production of information and resulting higher skills of critical thinking, and create the environment that could hinder the development of addiction-like behavior – the less frequent the practice, the lower likelihood of its transformation into addiction-like urge. In what follows, the paper offers three suggestions on limiting online multitasking opportunities.

The first recommendation is to remove access to mobiles devices and stationary computers at least in some lectures because this will reduce the time spent online and thus automatically should increase the quality of learning. In fact, students have been found to obtain higher academic scores when using pens and paper instead of using laptops for note-taking [22]. To ensure that students actually participate in the learning process in such classes rather than engage in non-mobile form of multitasking, it might be appropriate to inform them in advance about formative assessment tests to be distributed at the end of such classes. Obviously, this practice should be introduced incrementally over the duration of the entire course because otherwise it might cause the increase in the anxiety level, which might become counter-productive. However, the fear of counter-productivity should not stop attempts aimed at limiting online multitasking not only for the sake of students, but also for the development of universities – it has been found that students engaging in multitasking not only demonstrate poorer academic performance, but also report higher levels of anxiety and dissatisfaction with the educational process [23]. The second recommendation is to create such online tasks and activities, the completion of which would require mobilization of learning, research and response skills. This could be attained by introducing time-constraints. The third recommendation is to implement an informational campaign that would inform and convince students of potentially negative effects of multitasking. This should ultimately result in the change of perceived needs for multitasking, which, as some researchers have found out [5], can positively impact the multitasking behavior. Colleges and universities typically have psychology or social sciences departments, which have intellectual and professional resources to implement these recommendations incrementally and efficiently.

3 Students at Particular Risk of Multitasking

Although people in general and students in particular have been found susceptible to multitasking at least under certain conditions, some individuals appear to be more prone to multitasking than others [24], particularly those with higher levels of impulsivity and sensation seeking, poorer working memory capacity and attention deficit. In what follows, some details of such research have been presented. First, people with high levels of impulsivity and sensation seeking have been claimed to engage more in multitasking [18, 25] possibly due to the difficulty of blocking impulses or distractors [26], which is linked to lower levels of executive control in the frontal lobe, in other words, poorer executive functions in the brain [18]. In this case, the recommendation is to seek for professional help of medical staff or clinical psychologists who specialize in impulse control and sensation seeking behavior. This help might be beyond the scope of operations of colleges and universities. Second, the working memory of lower capacity has been found to show difficulty ignoring distractors [27] because it cannot withstand the multitasking impulse [9], which results from inability of such individuals to process information for a sufficient period of time [27, 28]. However, research has also demonstrated that the working memory capacity could be enhanced through proper training, which reduces the volume and frequency of multitasking [28]. This training could be developed by psychologists and cognitive scientists employed by colleges and universities. Third, individuals with attention-deficit problems and Attention Deficit Hyperactivity Disorder (henceforth ADHD) are particularly vulnerable to multitasking because their brain demonstrates the lack of ability to sustain longer attention spans caused by a genetic component often associated with the frontal lobe areas [29], which, in turn, promotes continuous switches from one distractor to another, including shifts caused by multitasking stimuli. This, in turn, might promote the Internet addiction [30]. In another study, ADHD patients have also been found vulnerable to multitasking but the cause of this vulnerability was attributed to their inability to control impulses [31]. The recommendation for this group of individuals is similar to that of the first group – seek neurological or psychological help. Importantly, sometimes the symptoms of ADHD in the adult population are not as clear as in children, which is why, sometimes these individuals as well as their teachers are not aware of this problem [29], which might impede their academic progress.

Another category of individual differences pertaining to the degree of engagement in multitasking involves the traits of Five-Factor Personality model – extraversion, agreeableness, conscientiousness, neuroticism and intellect [32]. These traits emerge as a combination of innate and developed features [33]. Overall, the literature review has pointed out to some correlations between multitasking and four traits of the model – neuroticism, extraversion, conscientiousness and intellect. More specifically, the university students who are high on neuroticism have been found to participate more in multitasking due to their frequent need to engage in sensation seeking behavior [5]. Individuals with low conscientiousness score have been found to engage more in multitasking and commit more errors when engaging in it [34]. High extraversion scores have been found to associate with more frequent multitasking that has been highly error-prone [34]. As for intellect trait, the lower the intellect score, the less efficient was multitasking [35].

Because intellect scores have been associated with cognitive skills, these results could be attributed to lower cognitive abilities of individuals to cope with the selection processes between goal-relevant and goal-irrelevant information [36]. The general recommendation for such individuals is to develop skills that could shift them to the opposite end of the trait continuum, e.g. from lower to higher conscientiousness scores. Although these features have innate components, their effects on individuals' life could be altered to a certain degree due to the fact that some other aspects of these traits have a flexible structure that allows for modifications in response to external and internal stimuli. This should eventually help such individuals limit their multitasking activities or at least enhance their ability to engage in multitasking. Psychologists employed by colleges and universities could help develop proper recommendations.

4 Conclusion

The paper attempted to determine if there was a link between the quality of academic performance and multitasking. The review of literature has suggested such a link exists. In accord with the conducted literature analysis, multitasking has been found to tend to result in poorer academic performance possibly caused by divided attention that causes errors of wrong actions, errors of the omission of actions and longer time periods required to complete the tasks. It has also been found out that some student groups are more vulnerable than others to continuous indulgence in multitasking and the commission of a greater number of errors when multitasking. The literature review has suggested that such groups of students exhibited the following traits: (1) poorer capacities of the executive functions, which are responsible for integration of information and allocation of cognitive resources for the completion of a task; (2) poorer capacities of working memory; (3) shorter attention spans; (4) features of attention deficit hyperactivity disorder; (5) inability to block impulses; (6) the need to engage in sensation seeking; (7) higher scores on neuroticism; (8) higher scores on extraversion; (9) lower scores on conscientiousness and (10) lower scores on intelligence. As this paper is based on the secondary data analysis, the next step in this research line could be to verify other research findings in experimental settings in other cultural contexts.

References

1. Schuler, C., Winters, N., West, M.: The future of mobile learning: implications for policy makers and planners. UNESCO Working Papers (2012)
2. Alexander, B.: Going nomadic: mobile learning in higher education. Formante **1**, 59–68 (2006)
3. Briz-Ponce, L., Pereira, A., Carvalho, L., Juanes-Mendez, J.A., García-Penalvo, F.J.: Learning with mobile technologies – students' behavior. Comput. Hum. Behav. **72**, 612–620 (2017)
4. Carrier, L.M., Cheever, N.A., Rosen, L.D., Benitez, S., Chang, J.: Multitasking across generations: multitasking choices and difficulty ratings in three generations of Americans. Comput. Hum. Behav. **25**, 483–489 (2009)

5. Wang, Z., Tchernev, J.M.: The "myth" of media multitasking: reciprocal dynamics of media multitasking, personal needs, and gratifications. J. Commun. **62**, 493–513 (2012)
6. Murphy, K., McLauchlan, S., Lee, M.: Is there a link between media-multitasking and the executive functions of filtering and response inhibition? Comput. Hum. Behav. **75**, 667–677 (2017)
7. Oberlander, E.M., Oswald, F.L., Hambrick, D.Z., Jones, L.A.: Individual difference variables as predictors of error during multitasking. Navy Personnel Research, Studies, and Technology (NPRST/BUPERS-1) (2007)
8. Buhner, M., Konig, C.J., Pick, M., Krumm, S.: Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. Hum. Perform. **19**(3), 253–275 (2006)
9. Pollard, M.A., Courage, M.L.: Working memory capacity predicts effective multitasking. Comput. Hum. Behav. **76**, 450–462 (2017)
10. Sana, F., Weston, T., Cepeda, N.J.: Laptop multitasking hinders classroom learning for both users and nearby peers. Comput. Educ. **62**, 24–31 (2013)
11. Chen, Q., Yan, Z.: Does multitasking with mobile phones affect learning? Rev. Comput. Hum. Behav. **54**, 34–42 (2015)
12. Bailey, B.A., Konstan, J.A.: On the need for attention-aware systems: measuring effects of interruption on task performance, error rate, and affective state. Comput. Hum. Behav. **22**, 685–708 (2006)
13. Barak, M., Lipson, A., Lerman, S.: Wireless laptops as means for promoting active learning in large lecture halls. J. Res. Technol. Educ. **38**, 245–263 (2006)
14. Shao, D.H., Shao, L.P.: The Effects of multitasking on individual's task performance. Int. J. Bus. Strategy **12**(2), 75–80 (2012)
15. Fox, A.B., Rosen, J., Crawford, M.: Distractions, distractions: does instant messaging affect college students' performance on a concurrent reading comprehension task? CyberPsychol. Behav. **12**, 51–53 (2009)
16. Fried, C.B.: In-class laptop use and its effects on student learning. Comput. Educ. **50**(3), 906–914 (2008)
17. Bowman, L.L., Levine, L.E., Waite, B.M., Gendron, M.: Can students really multitask? An experimental study of instant messaging while reading. Comput. Educ. **54**, 927–931 (2010)
18. Sanbonmatsu, D.M., Strayer, D.K., Medeiros-Ward, N., Watson, J.M.: Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. PLoS ONE **8**(1), 1–8 (2013)
19. Rosen, L.D., Carrier, M.L., Cheever, N.A.: Facebook and texting made me do it: media induced task-switching while studying. Comput. Hum. Behav. **29**, 948–958 (2013)
20. Yeykelis, L., Cummings, J.J., Reeves, B.: Multitasking on a single device: arousal and the frequency, anticipation, and prediction of switching between media content on a computer. J. Commun. **64**, 167–192 (2014)
21. Contreras-Castillo, J., Pérez-Fragoso, C., Favela, J.: Assessing the use of instant messaging in online learning environments. Interact. Learn. Environ. **14**(3), 205–218 (2006)
22. Wood, E., Zivcakova, L., Gentile, P., Archer, K., De Pasquale, D., Nosko, A.: Examining the impact of off-task multi-tasking with technology on real-time classroom learning. Comput. Educ. **58**, 365–374 (2012)
23. Wurst, C., Smarkola, C., Gaffney, M.A.: Ubiquitous laptop usage in higher education: effects on student achievement, student satisfaction, and constructivist measures in honors and traditional classrooms. Comput. Educ. **51**, 1766–1783 (2008)

24. Guastello, A.D., Guastello, S.J., Guastello, D.D.: Personality trait theory and multitasking performance: implications for ergonomic design. Theor. Issues Ergon. Sci. **15**(5), 432–450 (2014)
25. Kazakova, S., Cauberghe, V., Pandelaere, M., De Pelsmacker, P.: Can't see the forest for the trees? The effect of media multitasking on cognitive processing. Media Psychol. **18**, 425–450 (2015)
26. Ophir, E., Nass, C., Wagner, A.D.: Cognitive control in media multitaskers. Proc. Natl. Acad. Sci. **106**(37), 15583–15587 (2009)
27. Cornell, P., Riordan, M., Townsend-Gervis, M., Mobley, R.: Barriers to critical thinking: workflow interruptions and task switching among nurses. J. Nurs. Adm. **41**(10), 407–414 (2011)
28. Shipstead, Z., Redick, T.S., Engle, R.W.: Is working memory training effective? Psychol. Bull. **138**(4), 628–654 (2012)
29. Magnin, E., Maurs, C.: Attention-deficit/hyperactivity disorder in adulthood. Revue neurologique. **173**, 506–515 (2017)
30. Yen, J.-Y., Yen, C.-F., Chen, C.-S., Tang, T.-C., Ko, C.-H.: The association between adult ADHD symptoms and the Internet addiction among college students: the gender difference. CyberPsych. Behav. **12**(2), 187–191 (2009)
31. Gawrilow, C., Merkt, J., Goossens-Merkt, H., Bodenburg, S., Wendt, M.: Multitasking in adults with ADHS. Atten. Deficit Hyperactivity Disord. **3**, 253–264 (2011)
32. Privado, J., Roman, F.J., Saenz-Urturi, C., Burgaleta, M., Colom, R.: Gray and white matter correlates with the Big Five personality traits. Neuroscience **349**, 174–184 (2017)
33. Luan, Z., Hutteman, R., Denissen, J.J.A., Asendorpf, J.B., Van Aken, M.A.G.: Do you see my growth? Two longitudinal studies on personality development from childhood to young adulthood from multiple perspectives. J. Res. Pers. **67**, 44–60 (2017)
34. Rose, C.L., Murphy, L.B., Byard, L., Nikzad, K.: The role of the Big Five personality factors in vigilance performance and workload. Eur. J. Pers. **16**, 185–200 (2002)
35. Colom, R., Martinez-Molina, A., Shih, P.C., Santacreu, J.: Intelligence, working memory, and multitasking performance. Intelligence **38**, 543–551 (2010)
36. Uncapher, M.R., Thieu, M.K., Wagner, A.D.: Media multitasking and memory: differences in working memory and long-term memory. Psychon. Bull. Rev. **23**(2), 483–490 (2016)

Author Index

I. Kabashkin et al. (eds.), *Reliability and Statistics in Transportation and Communication*, Lecture Notes in Networks and Systems 36,
https://doi.org/10.1007/978-3-319-74454-4